PUBLIC ECONOMICS

principles and practice

peter abelson

Public Economics: Principles and Practice: On-Line Edition

Contents for Print PDF Publication 2018

Part 8 Public Finance and Taxation

Part 9 Taxation in Practice

Part 10 Multi-Government Systems

Public Economics: Principles and Practice

Preface to On-Line Edition

It is a pleasure to bring to you an on-line, open-access, edition of *Public Economics: Principles and Practice*.

Public economics is the study of the role of government in the economy. As I remarked in the Preface to the Third Edition in 2012: "It is hard to think of a more important and topical field of study in the social sciences". This is surely as true in this turbulent world today as it was then.

It is a core thesis of the book that government is a fundamental instrument of the social welfare of communities. As described in the book, government has traditionally played a fundamental role in both economic development and social welfare in Australia, as indeed it has in many other countries.

In taking this view of government, we are following the European tradition more closely than the American (US) one which tends to elevate markets and distrust government. Of course, the picture is complicated with many variations within these political cultures and with many reasons for and against both governments and markets, which we discuss at many points in the book.

On a personal note, I may say that when I completed the Third Edition six years ago in 2012, I had no intention to prepare another Edition. At that time, I had completed 47 years of professional life. This included ten exotic years of economic consulting work in 13 countries around the globe and 37 years of much more sedentary though very pleasant life in Australian Universities in Sydney (Macquarie and Sydney University).

I was then (in 2012) about to move out of academic life into local government as Mayor of my local community in Sydney (Mosman) for 5 years as well as part-time economic consultancy work, including for the NSW (state) Government, which was to prove a slightly tricky combination at times. This renewed experience in government was a source of great pleasure and fully confirmed for me the importance of government in community welfare.

At the same time, we were observing extraordinary and deeply disturbing economic and political developments around the world: increases in inequality within countries, anger with globalisation of trade and international movements of labour, the vacuous Brexit debates, the extraordinary election of a President of the United States with neither knowledge of nor aptitude for government, and within Australia a distrust of both major parties leading to more minority parties and weak government. A theme common to all these events was deep distrust of government.

I was therefore delighted when McGraw-Hill (publisher of the Second and Third Editions) agreed in December 2017 to revert the rights of Public Economics to me, thus allowing me to publish a revised version of the 2012 text in full on line.

In the short time available since then, I have edited and partially updated the 2012 edition. This includes a minor restructure with the original Part 9 (Social Welfare and Income Redistribution) advanced to Section 7. However, at this stage I have included few new academic references. And I am aware that many other improvements can be made.

Subject to sufficient interest, I plan is to revise the book regularly and to provide this as an ongoing, open access, publication. I invite, and would be most grateful for, any suggestions for improvements and updating.

I hope that this will contribute to an informed and inclusive discussion of these issues that are so important to our individual and collective welfare both here in Australia and in many other countries.

In preparing this online edition I have received valuable research assistance from Thomas Bennett. I also wish to acknowledge the great support that my wife, Jeanne, has given me once again while I have been preparing this online addition.

Dr. Peter Abelson

Sydney, Australia

April 2018

Public Economics: Principles and Practice

Preface to 3rd Edition 2012

Public economics is the study of the role of government in the economy. This study includes nearly all aspects of public expenditure and finance and government economic policies. By convention, it does not include macroeconomic management which is generally treated as a separate subject. Thus, public economics may also be described as the study of the role of government in economic efficiency and distribution.

In the language of economics, the study is both positive and normative. The positive part is the study of how government actually affects the economy. The normative part is the study of how government ought to affect the economy. These studies inevitably involve elements of political science and moral philosophy respectively.

In the preface to his influential book, *The Economics of Welfare*, Arthur Pigou observed nearly 100 years ago that: 'The complicated analyses which economists endeavour to carry through are not mere gymnastic. They are instruments for the betterment of life.' This view motivates many of us who work in the field of public economics. Our fundamental goal in studying public economics is to understand both how the arrangement of economic affairs in a country, whether by markets or by government, affect the welfare of individuals that make up society and how these arrangements can be improved. This study has been a central part of the study of economics or of political economy at least since the publication of Adam Smith's famous book, *The Wealth of Nations*, in 1776. Indeed, it is hard to think of a more important and topical field of study in the social sciences.

Content of *Public Economics: Principles and Practice*

In *Public Economics* I attempt to provide an understanding of the whole field of public economics. This includes the study of the role of markets and of government in the economy, methods of economic evaluation and their applications to public policy, the political economy of government policy making, the principles and practice of public expenditure, policies for social assistance and income redistribution, taxation principles and practice, and multilevel systems of government.

As is customary in public economics, the discussion is based on microeconomic principles. The content of public economics is determined more by its subject matter than by the introduction of new economic methods of analysis. A theme of the book is that basic economic principles apply to government as they do to markets although not always in quite the same way.

I also try to convey both the rigorous and consistent principles that underlie the economic approach to public economics and how these principles can be applied to practical issues. The principles and the practice are complementary. Without the principles, public policies would be poorly based. Without the practical applications, the principles would be of little interest or use. I hope that this book will help readers understand and appreciate how good economic policies can enhance the welfare of the communities in which we live.

In its structure and major content, this third edition is similar to the second. But, in parts, the content has been significantly revised. The world has changed remarkably in the last five years. There is much less

trust of markets following the global financial crisis and yet globalising market forces are stronger than ever. Also, in recent years, both the amount of economic data available and the volume of empirical studies have increased significantly. I have tried to incorporate an understanding of these contemporary forces and changes, as well as other technical revisions, into this book.

Readers of *Public Economics: Principles and Practice*

The book is written as an introduction to public economics. I expect that most readers will have studied economics for one or two years before undertaking a study of public economics or public finance. However, I have tried to produce a book that will also be accessible in the most part to readers with less background in economics, and interesting to those with more knowledge of economics. Readers in one or other of these categories may find selected parts of the book of interest.

The book will doubtless be of most interest to Australian readers, or Australian-based students, because I give mainly Australian examples. However, I also provide international comparisons throughout.

For teachers of public economics who use this text, PowerPoint slides and answers to the discussion questions are available from the publisher.

Acknowledgments

In preparing this and the previous two editions, I have received an enormous amount of assistance and accumulated numerous debts. I must especially thank Glenn Withers (Australian National University), Pundarik Mukhopadhaya (Macquarie University) and John Freebairn (Melbourne University). Glenn provided advice and encouragement over much of the development of the first edition. Pundarik co-authored three chapters (Chapters 3, 6 and 28, now 20) in the second edition and commented helpfully on several other chapters. John Freebairn commented in detail on the taxation Chapters 20 to 27 (now 25 to 32) in both the second and third editions. I was also helped greatly at various times by four outstanding research assistants—Vinita Doedhar, Raymond Li, Sandra Redmond and Amanda Sayegh.

I should also like to thank the many colleagues around Australia who have generously taken the time to review draft chapters of one or other editions of the book (and in some cases, more than one edition). These reviewers include: Francisco Azpitarte (Melbourne University), Fred Argy (past Director of the then Economic Planning and Advisory Commission), Jeff Bennett (Australian National University), Geoffey Brennan (Australian National University), Michael Brooks (University of Tasmania), Tony Bryant (Macquarie University), Helen Cabalu (Curtin University of Technology), Stephen Cheung (Sydney University), Harry Clarke (La Trobe University), Michael Coelli (Melbourne University), Michael Cosgrave (Australian Competition and Consumer Commission), Michael Dobbie (Macquarie University), Brian Dollery (University of New England), Stephen Dowrick (Australian National University), Joanne Epp (New South Wales Treasury), Hayley Fisher (Sydney University), Peter Forsyth (Monash University), David Gadiel (Director of Applied Economics), Partha Gangopadhya (University of Western Sydney), Noel Gaston (Bond University), Jenny Gordon (Productivity Commission) and anonymous colleagues at the Productivity Commission who reviewed various chapters, Eric Groom (NSW Independent Pricing and Regulatory Tribunal), Ross Guest (Griffith University), Ann Harding (University of Canberra), David Hensher (University of Sydney), Shane Johnson (Australian Treasury), Glenn Jones (ex–Macquarie University), Roselyne Joyeux (Macquarie University), Geoffrey Kingston (University of New South Wales), Andrew Leigh (ex–Australian

National University), Kwang Ng (Monash University), David Maynard (ex–New South Wales Treasury), Paul Miller (University of Western Australia), Neville Norman (Melbourne University), Rod O'Donnell (University of Technology), Chris Panousis (Griffith University), Johnathan Pincus (Adelaide University), David Prentice (La Trobe University), John Quiggin (University of Queensland), Maria Racionero (Australian National University), Jeff Sheen (Macquarie University), Noree Siddique (Flinders University), David Throsby (Macquarie University), Andrew Torre (Deakin University), Tom Valentine (ex–University of Western Sydney), Michael Warlters (New South Wales Treasury), Stephen Whelan (Sydney University), Peter Whiteford (University of New South Wales), Ross Williams (Melbourne University) and Erkan Yalcin (University of New England). These reviews were a great assistance.

Needless to say, none of the reviewers is responsible for any shortcomings in the book. My thanks are also due to the School of Economics, Sydney University, which has provided me with a convivial home for the past six years. I have also been assisted greatly in the preparation of the work at various times by Jane Oldroyd, Helen Boneham, Clara Liosatos and Laura Billington. I wish, especially, to thank Clara for working long hours over the last year in her spare time to make this publication possible and Laura for standing in at the last critical moments.

I am also grateful to McGraw-Hill's proof reader, Anne Savage, and Publisher Kate Aylett–Graham for their excellent and helpful assistance in the final stages of the production.

Above all, I thank my family and my wife, Jeanne, for their support and tolerance. They predicted the amount of time required to complete the third (and earlier) editions much more accurately than I did.

Peter Abelson

University of Sydney

May 2012

About the Author

Peter Abelson graduated from Oxford University and gained an M.Sc and Ph.D in Economics from London University. Peter started his career in international economic consulting from 1966 to 1974. In 1974 he joined the Economics Department at Macquarie University in Sydney, where he held a Personal Chair in Economics from 2001 to early 2006. From 1993 to 2006, Peter was Secretary of the Economic Society of Australia. From 2006 to 2012, Peter was a Visiting Scholar and lecturer in public finance at Sydney University, a part-time Principal Economic Advisor to the NSW Treasury and a Director of Applied Economics, an economic consultancy practice. Peter was a councillor on Mosman Council (Sydney) from 1979 to 1989 and was Mayor from 1983 to 1985. In 2012, he returned to Council as the first popularly elected Mayor of Mosman for a period of 5 years. He continued to work part-time as an economic consultant. Peter is married to Jeanne and they have three (adult) children, David, Julian and Antonia.

Part

1

Nature of Government

Chapter

1

Foundations of Public Economics

I think of the state as an association of individuals engaged in a cooperative venture, formed to resolve problems of coexistence and to do so in a democratic and fair manner.

Richard Musgrave

Government and the State ♦ General Functions of Government ♦ Economic Functions of Government ♦ Historical Perspective ♦ Public Policy and Principles of Economic Analysis

In 1998, the Center for Economic Studies at the University of Munich held a week-long workshop to discuss two major contrasting visions of the role of the state presented by two famous economists, Richard Musgrave and James Buchanan. In the words of Hans-Werner Sinn, Musgrave's book, *The Theory of Public Finance* written in 1959 was "a landmark in the field for decades… laying the foundations for the way public finance economists around the world think about the state".[1] On the other hand, in 1986 Buchanan received the Nobel Prize in Economics as the principal founder of the School of Public Choice. Their backgrounds and views were quite different.

Richard Musgrave spent the first 30 years of his life in Germany, had a strong social democratic background and was schooled in German public finance. As shown by the quotation at the head of this chapter, drawn from this workshop, Musgrave was a strong believer in the role of the state as a provider of community wellbeing, though as a refugee from Nazi Germany he was also deeply aware of how things could go terribly wrong.

On the other hand, by his own description, James Buchanan came from a long line of deeply independent Scots-Irish immigrants to Tennessee in the south of the United States, who bitterly recalled the defeat in the Civil War in the 1980s and for whom there could be "no sense of membership in a genuine national community". He then studied at the University of Chicago where "the Chicago economist learns how economies work rather than how economies might be controlled". These experiences led Buchanan, like the Chicago school, to advocate a strongly constrained role of government.

These differences lie at the heart of study of public economics.[2] Economic and social well-being require good government. A good government provides the institutions and rules that allow markets to flourish, effort to be rewarded and people to lead prosperous and healthy lives. It supplies essential public services, including law and order, economic infrastructure,

[1] The workshop and papers were published in Buchanan and Musgrave, 1999, *Public Finance and Public Choice: Two Contrasting Visions of the State*, MIT Press, Cambridge, USA. The quotations above are from this book.

[2] They also lie at the heart of the differences between the Obama and Trump administrations in the US.

and basic health and education services. It seeks to unify the community, assist the poor and protect the vulnerable.

To provide these goods, government must have wide-ranging powers. Government must have the power to defend the community from external aggression, to provide internal security, to make and enforce the laws necessary for economic activities and to provide the basic public goods that a society needs. Moreover, it must have the power to fund these collective activities. As John Stuart Mill remarked, there is 'in almost all forms of government agency, one thing which is compulsory: the provision of the pecuniary means. These are derived from taxation'.[3]

But governments may misuse their powers. Governments may act arbitrarily and oppressively, levy excessive taxes well over the value of services supplied, engage in corrupt activities, rule over stagnating or declining economies and increase rather than reduce the divisions in society.

These are complex matters. What are the objectives of government? How can governments achieve these objectives? What are the appropriate economic functions and policies of government? And what constraints should be put on government? While, as economists, our main contribution is to assess and propose suitable economic policies, in doing so we have to understand the political and social environment in which we are living and working.

Accordingly, in this opening chapter we start by discussing the relationships between the government, the state and the individuals who make up the state. We then move on to examine the main functions of government. The last part of the chapter outlines some of the major principles of economics that we will draw on throughout the book to analyse government policies and actions.

Box 1.1 Government and state

Government and state are often described by reference to each other. A government is the governing power of the state. On the other hand, a state cannot exist without a government. How then are government and state distinguished?

The state represents the political organisation of society. It is the way in which the members of a society associate and organise under a system of government for the purpose of reaching and implementing collective decisions. The primary purpose of the state is to maintain order and security, which it does by a system of law backed by force. The principal characteristics of the state are sovereignty and a defined geographical area within which the state is sovereign. In this area, the state has supreme and exclusive authority. The government is the practical embodiment of this authority.

The word 'state' is also used to describe some politically organised communities that are not sovereign states. Examples are the states of the United States of America and the Commonwealth of Australia, which are members of a federal union. In these countries, the federal union is the sovereign state. It alone has the authority to make treaties with foreign states and international bodies. Also, it usually has the superior authority in internal affairs.

Because government acts for the state, it has many of the state's characteristics, including its sovereignty powers. Government has a monopoly on the legitimate use of force through the military forces and the police. It has the power to compel all residents of the state (not only citizens) to obey the rules it makes on behalf of the state.

But there are important differences between the state and the government. All citizens are members of a state, not of government. States go to war; governments declare and conduct war. The head of the state may be different from the head of the government. The state, not government, owns public resources. A state has a high measure of continuity, although not necessarily permanence; governments can change regularly. Opposing the state is widely regarded as treason, but citizens can oppose the government. In democracies, the government's power of sovereignty is subject to the periodic consent of the people.

[3] Mill (1848), *Principles of Political Economy*, p. 944, in W.J. Ashley (ed.) (1909).

Government and the State

In political science, government is commonly defined as 'the governing power of the state'. The government directs the affairs of the state and is the vehicle for collective action. However, in a democratic state, the exercise of this power is subject to the periodic consent of the people in national elections. The relationship between government and state is described more fully in Box 1.1 (above).

The nature of government is revealed more pragmatically in the institutions of government. As shown in Figure 1.1, in a democracy government generally consists of three branches: the legislature, executive and judiciary. The legislature makes the laws; the executive develops government policy and carries out the business of government; the judiciary interprets and applies the laws. The legislature often consists of two houses of parliament: one house initiates legislation and the other reviews it.[4] The executive branch includes public administration and services, the armed forces and police, statutory authorities and public trading enterprises. The judicial branch generally comprises a hierarchy of courts.

Government services can also be divided into budget and non-budget sectors, based on the sources of finance. The **budget sector** consists of all activities that are financed out of the annual government budget. This includes most government services, the military, the police, and the courts. It also includes not-for-profit statutory authorities such as universities and public broadcasting authorities which are financed partly through the budget. These authorities are usually established by legislation and have separate legal identity and some autonomy, sometimes including revenue raising powers. However, their autonomy is constrained by their statutory responsibilities, their reliance on government funding and government's power (subject to the governing legislation) to appoint and dismiss directors.

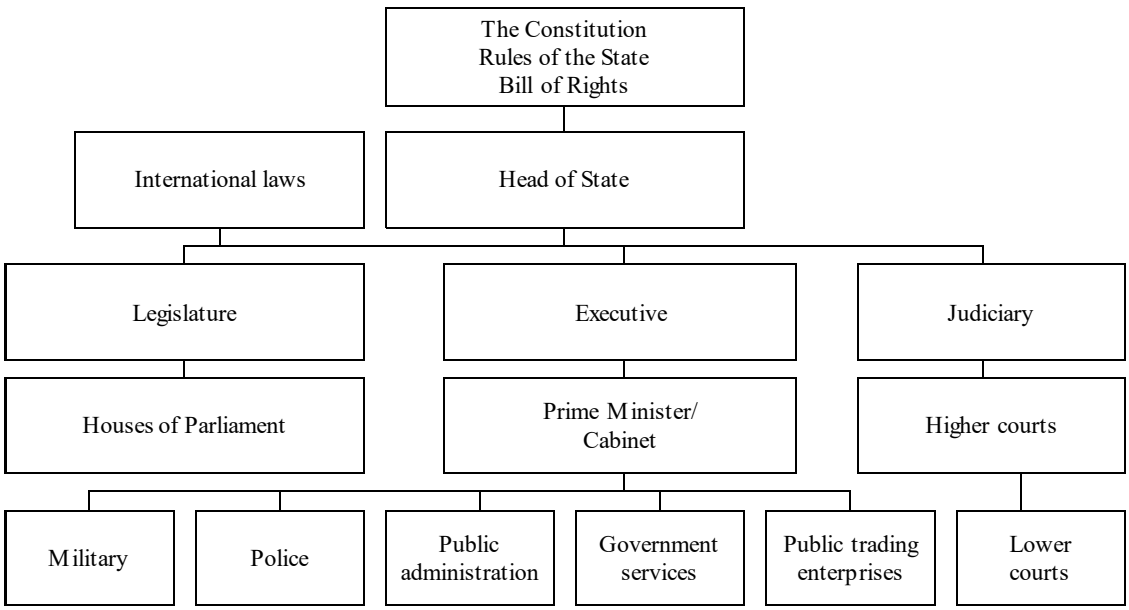


Figure 1.1 Typical institutions of government in a democracy

[4] There are exceptions, such as New Zealand, where the legislature consists of only one house.

The non-budget sector consists of public trading enterprises (PTEs) that finance all or most of their activities by trade revenues. PTEs are usually wholly owned by the state, but in some cases the state owns a majority share. The revenues generated by PTEs give them some autonomy. However, government generally controls the board of directors, regulates major investment and pricing decisions and may provide financial support through the budget.

So, what is government? In popular discussion the term 'government' usually refers to the agencies of government that 'govern the state'—the agencies which make rules and policies and provide the main civilian administrative services. Publicly employed teachers and broadcasters may not think of themselves as part of the government. However, when discussing economic issues, such as government expenditure, the term 'government' generally applies to all budget-sector services, including the armed forces, budget-financed statutory authorities and the courts. More broadly still, the **public sector** includes PTEs as well as budget-sector services. In this book we discuss both the role of government and the broader economic issues of the whole public sector.

Constraints on government

In nearly all states, government is based on a written constitution.[5] The constitution represents a superior law that defines the basic institutions and rules of the state, and the methods of appointment and powers of the head of state and the government. The constitution is a critical legal constraint on government actions. Constitutions are generally hard to change, with change requiring a high degree of public assent. In Australia, a change to the constitution requires the support of the majority of all voters and the support of the majority of voters in at least four of the six states. Along with a constitution, most countries also have a bill of rights that provides individuals with legal protection against oppressive government action. Australia is one of the few countries that does not have such protection of individual liberties.

A related constraint on the power of government is the **separation of powers** between its three main branches. This separation has long been viewed as essential for the avoidance of tyranny by government.[6] The independence of the judiciary is especially important. This independence enables it to strike down laws that are contrary to the constitution and to punish members of the executive for actions contrary to law, including illegal use of power against citizens. The separation of powers has been adopted most fully in presidential forms of government, as in the United States and France. In these countries, the legislature and the head of the executive branch (the President) are elected in separate elections. Members of the legislature are not normally members of the executive. By contrast, in a parliamentary system, as in the UK and Australia, members of the executive are generally members of the legislature. The governing party controls both the legislature and the executive arm of government. However, this control is rarely absolute. When the legislature consists of two chambers, the second chamber usually has some legislative power and the executive branch of government may not control both chambers. Also, the executive branch usually has many departments. Both ministers of departments and officers within departments often fight fiercely over policy control and administrative territory.

A federal form of government is a further important constraint on the power of central government. In a federation, such as the United States and Australia, power is shared between the central government and the states that make up the federation, with the powers of the states supported by the constitutions of the respective countries. The federal structure of government dilutes the power of the centre, encourages competition between government authorities and sets up more checks and balances within the overall government structure. In a

[5] All countries, except Israel, New Zealand and the United Kingdom, have a written constitution.
[6] See Montesquieu, *De L'esprit des Lois* (1748, *The Spirit of the Laws*).

unitary state, like France and the UK, the central legislature has supreme authority to make laws for the whole country.[7]

Finally, national governments are constrained by international laws and agreements. Whether for reasons of good international governance or self-interest, governments have been increasingly signing international agreements, including agreements on economic issues such as trade, the environment including climate change, and labour conditions. For example, Australia's trading laws are derived almost entirely from treaties entered into through the (then) General Agreement on Tariffs and Trade (now World Trade Organization) and at the time of writing it appears that Australia will join the Trans Pacific Partnership trade agreement. On the other hand, Brexit and Trump's America First policies are currently undoing many long-standing international agreements.

General Functions of Government

Views on the general functions of government depend critically on (1) how people regard the relationship between the individual and the state and (2) how people view the relative effectiveness of markets and government. We discuss these two issues in turn.

Historically there have been two main views of the state: the organic and the mechanistic. In the **organic view**, society is a natural organism like a human body. Each individual is a part of this organism, but the body has an existence in its own right. The organic model implies that the state has an underlying being and value. Thus, Aristotle (c. 340 BC) wrote that 'the city or state has priority over the household and over any individual among us. For the whole must be prior to the parts … It is clear that the state is both natural and prior to the individual'.[8] It follows that government should act in the interest of the state as well as for individuals. However, in the 20th century these views were associated with the fascist regimes in Germany and Italy and the totalitarian regimes of Stalinist Russia and Maoist China and became discredited. In these countries, government viewed itself as synonymous with the state. Opposition to government was viewed as opposition to the state. There were few constraints on government power and individual interests were sacrificed to the supposed interests of the state.

Organic state
The state has value in itself

The **mechanistic model** of the state is more democratic. In this model, individuals form a nation for the purpose of making collective decisions. The state is a mechanism for improving individual welfare. Government is an instrument of the people who make up the state and a means of enabling collective actions. The government is subject to law. Government actions and social outcomes are judged by their effects on the welfare of individual members of society. In the words of Mill (1859), 'the worth of the State, in the long run, is the worth of the individuals comprising it'.

Mechanistic state
The state is a means for improving the welfare of the citizens

As described in Box 1.2, leading political writers have advocated a mechanistic model of the state for over 300 years. Today, it is the dominant model. Most countries have a constitution that embodies two key principles—the sovereignty of the people and respect for human rights. The state is designed to serve the people, not the converse. This model of the state is also the basis for most economic analysis. Government actions are judged by whether they enhance the welfare of individuals. This is the approach adopted in this book.

However, the organic view of the state is not solely of historic significance. Some Islamic states may be regarded as versions of an organic society. Even in secular states, many citizens view society as more than the individuals that comprise it and consider a strictly individualistic approach to collective issues an unattractive and inappropriate view of society.

---

[7] The UK government has recently devolved significant executive powers to local parliaments for Scotland and Wales, but it has retained major taxation and monetary controls.

[8] Aristotle, *The Politics,* Penguin Classics (1962), p. 29.

Box 1.2 Origins of the mechanistic state

Thomas Hobbes (1651) was one of the first writers to articulate a mechanistic model of the state in his famous book *Leviathan*. Writing in turbulent times in England, Hobbes argued that, because men are largely selfish and competitive, the natural condition of mankind is 'war of every man against every man' in which there is 'continual fear, and danger of violent death; and the life of man, solitary, poor, nasty, brutish, and short'.[9] A state ruled by a sovereign (man or assembly) with absolute authority is the only way to avoid anarchy. In accepting the benefits of the state, citizens were deemed to give up their natural rights and to be under an obligation of obedience to the sovereign. However, men give up their natural rights only for an implicit 'covenant' with the sovereign for protection. If the sovereign fails to provide protection, men are not obliged to continue to obey the sovereign.

John Locke (1690), in *The Second Treatise of Government*, produced a more attractive model of the state. Locke held that man has three main natural rights: to life, liberty and property. Men establish a state whose function is to protect these rights. He argued, moreover, that the state rests on an implied social contract. The government is entrusted with power to protect these natural rights, but its powers are limited broadly to these responsibilities.

If the government fails to protect man's natural rights or oversteps its responsibilities, the citizens are entitled to establish a different government. While Locke was writing, the British parliament established by legislation that the executive (King William II) could not tax without its consent. Thus was born the fundamental democratic principle that no man can be taxed without his consent or that of his representative.

Nearly 100 years later, the authors of the American Declaration of Independence (1776) declared in ringing tones that all men are endowed with:

> certain inalienable rights, that amongst these are life, liberty and the pursuit of happiness; that to secure these rights, governments are instituted among men, *deriving their just powers from the consent of the governed;* that whenever any form of government becomes destructive of these ends, it is the right of the people to alter or abolish it …' (author's italics).

Shortly afterwards, the *French Declaration of the Rights of Man and of the Citizen* (1789) declared likewise that the purpose of political association is to preserve the natural rights of man.[10]

Alternative views of the role of government

Within the dominant mechanistic model of the state, there is a wide range of views about the appropriate role of government. These views depend on judgements about the effectiveness of markets and government and on beliefs about individual freedom and social justice.

At one end of the political spectrum, **libertarians** strongly support private property and advocate very limited government control over markets. For example, Nozick (1974) defends private property on moral grounds: he argues that people have natural rights, liberty is the ultimate political good and state intervention is morally wrong except in limited circumstances to protect natural rights. On the other hand, empirical libertarians, such as Hayek and Milton Friedman, argue that most markets are efficient and that government intervention usually reduces total welfare. They criticise provision of support above subsistence because the resulting institutions, notably taxation, achieve little justice and reduce economic efficiency. In *Free to Choose*, Milton and Rose Friedman (1980, p.158) conclude that welfare programs "weaken the family; reduce the incentive to work, save and innovate; reduce the accumulation of capital; and limit our freedom".

At the other end of the spectrum, **socialists** advocate a command and control economy. Socialists see the distribution of rewards in market systems as highly unequal and unjust. They view private markets as instruments of exploitation rather than as voluntary exchanges. Owners of capital exploit labour and use resources to meet the demands of the rich. In their

[9] Hobbes (1651), *Leviathan*, pp. 88–89, R. Tuck (ed.) (1991)

[10] Rousseau's *The Social Contract* (1762) was a major influence on the Declaration. But Rousseau viewed man's rights as social, not natural. He argued that sovereignty resides in the people and that laws should reflect the general will of the people.

view, government generally allocates resources to meet social needs more efficiently and fairly than do markets. Socialists generally support strong executive government, with fewer constraints imposed by the separation of powers within government.

Between these extremes, there are many views about the role of government that cannot be readily classified. Simplifying considerably, we distinguish two further groups: conservatives and social democrats. **Conservatives**, like libertarians, regard markets as generally efficient. They question active government macroeconomic policies and provision of goods. They generally favour constraints on an activist government and the separation of powers. Because of the element of compulsion in government action, conservatives usually place the burden of proof is on government intervention rather than on markets. However, they are more inclined than libertarians to acknowledge that government can do some things more effectively than the private sector. They also consider that the state has an obligation to support the poorer groups in society and that it is in the state's interest to do so to avoid civil disorder.

Social democrats recognise that markets are often productive but place more emphasis on the failure of markets to allocate resources efficiently and the unequal outcomes. They view public ownership and private property as pragmatic issues. Government should adopt the ownership mix that best meets society's goals. Private property is viewed as a means towards achieving social goals rather than as an end in itself. The social democratic tradition emphasises equity. Redistribution is viewed as an appropriate and desirable function of the state. For example, in *A Theory of Justice* John Rawls (1971) argues that government has a responsibility not just to relieve poverty, but also to achieve a 'just' distribution of income. Many Nobel prize winners in economics, such as James Meade, James Tobin and Amartya Sen, hold a similar view.

Table 1.1 summarises the views of each of these groups. Much of the debate in public economics about the role of government revolves around these views of the efficiency and fairness of markets and government.

Table 1.1 Summary of views of the role of government

Political classification	*Freedom*	*Markets*	*Equity*	*Role of government*
Libertarians	Government a major threat to freedom	Markets are very efficient	Government often fails to improve equity	Very limited
Conservatives	Government may infringe freedom	Markets are generally efficient	Government should alleviate poverty	Moderate
Social democrats	Government can enhance freedom of less well off	Markets are often inefficient	Government should ensure minimum income for all and reduce inequality	Considerable
Socialists	Government frees people from want	Markets are very inefficient	Government should provide full equity	Very high

Economic Functions of Government

To discuss the economic functions of government it is useful to start with a simple model of the economy. As shown in Figure 1.2, resources are transformed into goods and services via two major interacting sets of markets, namely factor and product markets. Product markets drive the demand for factors of production (land, labour and capital). On the other hand, incomes generated in factor markets drive the demand for products. In first year texts, this is described as the circular flow of income.

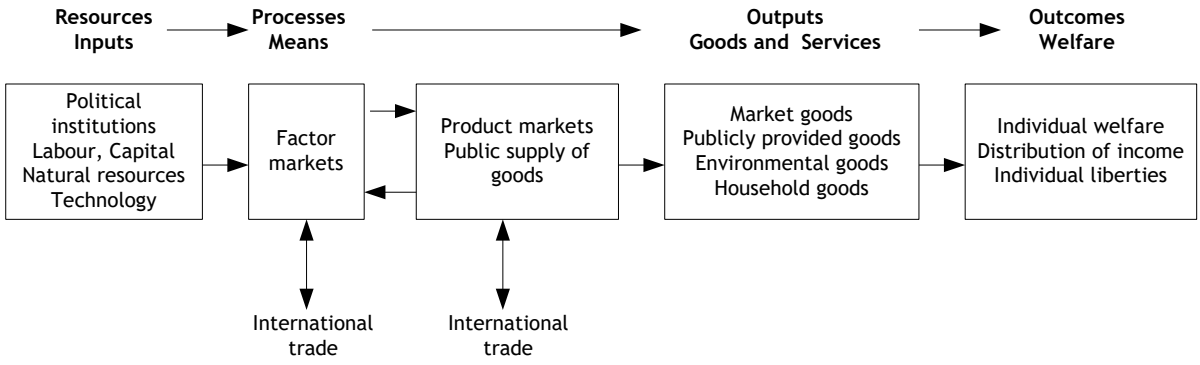


Figure 1.2 The economic process

Goods may also be produced in the public or household sectors. Moreover, factors of production and goods may be imported or exported. However, what matters ultimately is individual welfare. This depends not only on the quantity of goods produced, but also on income distribution, non-market goods and political freedoms.

Government can intervene in the economic process in various ways. It can regulate any factor or product market, regulate international trade and produce goods in lieu of, or in addition to, market output. It can redistribute ownership of resources and it can redistribute incomes earned in factor markets.

Establishing institutions and ground rules. The starting point for any discussion of these roles is that government must establish the institutions and rules that allow markets to work. Markets require an established foundation of law and an effective system of property rights.

Property rights include the right to use an asset, to allow or exclude its use by others, to collect the income generated by the asset, and to sell or otherwise dispose of the asset. In the words of the World Bank (1996, pp. 48–49):

> Property rights are at the heart of the incentive structure of market economies. They determine who bears risk and who gains or loses from transactions. In so doing they spur worthwhile investment, encourage careful monitoring and supervision, and create a constituency for enforceable contacts. In short, fully specified property rights reward effort and good judgement, thereby assisting economic growth and wealth creation.

Laws are required to define and protect property rights; to set rules for exchanging these rights; and to enforce contracts. They are also required to establish market structures and to establish rules for entry into and exit out of markets. Enforcement of these laws requires a fair and predictable judicial process. In their eloquent book, *Why Nations Fail*, Acemoglu and Robinson (2013, p. 43) show that "economic institutions are critical to determining whether a country is poor or prosperous". They also show that, critically, the political institutions should be inclusive to reward the many and not extractive to reward the few.

International experience in the 1990s showed the importance of the rule of law for market operations. In what was known as the "Washington Consensus", international agencies such as the International Monetary Fund routinely prescribed structural reform and competitive markets to cure underdevelopment, especially in Latin America and the then command and control economies in Eastern Europe. They encouraged privatisation (selling state-owned businesses) and liberalisation. However, these policies failed because the basic institutions and laws were inadequate, the political institutions were extractive and public corruption was common in many of these countries. By the late-1990s, the policy emphasis was on first getting appropriate institutions and laws in place (Rodrik, 2015, pp. 159-167).

More dramatically and recently, a major cause of the meltdown in financial markets between 2007 and 2009 was the failure of governments to effectively regulate major financial institutions. Manifestly the private institutions failed to self-regulate and the rescue of the financial system cost taxpayers many billions of dollars and, as we will see in Chapters 2 and 29, put many governments heavily into deficit and debt.

Given that markets are operating under a functioning political and legal system, what economic services should government provide? The common approach in economic analysis is to assess which goods are supplied inefficiently by markets and might therefore be provided by government. This may appear a biased approach. Why not ask which services government can supply efficiently and leave the rest to markets? There are two main answers.

First, if individuals can achieve their goals through voluntary trades in decentralised markets, government direction of individuals or firms is unnecessary. In voluntary trades, all parties expect to benefit. Collective action that simply replicates the actions of free individuals serves no purpose. When compulsion exists, it is always possible that someone is coerced into an exchange against their will. Thus, the general function of government is to accomplish things by collective action that individuals find difficult or impossible to accomplish separately.

Second, competitive markets are very good at producing the goods that people want. Markets coordinate decentralised decisions by generating price signals that reflect demands and relative scarcities without the need for a central planner estimating all these prices. As Adam Smith observed: 'It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.'[11] Moreover, it does not matter that each individual:

> neither intends to promote the public interest, nor knows how much he is promoting it … He intends only his own gain, and he is in this, as in many other cases, led by *an invisible hand* to promote an end which was no part of his intention … Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of society more effectually than when he really intends to promote it (author's italics).[12]

Smith's famous metaphor of the invisible hand guiding the competitive economy has now been formalised by economists into the First Theorem of Welfare Economics (see Chapter 3). This theorem states that economies with perfectly competitive markets produce the highest possible level of output and welfare, given the resources and technology available. Although few markets are perfectly competitive, markets generally do quite a good job at producing the goods that people want.

Three weaknesses of markets. Despite the invisible hand, markets have three main sets of weaknesses: (1) Market failures: under certain conditions markets allocate resources inefficiently and fail to supply goods that consumers want; (2) Income inequality: markets often result in inequitable distribution of income; (3) Macroeconomic instability: markets may not produce full employment and price stability.

Market failures
Situations in which markets allocate resources inefficiently

There are four main forms of **market failure**: public goods, externalities, imperfect competition and information failures. Markets undersupply public goods, such as national security, health and education, which have widespread community benefits. Typically, firms cannot charge for all the benefits that accrue from providing these goods. On the other hand, markets oversupply goods that have damaging (externality) effects on firms or individuals that are not involved directly in the exchange, as occurs with environmental damages. Without regulation, markets produce excessive levels of pollution. Third, many markets are not competitive. As Adam Smith famously observed: 'People of the same trade seldom meet

[11] Smith (1776), *The Wealth of Nations*, p. 14, in E. Cannan (ed.) (1937).
[12] Smith (*ibid*., p. 423).

together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices'.[13] When markets are not competitive, firms tend to charge excessive prices, restrict output and use excessive resources. Fourth, efficient trading requires that the trading parties are well informed. When poorly informed, one or other party may be made worse off by the trade.

Turning **to income distribution**, in competitive markets workers receive approximately the value of their marginal product. However, because of differences in abilities or opportunities, or simply from accident of birth, the value of the marginal products and therefore the income of individuals vary greatly. All markets, competitive or otherwise, tend to produce an unequal distribution of income. Indeed, there is no guarantee that markets will provide everyone with an adequate living to avoid destitution.

Macroeconomic instability typically manifests in price or employment instability. As the Great Depression of the 1930s and the recent global financial crisis showed, markets do not automatically self-correct. Depressed private investment and consumption, along with fear of lending, may create inadequate demand and underemployment. John Maynard Keynes (1936) pointed out that, even if wages were to fall in flexible labour markets, lower wages will not produce full employment in the short run if the aggregate demand for goods is deficient. Less dramatically, markets may produce inefficient fluctuations in economic activity over business cycles.

Three potential government functions. These three forms of market weakness provide three potential economic functions for government—allocation, redistribution and stabilisation functions.[14] The allocation function requires government to respond to market inefficiencies: to provide public goods, regulate externalities, promote competitive markets or regulate market power and protect individuals from information failures. The redistribution function requires government to alleviate poverty and reduce income inequality. The stabilisation function involves management of aggregate demand to achieve price stability and high levels of employment and to reduce fluctuations in output over the business cycle.[15]

However, the extent to which government should carry out any economic role depends not only on the weaknesses of markets but also on the ability of government to deliver improved outcomes. Government actions may be costly or unfair. Government has limited information, is not always efficient and has limited control over private responses to its actions. The key issue is therefore whether government intervention, on balance, increases social welfare.

Determining whether government actions increase social welfare is complicated because government actions often have conflicting outcomes. Efficient use of resources may increase income inequality. On the other hand, government actions to redistribute income may reduce the total income available. Attempting to resolve the trade-off between these outcomes is a major issue in public economics. Moreover, governments themselves may not act efficiently.

Historical Perspective

Economists' views of the economic functions of government, like popular views, have changed greatly over the last two centuries. For much of the 19th century, most economists advocated a limited role for government. Then, for some 100 years from about 1870 to 1975, economists advocated that government should exercise an increasing role in economic management. In the last quarter of the 20th century, as command and control economies collapsed, many economists viewed markets as more effective than government and sought to

[13] Smith (*ibid*., p. 128).

[14] Following Musgrave (1959), government is often described as having three economic branches: an allocation, distribution and stabilisation branch.

[15] As noted in the Preface, the stabilisation role is the subject of macroeconomic texts and is not discussed in detail in this book.

limit the role of government. In the first decade of the new millennium, the global financial crisis along with growing global environmental concerns led to another reversal, with government widely viewed as an essential instrument of global welfare. However, in the last few years popular distrust of governments has again grown, and this has doubtless influenced the thinking of economists.[16]

Mill, perhaps the most influential economist in the 19th century, held that:

> Restricting to the narrowest compass the intervention of a public authority in the business of the community ... laisser-faire, in short, should be the general practice: every departure from it, unless required by some great good is a certain evil.[17] ... The great majority of things are worse done by the intervention of government, than the individuals most interested in the matter would do them, or cause them to be done, if left to themselves because government has one great disadvantage: an inferior interest in the result.[18]

Notwithstanding his advocacy for limited government, Mill agreed with all Adam Smith's proposed government functions (Box 1.3). He regarded the provision of security, the general rule of law and the specific laws of property and contracts as vital for capital accumulation and economic growth.[19] He supported government provision of a wide range of public goods, either because of the general expediency and efficiency of collective action or because private providers could not be remunerated by user fees. These goods included the coining of money,

Box 1.3 Adam Smith's views on the functions of government

Adam Smith's famous book, *The Wealth of Nations*, published in 1776, was an intense reaction against the policies of the 'mercantilists' who then dominated economic thinking. Mercantilists argued that the government should control trade. They called on government to promote industry and exports and to protect the economy from imported goods. Smith strongly opposed these protectionist policies, foisted upon a 'venal parliament by rapacious merchants and manufacturers', along with the complex apparatus of controls that these commercial interest groups imposed on other individuals, including widespread limitations on apprenticeships, which were a direct restraint on trade. Smith believed that competitive markets would produce most of the goods that people want along with the invention and capital accumulation to create a steady increase in wealth. However, he held no illusions about the actual nature of markets. 'Whoever imagines that masters rarely combine is as ignorant of the world as of the subject. Masters are always and everywhere in a sort of tacit, but constant and uniform combination, not to raise the wages of labour above their actual rate.'[20]

It was the government's job to provide an appropriate institutional structure for markets and to remove any restrictions on competition and trade.

Smith held that government has three other main duties. First, to protect society from external threats and violence. Second, to administer justice and to protect every member of society from internal injustice or oppression, including commercial injustices. Third, to erect and maintain 'certain public works and certain public institutions, which it can never be for the interest of any individual, or small number of individuals, to erect and maintain; because the profit could never repay the expense of any individual, or small number of individuals, though it may frequently do much more than repay it to a great society'.[21] Although Smith thought that transport infrastructure (such as roads, bridges, canals and harbours) could fairly be paid for by users, he opposed private provision because monopolists with private turnpikes would abuse the system. He also supported substantial public expenditure on education for youth and for persons 'of all ages'.[22] This was necessary for a stable and strong country, the understanding and acceptance of law and the wellbeing of society as a whole.

---

[16] These are of course generalisations. Many exceptions can be found.
[17] Mill (1848), p. 950 in the 1909 edition.
[18] Mill (*ibid*., p. 947).
[19] Mill (*ibid*., p. 881). There needs to be certainty that 'those who produce shall consume and that those who spare to-day will enjoy to-morrow'.
[20] Smith (*ibid*., pp. 66–67).
[21] Smith (*ibid*., p. 651).
[22] Smith (*ibid*., pp. 739–740).

prescribing standard weights and measures, surveying and mapping, raising dykes to reduce flooding, irrigation works, street paving and lighting, roads, canals and harbours, hospitals and support for scientific exploration and research academies. Nor did the role of government stop there. Mill acknowledged that the principle of consumer sovereignty could not apply without many exceptions.[23] Government should look after those who cannot protect themselves or who fail to choose wisely in life. Moreover, Mill strongly supported public provision of public education. He argued that all society suffers from the consequences of the ignorance and want of education in their fellow citizens.

Marx (1867) held a more complex view of government. He held that governments were partly responsible for exploitation of the working classes. However, by taking industry into public ownership, government could radically change property relationships. This would stop exploitation of labour. It would also avoid the collapse of the economy as the return on capital would fall with the growth of capital, thus causing private investment to fall and unemployment to rise. State ownership and control of resources and the means of production were necessary to maintain capital investment and ensure full employment of labour.

Increasing the role of government. During the first three-quarters of the 20th century there was increasing public and intellectual support for larger government. The widening of the popular vote to full adult suffrage in many countries increased the legitimacy of government. It also increased the groups in society who could make claims directly on government.

Wars had significant impacts. In both the First and Second World Wars many governments, generally with public support and some success, controlled a high proportion of their country's economic resources including labour and allocated it in accordance with national priorities, introduced extensive price controls and rationing, and expanded the tax base.

In 1920, the first major work devoted to welfare economics, Arthur Pigou's *Economics of Welfare*, was published. Pigou was concerned not only with poverty and inequality but also with market misallocations of resources. He identified as major problems negative externalities such as industrial pollution, and the misuse of natural resources such as fisheries, forests and farming land. He also argued that people consume too much in the present and save too little. He advocated that government should tax negative externalities, protect the environment and support long-term public projects such as water supply and afforestation.

In the 1930s, the Great Depression had a major impact on views of market economies. In the United States, the world's largest capitalist economy, unemployment reached 25 per cent and national output fell by 30 per cent from its peak in 1929. These falls in employment and output were widely attributed to capitalist speculation and banking failures (Galbraith, 1955). On the other hand, the recovery in the late 1930s was attributed to government expenditures in both democratic states like the United States and fascist states like Germany. President Franklin D. Roosevelt introduced a raft of public works and social welfare programs, known as the 'New Deal', to alleviate poverty, unemployment, sickness and old age. Keynes (1936) provided the intellectual support for increased public expenditure. With the collapse of private consumption and failure of private investment to respond lower interest rates, only increased public expenditure would lift aggregate demand sufficiently to drive the economy back to full employment. The Depression scarred public attitudes towards capitalism. In 1943, Joseph Schumpeter wrote:

> The public mind has by now so thoroughly grown out of humour with it as to make condemnation of capitalism and all its works a foregone conclusion—almost a requirement of the etiquette of discussion. Whatever his political preference, every writer or speaker hastens to conform to this code and to emphasise his critical attitude, his freedom from 'complacency', his belief in the

[23] Mill (*ibid*., p. 575) 'The proposition that the consumer is a competent judge of the commodity can be admitted with only numerous abatements and exceptions.'

> inadequacies of capitalist achievement, his aversion to capitalism and his sympathy for anti-capitalist interests.[24]

After the Second World War, several developments increased the economic role of government. First, many states increased their control over production. China and most Eastern European countries followed the then Soviet Union and became command and control economies. Also, governments in many mixed economies took over ownership of major industries such as coal, steel, railways and power generation. Control of these basic industries was thought necessary to secure capital investment, full employment and growth, to protect workers and to control the prices of necessities.

Second, many mixed economies developed strong welfare cultures and services. The *Beveridge Report* (1942), named after its author, which argued that governments should establish a comprehensive social security fund to provide "from cradle to grave" for the poor, unemployed, sick and elderly, was highly influential in the UK where it was produced and elsewhere. For the next 30 years, many governments expanded the suite of welfare services along with income support and public health and education services.

Also, the 1950s and 1960s were a period of decolonisation and independence for many developing countries. Many newly independent governments saw their role in development as pre-eminent. The state would mobilise and direct resources and eliminate social injustices. Markets were part of the discredited colonial system that enriched the coloniser and impoverished the colonised. Prebisch (1970), Head of the Economic Commission for Latin America, articulated a popular school of thought that development would be built on government-owned and supported industries, constructed around high import tariffs and quotas. Also, many countries followed the Indian model, with its centrally planned development within a democratic framework.

Reducing the role of government. By the late 1970s, as government spending in many countries exceeded 40 per cent of gross domestic product (GDP), and income transfers accounted a high proportion of tax revenues, attitudes towards government spending and taxes started to change. Increased public expenditure appeared to reduce rather than increase rates of economic growth. And the command and control economies of the Soviet Union, Eastern Europe and elsewhere were generally performing poorly. US Presidents Ronald Reagan and George H.W. Bush (1981–92) and the UK Conservative governments (1979–97) were strongly committed to smaller government and lower taxes. In his inaugural Presidential address in 1981, Reagan proclaimed that 'government is not the solution to our problem; government is the problem'. In Australia, Labour governments from 1983 to 1996 and Liberal Coalition governments from 1996 to 2007 favoured, in varying degrees, reforms to encourage competitive markets and restrain the growth of public expenditure (see Chapter 2).

Following the lengthy economic boom of the 1990s into the early 2000s, markets were widely perceived to be innovative and productive. While basic public institutions and laws were required, Government was widely perceived to be a constraint on economic growth.

The development of public choice theory tended to support these critical views of government. Public choice theory views government not as an abstract benevolent entity that automatically pursues the public interest but as an agency run by individuals who pursue their own interests subject to constraints of elections and various governmental processes. Arrow's Impossibility Theorem (Arrow, 1951) showed that all methods of making collective decisions, typically by voting, fail quite mild tests for satisfactory process or outcome. Downs (1957) described how elected representatives may pursue their own objectives and make public decisions contrary to the interests of the governed. Niskanen (1971) showed how the private interests of public servants may expand and bias public output. In the last 50 years

[24] Schumpeter (1943, p. 63).

public choice theory has been greatly extended and used increasingly to interpret government behaviour (see Chapters 9 and 10).

The new millennium has witnessed extraordinary changes, major uncertainty, the splintering of countries and parties, and no clear overall direction for the role of government.

Following many years of economic growth and confidence in markets, in 2007 a tsunami hit the global financial system in 2007. Within 12 months, asset values on the world's largest stock markets halved, giving the lie to the claim that equity markets followed efficient pricing principles. Several of the world's largest financial institutions (Lehman Brothers, Merrill Lynch, Bear Stearns, AIG Insurance, Freddie Mac and Fannie May housing mortgage businesses, the Royal Bank of Scotland) and businesses (General Motors and Chrysler) went bankrupt and were so large as to threaten international financial stability. Many more large companies (including Goldman Sachs) would likely have been bankrupted without massive fiscal and monetary interventions from many of the world's largest governments. There were major management failures as markets gave managers incentives to take financial risks where they gained the rewards of success, but others would bear the costs of failure. As revealed by the (US) Financial Crisis Inquiry Commission (2011), Wall Street bankers engaged in numerous irresponsible, self-serving, deceitful and unethical practices. These practices were implicitly endorsed by the major credit rating agencies (Moody's, Standard and Poor's and others) who failed to manage their conflicts of interest (being paid by the same large companies that they were rating) and systemically tolerated by the government regulators. Manifestly, under-regulated markets failed. As Reinhart and Rogoff (2010a) showed, financial crises are not one-off exceptions; rather they are endemic features of capitalist economies.

Concurrently global concerns about climate change were influencing views of markets and government. The private interests of carbon-producing energy suppliers and consumers were clearly inconsistent with the collective interest in a stable environment. But governments everywhere were finding it difficult to articulate and implement practical climate change policies either domestically or internationally.

On the other hand, there has been increasing inequality in many countries (Atkinson, 2015). This has led to growing opposition to globalisation and notably to free trade, free movement of labour and to working under international rules. There is pressure on governments to restrict markets and to minimise free movement of labour across national borders. Thus, the British people voted (narrowly) to leave the European Union (Brexit – British Exit) and the Americans elected a President on an America First ticket who also believed in minimal government while at the same time promising to help the left-behind groups in the US.

Thus, after nearly two decades of the new millennium, economists and political scientists are still trying to work out an optimal role for government in an uncertain and unequal world.

Public Policy and Principles of Economic Analysis

Finally, in this introduction to public economics we consider how economic analysis can contribute to public policy making. The issues are complex and diverse. They may involve, among other things, how much to spend on hospitals and roads, whether and how to regulate markets, what assistance to provide to different groups in society or how to raise revenue. But, whatever the issue, four main questions must nearly always be answered.

1. What are the policy objectives?
2. What are the options for achieving these objectives?
3. What are the implications of each option?
4. Which is the preferred option?

There is an important difference between these questions or at least between the types of answers that may be given. Understanding this difference is a crucial theme of public economics.

Questions 1 and 4 involve value judgements. The answers require judgements about what is desirable: judgements that society would be better off with one objective or outcome than with another. Judgements of desirability, of good and bad, are ethical or moral judgements. Such judgements are not true or false—they are matters of personal values or views. Economic analysis that requires ethical judgements is known as **normative economics**. On the other hand, questions 2 and 3 deal with facts, with how an economy works. Question 2 deals with how objectives can be achieved. Question 3 deals with the consequences of government policies. These are empirical questions and the answers are true or false, or accurate or inaccurate (although establishing accuracy may not be easy). Economic analysis that deals with facts is described as **positive economics**.

Normative economics
Economic analysis that involves a value judgement

Positive economics
Economic analysis of how an economy actually functions

The branch of economics known as welfare economics provides a powerful framework for identifying social objectives and a consistent method for evaluating social choices (see Chapters 7 and 8). However, economists have no special expertise or authority to make ethical judgements and we must distinguish carefully between matters that require such judgements and matters of evidence or fact.

Nevertheless, economic analysis is nearly always required to identify the means available to achieve desired social objectives and the implications of these measures. Economic analysis is needed even when government plans to redistribute income, for example to the unemployed or the elderly. In such cases, estimates of the costs and consequences of the options are generally required.

Some applications of economic principles to the public sector

Although economic principles were developed mainly to explain behaviour and outcomes in markets, most of these principles apply also to the public sector. In particular, microeconomic analysis—the principles of demand and supply and the role of markets—is essential for developing and evaluating public policy. Some examples are highlighted here. Many more arise in later chapters.

Opportunity cost
The value of output or leisure forgone by using a resource in one way rather than another

Scarcity and opportunity cost. All economics is concerned with scarcity. Scarcity implies choice and choice implies opportunity cost. A consumer chooses a good because she believes that its value exceeds that of the good forgone (the opportunity cost). Nearly all public policy involves an opportunity cost. Public expenditure on one service reduces it on another. Regulations to protect the environment almost always involve forgoing some other goods. The concept, and estimation, of opportunity cost are fundamental to public policy.

The value of trade. Trade is the basis for extracting gains from markets. When parties trade in factor or product markets, both parties to the trade expect to gain. Similar gains occur in regional or international trade. Trade between regions or between countries can make each region or country better off and the households in each region and country on average better off. However, issues arise when third parties lose out as they may do (see Chapter 34).

The role of incentives. Incentives matter in the public sector as they do in the private. Consider Adam Smith's description of university teaching 200 years ago:

> It is in the interest of every man to live at much as ease as he can; and if his emoluments are to be precisely the same, whether he does or does not perform some very laborious duty, it is certainly his interest … either to neglect it altogether, or if he is subject to some authority which will not suffer him to do this, to perform it in as careless and slovenly manner as the authority will permit.[25]

[25] Smith *(ibid.*, p. 718).

The economic analysis of government recognises that politicians and public servants are guided by private motives as well as by social ideals.

The role of prices. Pricing principles apply to many public sector activities, including to non-marketed goods. When there is no price, there is almost always excess demand, as occurs for most health care services and urban roads. On the other hand, the supply of factors of production generally falls with lower prices. Thus taxes that reduce take-home income usually reduce labour supply. But price theory also tells us that all price changes can be decomposed into a substitution (relative price) effect and an income effect and that these may have opposite effects on labour supply. The analysis of taxation requires an understanding of how individuals react to changes in prices.

The principles of efficient production. These principles apply to public production as to private. The efficient use of factors of production requires that they be employed up to the point where the value of their marginal product equals their marginal cost. Also, the division of labour (specialisation) generally increases efficiency. These principles should inform public methods of production as they do private methods.

Working with markets. Governments should generally work with markets rather than against them. Markets represent powerful forces of demand and supply. Government attempts to control interest rates are a classic example of futile attempts to repeal the law of markets. As Adam Smith observed:

> In some countries the interest of money has been prohibited by law … This regulation, instead of preventing, has been found from experience to increase the evil of usury; the debtor being obliged to pay not only for the use of money, but for the risk which his creditor runs by accepting compensation for that use. He is obliged … to insure his creditors from the penalties of usury … No law can reduce the common rate of interest below the lowest ordinary market rate.[26]

More positively, governments can use market instruments, notably taxes and subsidies, to bring about desired policy outcomes, for example to control pollution or to encourage private health insurance. Market instruments often achieve desired policy outcomes more effectively than regulations (see, for example, Chapter 13).

Social Welfare. Certainly, determining public policy is considerably more complex than explaining private actions. To explain private decisions, economists assume that individuals aim to maximise their utility and that firms maximise profits. In assessing public policy, the over-riding objective is assumed to be to maximise social welfare, which is assumed in turn to depend in some complex way on the welfare of all individuals in the community. The objective of maximising social welfare in one or other form underlies all economic evaluation of public policy (see Chapters 7 and 8).

Social welfare
The collective welfare of society based on the welfare of the members of that society

Rational decision making. Rational decision making in the public sphere, as in the private, requires that decisions be based on marginal benefit and cost. Firms maximise profits by producing up to the point where marginal revenue equals marginal cost. Consumers maximise welfare by consuming goods up to the point where marginal private benefit equals marginal cost. In the public sector, as an optimising principle, government should aim to ensure that the marginal social benefit of policies and projects exceeds or at least equals marginal social cost. When this is not achieved, it is always possible to make at least one person better off without making anyone else worse off.

In summary, governments have much more complex social objective than individuals exchanging labour or goods in market. However, most principles of economics apply to the public sector as strongly as to the private. Notably, the principles of choice, opportunity cost and trade, the role of prices in determining demand and supply, the effects of incentives on behaviour and the principles of efficient production apply equally in the public sector as in the private.

[26] Smith (*ibid.*, pp. 339–340).

Summary

- Government exercises the sovereignty powers of the state and can control, by regulation or taxation, the activities of firms and individuals within the territory it governs. However, the mechanistic (democratic) view of government implies that government is expected to act in the interests of the citizens.
- In economic terms, general government includes all services financed by the government budget. The public sector includes public trading enterprises as well as budget-sector services.
- The major economic functions of government include establishing the institutions and rules that allow markets to function, the allocation of resources when markets are inefficient, provision of social welfare and macroeconomic management.
- Views of the role of government have changed numerous times over the last 200 years. Following the global financial crisis in 2008-09, growing inequality and the threat of climate change, many people regard government as essential to economic prosperity and social welfare.
- However, just as markets can fail, so can government. And there is also wide distrust of government.
- In economic analysis, the aim of public policy is typically to maximise social welfare, which is a function of the welfare of all individuals in the community. This requires normative and positive economics. Normative economics deals with what is desirable, which depends on ethical judgements. Positive economics explains how social objectives can be achieved and the implications of public policies.
- Notwithstanding the social welfare objectives of most governments, the major principles of economics apply to government as to markets. These include the concepts of scarcity and opportunity cost, the role of trade and prices, the role of incentives and the principles of efficient production.

Questions

1. What is the difference between the state and the government?
2. Why is the rule of law not necessarily the same as the rule by government?
3. What would be the features of a minimal state? Why are property rights essential to well-functioning economies? Explain why legal enforcement of property rights may be viewed as a public good.
4. Are government decisions on what citizens can eat, drink and inhale consistent with a mechanistic view of government?
5. What is the invisible hand? How does it work?
6. What are the main economic functions of government? What is the basis for these functions?
7. Was the global financial crisis that erupted in 2007 due to market failures? If so, what failures? Are financial crises endemic to capitalist societies?
8. Suppose, as economists often do, that the aim of public policy is to maximise social welfare. What would be the main components of social welfare? Should the welfare of citizens and foreign residents have the same weight in a social welfare function?
9. What are some differences between not-for-profit statutory authorities and public trading enterprises?
10. Which of the following are normative statements? Why?
    i. Reducing unemployment will increase the rate of inflation.
    ii. Lower interest rates will reduce the number of unemployed persons.
    iii. Redistributing income from the rich to the poor increases social welfare.
    iv. Economists should not make normative statements.
11. Rational decision making implies that decisions should be based on marginal considerations. Can this principle be applied to the public sector? If so, how?
12. Give some examples of how price theory may apply to public decisions.
13. What arguments can be put for government subsidies for university students? Are these positive or normative arguments?

Further Reading

Acemoglu and Robinson, J.A. (2012), *Why Nations Fail, The Origins of Power, Prosperity and Poverty*, Crown Publishers, Random House, USA.

Atkinson, A.B. (2015), *Inequality, What can be done?* Harvard University Press, Cambridge, MA.

Buchanan, J.M. and Musgrave, R.A. (1999) *Public Finance and Public Choice: Two Contrasting Visions of the State*, MIT Press, Cambridge, MA.

Easton, S. and Perin, P. (2010) 'Market efficiency and the global financial crisis', *Australian Economic Review*, 43, 464–468.

Financial Crisis Inquiry Commission (2011) *The Financial Crisis Inquiry Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapters 1 and 3, Worth Publishers, New York.

.

Hillman, A.L. (2003) *Public Finance and Public Policy*, Chapter 1, 'Markets and Property', Cambridge University Press, Cambridge, UK

Holcombe, R.G. (1998) 'The foundation of normative public finance', pp. 1–42, in F. Thompson and M.T. Green (eds), *Handbook of Public Finance*, Marcel Dekker, New York.

Mill, J.S. (1848) *Principles of Political Economy*, Book 5, Chapters 1, 5–8 (pp. 159–166, 256–367), in J. Riley (ed.) (1994) Oxford World Classics, Oxford University Press, Oxford.

Musgrave, R.A. (1985) 'A brief history of fiscal doctrine', pp. 1–60, in A.J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics*, Vol. 1, North Holland, Amsterdam.

Reinhart, C.M and Rogoff, K.S. (2010a) *From Financial Crash to Debt Crisis*, Working paper 15795, National Bureau of Economic Research, New York.

Rosen, H.S. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapters 1 and 2, McGraw-Hill, New York.

Chapter

2

Government in Practice

Now, what I want is, Facts. Facts alone are wanted in Life.

Mr Gradgrind in *Hard Times*, Charles Dickens

The Legal Basis of Government ♦ Measures of Government Size ♦ Government Regulation ♦ Government Expenditure and Revenue ♦ Government Budgets: Deficits, Debt and Assets ♦ International Comparisons ♦ Determinants of Government Expenditure

Few economists would agree with Mr Gradgrind that all we need to know is facts. However, theory without facts is of little use. To understand the role of government, we need to know the legal basis of government, concepts of government size, the nature of government regulation, and the nature of government expenditure and revenue. Central to understanding government is an understanding of the annual budget.

This chapter addresses these and related issues. We start by outlining the legal basis of government in Australia. We then discuss concepts of government size and regulation. The central part of the chapter describes government expenditure, revenue and debt in Australia. Some international comparisons of government expenditure are also provided. The last part of the chapter discusses the major drivers of government expenditure.

The Legal Basis of Government[1]

The Australian Constitution, as laid out in the *Commonwealth of Australia Constitution Act 1900*, provides the legal basis for government in Australia.[2] The Constitution created a federation of the six then British colonies. It created the Commonwealth government (now known as the Australian government) and established the rules governing relations between the Commonwealth and the states (as the colonies became). It also established the powers of the Commonwealth parliament, the executive and the judicature. Unlike many other constitutions, the Constitution does not provide a formal bill of rights for individuals. Individual rights were assumed to be protected by common law and the exercise of democracy.[3] Box 2.1 describes some other important features of the Australian Constitution.

[1] This section draws on Saunders (1997).

[2] The Act came into force on 1 January 1901.

[3] The Constitution declares some individual rights, including freedom of religion and just compensation for property acquired by the Commonwealth. Parliament has now passed laws against racial and sexual discrimination. Australia is also a signatory to the Universal Declaration of Human Rights (United Nations).

Box 2.1 Features of the Australian Constitution

The Australian Constitution responded to local political issues and drew on UK, US and Swiss practices. It drew the system of parliamentary government from the United Kingdom. The elected parliament would be responsible for appointing the government (because government must have the confidence of the parliament). Parliament would determine public policies and budgets, which would be implemented by the executive arm of government.

From the United States, the Constitution drew the concept of two elected parliamentary houses. The House of Representatives (the 'Lower House') is elected by direct popular vote on the principle of one person, one vote. The Senate (the 'Upper House') would provide the states with power by providing each state with equal representation in the Senate. As in the United States, the judiciary is formally independent and responsible for interpreting and protecting the Constitution.

From Switzerland was drawn the idea of a referendum to approve changes in the constitution.

Changes to the constitution require the approval of a majority of electors nationally and a majority in at least four of the six states.

In practice, government does not always reflect these principles. Parliament exercises limited control over the executive arm of government. Although the Constitution requires that parliament impose taxation, the executive branch prepares the all-important annual budget.[4] The executive initiates most policies, which are usually approved by parliament, at least by the Lower House. It decides when elections will be held (within a maximum three-year period) and when parliament will meet.[5] The executive is sometimes constrained by the Senate where it may not hold a majority. But, contrary to the intentions of the Constitution, the Senate is generally dominated by political parties rather than by the states. Sometimes small parties with two or three elected members representing minority views (often quite strongly) hold the balance of power in the Senate and require special concessions in return for agreeing to pass bills.

Under the constitutional process, the states granted powers to the Commonwealth but retained residual (ungranted) powers. For example, the states initially controlled income taxes. They also controlled economic activities and labour markets that took place entirely within their area. As nationhood became more important, particularly in wartime, the states ceded important powers to the Commonwealth, notably the income tax in 1942 following a national referendum. In 1946, another referendum greatly extended the Commonwealth's power to redistribute monies to individuals and so to create a welfare state. Commonwealth powers over the economy also increased as trade and labour markets crossed state borders. Although the Commonwealth government is increasingly powerful because of its ownership of most tax revenues, the federal nature of the country introduces political competition between the Commonwealth and the states which provides some constraint on the power of the Commonwealth.

The Constitution devotes little space to economic issues. But what it does say about them has important consequences. The main sections dealing with economic matters are Chapter 1 (Section 51), and Chapter 4. Section 51 describes the economic powers of the Commonwealth government. Chapter 4 describes the main principles governing finance and trade in the federation.

Some main economic points of the Constitution inclusive of amendments are:

- Monetary policy. The Commonwealth government has exclusive power to issue coins and currency in Australia.
- Taxation. In effect, following the 1942 referendum, the Commonwealth controls income taxes because the Commonwealth can withhold financial assistance to the states if they levy income tax. Also, under Section 90, only the Commonwealth can raise customs and

[4] The Senate has limited powers to amend money bills.

[5] Formally, the Governor General makes these decisions, but he (or she) normally acts on the advice of the Prime Minister.

excise taxes. In 1997 the High Court interpreted 'excise' widely to mean any tax imposed up to and including point of sale.[6]

- Borrowing. Both the Commonwealth and the states have the right to borrow to finance budget deficits.
- External trade. There would be a single external tariff for any good.
- Trade between states shall be free and unrestricted.
- A uniform law for commerce across Australia. The Commonwealth has the power to make laws for trading and financial corporations. This has paved the way for most trade practices legislation and substantial workplace legislation.
- Labour markets. The Commonwealth can conciliate and arbitrate for the prevention and settlement of industrial disputes extending beyond the limits of any one state.
- Redistribution of public monies to the states. Following some transition arrangements, the Commonwealth would determine the distribution of its revenues to the states.
- Redistribution of public monies to individuals. The Constitution initially gave the Commonwealth very limited power over social security for invalid and old age pensions. By a referendum in 1946, the Commonwealth gained greatly increased social security powers, including provisions for unemployment, sickness, families and students.

Measures of Government Size

The size of government (GS) is measured most often by the ratio of general government expenditure to gross domestic product (GDP).

$$GS_1 = \text{Total general government expenditure/GDP} \qquad (2.1)$$

Total general government expenditure is the sum of current and capital government expenditure on goods and services and transfer payments by all levels of government (Commonwealth, state and local government) that are financed through the annual budgets (see Table 2.1). Current expenditure includes the cost of publicly produced goods (mainly wages) and the purchase of goods from the private sector (e.g. medical supplies) that are used for current consumption. It also includes contributions to statutory authorities and subsidies to some public trading enterprises. Capital expenditure is expenditure on physical capital formation, such as hospitals and school buildings. Of course, intra-governmental flows must be netted out. Transfer payments are cash transfers from taxpayers to recipients. They include personal income benefits such as unemployment benefits and pensions, personal subsidies in kind for example for housing, business subsidies and interest payments.

Expenditure is related to GDP because GDP represents the productive capacity of the economy.[7] In 2016–17, general government expenditure in Australia totalled 34.6 per cent of GDP. GS_1 is a popular measure of the size of government because it can be readily estimated, includes all goods and services and transfers financed by government, and largely determines taxation requirements. It is also convenient for international comparisons.

Another expenditure measure of government size is:

$$GS_2 = \text{General government expenditure on goods and services/GDP} \qquad (2.2)$$

GS_2 excludes transfer payments. This measure is consistent with the concept of GDP, which also excludes transfer payments. In 2016–17, GS_2 was 22.0 per cent of GDP. Expenditure on goods and services is sometimes described as exhaustive expenditure because it exhausts the purchasing power of the money spent. In contrast, transfer payments redistribute resources; they do not use them.

---

[6] Ha v New South Wales (1997)

[7] For this purpose, GDP is usually measured at factor cost, which excludes indirect taxes and subsidies.

Table 2.1 General government expenditure in Australia in 2016–17

General government expenditures	*$billion*	*% of GDP*
Expenditure on goods and services		
Current expenditure	325.0	18.5
Capital expenditure	60.8	3.5
Total	385.8	22.0
Transfer payments		
Personal benefits	130.3	7.4
Interest payments[a]	26.7	1.5
Business subsidies	23.0	1.3
Other transfers[b]	42.0	2.4
Total	222.2	12.7
Total general government expenditure	607.9	34.6
Gross domestic product	1754.7	100.0

(a) Including interest on unfunded superannuation liabilities

(b) For example subsidies for expenses such private housing rents, private health insurance, child care etc.

Source: ABS, 2017, *Australian System of National Accounts 2016—17*, Tables 2, 30 and 32, Cat. No. 5204.0.

Thus, GS_2 shows how government influences directly the use of resources. However, this measure omits slightly over a third of government expenditure and does not show government's financing requirements.

General government expenditure does not include expenditure by public trading enterprises (PTEs). Inclusion would involve some double counting because PTEs sell goods to government as well as to private firms. To avoid double counting, national income accounts identify PTE capital expenditure plus dividend and interest payments as a separate item.[8] Accordingly, for a more comprehensive measure of the size of the public sector than GS_1 or GS_2, we can adopt a third measure:

$$GS_3 = (\text{Total general government expenditure} + \text{PTE capital outlays})/\text{GDP} \qquad (2.3)$$

The GS_3 measure is not often used. Although PTEs are part of the public sector they generally act commercially and are similar in some ways to private firms. Problems of definition arise when government is part owner of an enterprise. Also, consistent international comparisons are hard to obtain.

A fourth measure of government size (GS_4) focuses on the role of government as a producer of goods. A common measure is the ratio of government employment to total employment.

$$GS_4 = \text{Total government employment/total employment} \qquad (2.4)$$

Because government purchases various goods and services from the private secor, GS_4 is generally lower than GS_2. In Australia, government at all levels employs about 16 per cent of the total workforce, including employees in welfare administration. Unlike expenditure measures of government size, employment measures do not show the extent to which economic resources are subject to government control.

[8] In *the Government Finance Statistics* (ABS), PTE outlays include fixed capital expenditure outlays, interest and other property income payments, capital grants and advances and net expenditure on stocks, land and intangible assets.

Finally, government size is sometimes measured by the ratio of tax revenue to GDP.

$$GS_5 = \text{Total tax revenue/GDP} \quad (2.5)$$

GS_5 shows the size of the tax burden. However, tax revenue is generally less than general government expenditure, with the gap met by other revenue measures, such as user charges, or by borrowing. Indeed, the gap between taxation and expenditure may be several percentage points of GDP. Therefore, tax revenue gives only a partial picture of government activity.

Government Regulation

Government spending is only a partial measure of government control over an economy. Even a low-spending government can regulate almost any area of private business activity. In Australia the government has long exercised significant control over the economy (see Box 2.2). *The Economist* (19 March 2011) noted that while President Bush was in office in the United States from 2000 to 2008 over 1000 pages of federal regulations were added each year and that a quarter of a million Americans have jobs devising and implementing federal rules.[9] Japan is another country where public expenditure is low and regulation is high.

Box 2.2 Government regulation in Australia

Government has played a major role in the development of Australia. In the 19th century the public sector was mainly instrumental in developing energy, transport and communications across the continent.

Following federation in 1901, government regulations increased. Government adopted a high-tariff protectionist stance along with an industrial welfare policy designed to ensure payment of adequate wages. In 1907, in the Harvester judgement, Justice Higgins laid down that the basic wage should be sufficient to provide for the basic needs of a family of five, estimated at 42 shillings a week. Australia was one of the few countries in the world with compulsory arbitration of wages and key work conditions. The government also introduced various welfare policies, including age pensions, workers' compensation for injury and some unemployment relief. Australians generally looked on government to provide social justice and 'to soften' market forces in the form of international prices or to 'elude them or master them'.[10] After the Second World War, the role of government increased further. The 1945 White Paper committed the Commonwealth government to guarantee full employment and to intervene as necessary to achieve that guarantee.

In the 1950s, government introduced import controls to respond to balance of payments difficulties and to maintain employment. It also provided subsidies to many import-competing manufacturing businesses. The 1972–75 Labor government further expanded the role of government by greatly increasing public expenditure, especially in health and education, control of foreign investment and protection of the environment. Despite advocating smaller government, the Liberal government (1975–83) did little to reduce public expenditure or regulations. In 1984 the Brookings Institute concluded its review of the Australian economy with the observation that 'Australia's public policies are greatly influenced by the national mistrust of market outcomes' (Caves and Krause, 1984, p. 400).

Since the mid-1980s there has been substantial liberalisation of the economy. Capital markets were liberalised by the floating of the dollar, the lifting of controls over foreign exchange transactions and deregulation of the domestic banking sector. Effective tariffs have been reduced from 35 per cent in the 1970s to 5 per cent today. Many public enterprises have been privatised and many public services contracted out. Telecommunications and airlines have been largely deregulated.

Today both major political parties are committed to promoting market competition. However, as we see in Chapter 14, maintenance of competitive markets requires regulatory policies. And Chapter 29 (Box 29.3) describes ongoing regulation in labour markets.

[9] Under President Trump, many regulations are being undone. At this point, the verdict on this is an open one.
[10] Hancock (1961, p. 67). This classic study, *Australia*, was first published in 1930.

Regulations are diverse and numerous. They include price and quantity controls, product standards, regulations of medicines and foods, occupational health and safety regulations in labour markets, environmental and urban planning controls, as well as public safety measures such as closed-circuit cameras related to the wars on terror and drugs.

There is no simple measure of the extent of government regulation. It is hard even to count the number of regulations without defining what counts and what does not. The World Bank (1997) suggested three indicators of the amount of government intervention in an economy: the openness of the economy (the share of trade in GDP), the overvaluation of the currency (based on the black-market exchange rate) and the gap between local and international prices. Other authorities, such as the influential journal *The Economist*, advocate that the estimated cost of government regulations should be published in an annual regulatory budget. This would be a very complex exercise especially if benefits are included. Although economists can estimate the price and output impacts of individual regulations and their welfare costs and benefits, it is difficult to estimate the total gross or cost of regulation in any country.

Public expenditure substitutes

Often government requires the private sector to make certain expenditures that substitute for public expenditure. These expenditures are known as **coerced private expenditures**. For example, governments in Australia require private firms to contribute to employee retirement schemes, provide redundancy payments, pay workers compensation for injury and invest in pollution control. An extreme example of coercion is military conscription where government pays conscripted labour far less than its economic opportunity cost. Jury service is another example. When there are large coerced private expenditures, public expenditure becomes a poor measure of the real influence of government on the economy.

Tax expenditures
A term for tax concessions

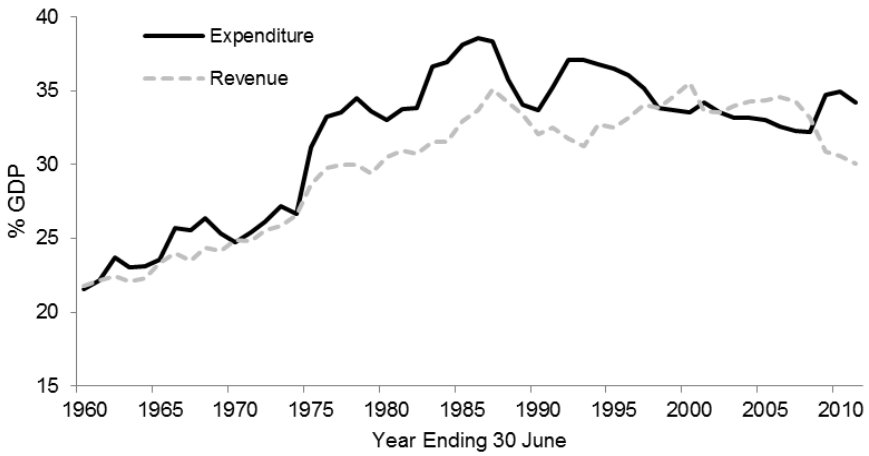
Tax expenditures are another major form of expenditure substitute. Tax expenditures are tax concessions. These concessions create revenue losses that have a similar effect on the budget as government expenditures. Suppose that some farmers spend $100 million on plant and equipment and that their marginal tax rate is 30 per cent. Government could assist the farmers by (1) giving them a grant of, say, $30 million or (2) allowing them to deduct their capital expenditure against current income, reducing their combined taxable income by $100 million and their tax payments by $30 million. Either way, the farmers would gain $30 million, and the cost to government (i.e. to other taxpayers) is $30 million. Alternatively, instead of providing allowances to families with dependent children, government could allow an equivalent amount in tax deductions. There are many forms of tax concessions, including tax exemptions or deferrals, expenditure deductions, accelerated depreciation and so on. Any such concession represents a call on the budget similar to direct outlays.

To assess the size of a tax concession, it is necessary first to establish what constitutes 'normal tax'. This can be difficult. In a detailed review. the Australian Treasury (2017) estimated that large Commonwealth tax expenditures totalled $149 billion in 2016-17, equal to a third of Commonwealth outlays or some 8.5 per cent of GDP. This included an estimated $61.5 billion for exempting homes from capital gains tax (CGT) and discounting CGT for investment housing properties, and $32.9 billion on concessions to superannuation contributions and earnings.

Government Expenditure and Revenue

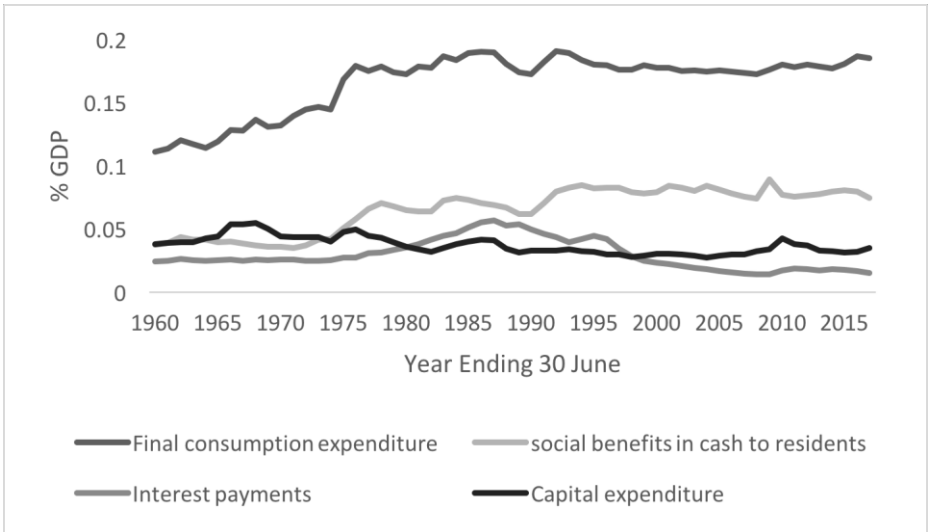
As shown in Figure 2.1 overleaf, government spending by all levels of government rose from less than 25 per cent of GDP in the mid-1960s to about 37 per cent of GDP in the late 1980s. Between then and 2007, general government expenditure oscillated but tended down to about 33 per cent of GDP. Following the global financial crisis, it moved sharply upwards to 35 per cent of GDP in 2009–10 and it has remained about that level in each year since then.

Figure 2.2 shows the main components of general government expenditure: consumption, income transfers, capital expenditure and interest payments. Consumption expenditure, the largest component, rose from 12 per cent of GDP in the mid-1960s to 20 per cent of GDP in the mid-1980s, but fell back to about 18 per cent in recent years. Personal income benefits, the second largest component, have increased from 4 per cent of GDP in the mid-1960s to about 8 per cent of GDP today. This does not include **in-kind** transfers tied to housing, health, child care and so on (see Table 2.1). Since the early 1980s, government fixed capital expenditure has generally been between 2 and 3 per cent of GDP. On the other hand, interest payments have fallen from a high of 5 per cent of GDP in the late 1980s to less than 2 per cent today due to government policies due mainly to lower interest rates and partly to a reluctance to borrow.



Source: ABS, *Time Series Spreadsheets*, Cat. No. 5204.0

Figure 2.1 General government expenditure and revenue (as a percentage of GDP)



Source: ABS, *Time Series Spreadsheets*, Cat. No. 5204.0

Figure 2.2 General government expenditure components (as a percentage of GDP)

Traditionally, public trading enterprises supplied most infrastructure services in Australia, including power, water, transport and communications services. Also, the states owned many banking and insurance businesses. In the early 1980s, PTE gross capital expenditure was over 6 per cent of GDP. Following the sale of many PTEs to the private sector, PTE capital outlays have fallen to-day to only about 1.5 per cent of GDP. Thus, public sector expenditure overall in 2016-17, including PTE capital expenditure, was 36.1 per cent of GDP.

Table 2.2 shows general government expenditure by function in 2015–16. The largest government programs are social security and welfare, health and education, which account for 28.1 per cent, 19.4 per cent and 14.6 per cent of total expenditure respectively. Within social security and welfare, there are several large programs including assistance to families, the unemployed, people with disabilities and the elderly (see Chapter 23). Other major expenditure areas are transport and communications, defence, public order and safety, and housing and community amenities, as well as general public services.

Table 2.2 General government expenditure by function 2015-16 ($billion)

Expenditure category	*Commonwealth*[a]	*State and local government*[b]	*Total*	*% of total expenditure*
Social security and welfare	152.0	20.1	168.8	28.1
Health	69.3	68.0	116.4	19.4
Education	32.3	58.4	87.9	14.6
General public services	22.7	13.1	34.8	5.8
Public debt transactions	16.0	8.6	38.2	6.4
Transport and communications	7.2	29.7	29.2	4.9
Other economic affairs	9.2	5.3	14.2	2.4
Defence	26.2	0	26.2	4.4
Public order and safety	4.8	25.1	29.3	4.9
Housing and community amenities	7.3	18.9	22.9	3.8
Recreation and culture	3.4	10.4	13.6	2.3
Other	80.5	13.8	19.4	3.2
Total	430.9	271.6	600.8	100.0

(a) Excludes GST payments to the states.

(b) The sum of Commonwealth and state and local government expenditures exceeds total government expenditure due to transfers between jurisdictions.

(c) Total slightly different from Table 2.1 with data from later source.

Source: ABS, *Government Finance Statistics*, Cat. No. 5512.0, 2015–16.

Table 2.3 Sources of general government revenue in 2015-16

Sources of revenue	*Total government ($billion)*	*% of total revenue*
Taxation revenue	464.7	80.5
Sales of goods and services	55.0	9.5
Interest income	8.5	1.5
Dividend income	11.7	2.0
Other	37.6	6.5
Total	577.7	100.0

Source: ABS, *Government Finance Statistics*, Cat. No. 5512.0, 2015–16.

Government revenue

Government raises most of its revenue via taxation, but a significant and increasing proportion by other means. As shown in Table 2.3 above, in 2015–16 government obtained 80.5 per cent of its revenue through taxation. It obtained just under 10 per cent in sales of goods and services, nearly 4 per cent in interest and dividend income, and 6.5 per cent by a variety of through other charges (fees, fines, etc.).

The major sources of taxation revenue are detailed in Table 2.4. In summary, the main sources of tax revenue were:

- Income taxes, including company tax 57.0 per cent
- Employers' payroll and labour taxes 5.0 per cent
- Taxes on goods and services (including motor vehicles) 27.4 per cent
- Taxes on wealth 10.6 per cent

Table 2.4 General government taxation revenue in Australia in 2015–2016 ($billion)

Sources of taxation	*Commonwealth*	*State and local*	*Total government*	*% of total revenue*
Taxes on income				
Individuals	192.1	–	192.1	41.3
Enterprises	71.2	–	71.2	15.3
Non-resident	1.8	–	1.8	0.4
Total	265.1	–	265.1	57.0
Employers' payroll and labour taxes	0.7	22.7	23.3	5.0
Taxes on provision of goods and services				
General taxes (sales taxes)	1.5	–	1.5	0.3
Goods and services tax	60.3	–	60.3	13.0
Excise tax	22.4	0.1	22.5	4.8
Tax on international trade	14.1	–	14.1	3.0
Taxes on gambling	–	6.1	6.1	1.3
Taxes on insurance	–	5.7	5.7	1.2
Total	98.3	11.9	110.2	23.7
Taxes on use of goods and performance activities				
Motor vehicle taxes		9.9	9.9	2.1
Other	5.8	1.5	7.3	1.6
Total	5.8	11.4	17.3	3.7
Taxes on wealth				
Land taxes		7.2	7.2	1.6
Municipal taxes	–	16.9	16.9	3.6
Taxes on capital and financial transactions	–	22.5	22.5	4.8
Other	–	2.9	2.9	0.6
Total	–	49.6	49.6	10.6
Total taxation revenue	369.9	95.5	465.5	100
Percentage of total (%)	79.5	20.5		

Sources: ABS, *Public Sector Accounts, Taxation Revenue*, Australia, Cat. No. 5506.0, 2015-16.

A major feature of the tax revenue is the high proportion levied by the Commonwealth. The Commonwealth collects just under 80 per cent of all taxes, but its own purpose spending accounts for little over half of all government expenditure. The state and local government deficit is funded mainly by revenue from the goods and services tax (GST), which the Commonwealth passes on to the states, and by other grants from the Commonwealth. The imbalance of expenditure and revenue at each level of government is known as **vertical fiscal imbalance** (for further discussion see Chapter 33).

Government Budgets: Deficits, Debt and Assets

As we saw in Figure 2.1 total government expenditure usually exceeds current revenue. This creates a budget deficit. That deficits often occur is not surprising. Expenditure includes current and capital expenditure. Naturally governments may borrow to fund some or all capital expenditure, with repayments made as the capital assets provide services, rather than pay for capital assets out of current revenue. This means creating debt as well as assets. To understand what is happening, the key financial terms (expenditure, revenue, deficit, debt and assets) need to be defined.

Expenditure and revenue are flow concepts that apply to a given period. However, there is an important distinction between cash and accrual accounts. **Cash accounts** are based on payments made or received in an accounting period. **Accrual accounts** are based on expenses incurred and revenue earned in that period regardless of whether payment has been made or received. Most private companies use accrual accounts. Australian governments changed from cash to accrual budgeting in the financial year 1999–2000, as did the UK government.

Accrual accounting
Expenses and income are recorded when they are incurred not when they are paid

Five key accrual terms. We now define five key terms in an accrual framework, including two definitions of budget outcomes.

Current expenditures are expenses incurred to provide services and transfers in a designated accounting period such as a financial year. Employee expenses include entitlements such as superannuation and long-service leave that are accrued as well as wages and salaries paid out in the year. Other operating expenses include depreciation of all assets, *including previously purchased assets*, but not expenditure on purchase of new assets. Transfer payments include personal benefit payments and interest payments.

Current revenue is the revenue due in an accounting period from taxes, sales of goods and services, interest and dividends, and fees and fines. It also includes operating grants from another level of government.

The **net operating balance** equals current operating revenue minus current expenses. It is broadly similar to a company's net income (i.e. its profit or loss).

Net operating balance
Current operating revenue minus current expenses

Net capital investment is capital expenditure on new physical assets and inventories less depreciation. In an accrual budget, net capital investment does not include the purchase or sale of financial assets.

The **overall budget balance** equals the net operating balance minus net capital investment This is the budget that is commonly referred to as being in deficit or surplus. The balance is also described sometimes as the net borrowing or lending balance. This measure of budget balance is not strictly an accrual concept because it includes all capital expenditure and does not include depreciation.[11]

Overall budget balance
Net operating balance minus net capital expenditure

Two key cash terms. In addition, there are two important cash-based measures of the overall budget balance. The **underlying cash balance** is the budget balance based on cash payments

[11] Some experts (e.g. Gruber, 2016, Chapter 4) contend that it is too difficult, and not practical, to distinguish between consumption and capital expenditure. Our view is that this is a critically important expenditure distinction with major policy implications (see Chapter 29 below).

and receipts rather than on accrued payments and receipts. It includes expenditure on new assets such as schools, but not the purchase or sale of financial or existing physical assets. It ignores accrued financial liabilities, such as pensions, and depreciation of plant and equipment.

The **headline cash balance** equals the underlying cash balance plus sales and purchases of financial and physical assets. This allows for the sale of physical assets to fund and apparently avoid a budget deficit. However, from the perspective of the government's balance sheet, it is immaterial whether government sells a physical asset worth $x million or takes on a financial liability of $x million.

Summary. The main budgetary concepts along with the Commonwealth government's (then) estimated budgets for 2016-17 and 2017–18 are shown in Table 2.5. These figures exclude GST revenue, which the Commonwealth counts as a state tax (though this is contrary to law and the view of the Australian Bureau of Statistics). In these years all four estimated measures of budget balance were in deficit. This is not always the case. In some years one or more of the four measures of budget balance may be in surplus and the others in deficit. This provides politicians with a perfect opportunity to confuse the public!

Accrual measures generally provide a more informative picture of the budget than do cash budgets because they reflect real economic activity over a period and they exclude financial transactions that have no effect on the net worth of government. Therefore, the preferred measures of budgetary outcome are the (accrual) net operating balance or budget balance. However, cash flow forecasts are required, especially for short-term management of the budget, to establish working capital needs.

The government's balance sheet: assets and liabilities (debts)

We turn now to the government's balance sheet. The main components for the whole Australian public sector are defined and shown in Box 2.3. On the one hand, government (on behalf of the state) owns various financial and non-financial assets. Financial assets include equity assets. Non-financial assets include produced and non-produced assets (land). On the other hand, borrowing to fund a budget deficit creates a financial debt (liability).[12]

Table 2.5 Summary of Commonwealth government budget estimates for 2016-17 ($billion)

Budget figure	*Basis*	*2016-17*	*2017-18*	*Comments*
Total current revenue[a]	Accrual	412.1	444.4	Includes all current revenues
Total current expenses	Accrual	450.8	464.3	Real operating expenses
Net operating balance (NOB)	Accrual	-38.7	-19.8	Current revenue – current expenses
Net capital investment (NCI)[b]	Accrual	2	0.5	Capital expenditure – depreciation
Budget (fiscal) balance	Accrual	-40.7	−20.3	NOB – NCI = net lending/borrowing balance
Underlying cash balance (UCB)	Cash	-37.6	-29.4	Budget balance on a cash basis
Headline cash balance	Cash	-51.1	-48.4	UCB + financial purchases and sales

(a) Includes expected earnings from Future Fund.

(b) Equals net acquisition of non-financial assets.

Source: Treasurer, *Budget Strategy and Outlook*, 2016–17, Budget Paper No.1, Statement 3, Table 3.

[12] Government can fund a deficit by printing money, which is known as an inflation tax (Chapter 25).

Box 2.3 Key balance sheet concepts and numbers for Australian public sector at 30 June 2016

Assets/liabilities	*Definition/components*	*$billion*
Assets	Are controlled and provide future economic benefits	
Financial assets	Cash, deposits, investments, loans, accounts receivable, equity assets	761.9
Non-financial assets	Produced assets and non-produced assets (land), other non-financial assets	1716.7
Total assets	Financial plus non-financial assets	2478.7
Liabilities	Amounts owed to other parties	
Financial liabilities	Deposits held, advances received, borrowing, accounts payable	908.9
Other liabilities	Unfunded employee entitlements, other provisions	720.2
Total liabilities	Financial plus other liabilities	1703.4
Net measures		
Net public debt	Sum of deposits held, borrowing and advances received less sum of cash, financial assets (excluding equity) and advances paid	380.4
Net financial liability	Total liabilities minus financial assets	-941.4
Net public worth	Total assets less total liabilities	775.3

Source: ABS, *Government Finance Statistics*, Cat No 5512, 2015-16.

Table 2.6 Public net worth and net debt as percentage of GDP

Measure	*2007-08*	*2008-09*	*2009-10*	*2010-11*	*2011-12*	*2012-13*	*2013-14*	*2014-15*	*2015-16*
Net public worth	75.6	74.7	72.6	64.6	48.9	55.3	52.8	52.2	46.7
Net public debt	-7.3	-3.5	2.3	9.5	15.2	16.1	18.1	20.3	19.1

Source: ABS, *Government Finance Statistics*, Cat No 5512, 2015–16.

Total financial debt is the stock of liabilities that accrue from running deficit budgets. These liabilities include deposits received, advances paid, accounts receivable, borrowings, unfunded employee entitlements and various provisions (insurance technical reserves).

Net public debt
Financial liabilities minus financial assets (excluding equity)

There are three important summary measures. One which gets a lot of attention is the **net public debt**: this is the sum of deposits held, borrowing and advances received less the sum of cash, financial assets (excluding equity) and advances paid. The net public debt, combined with the interest rates attached to the liabilities and assets, determines the net interest payment to be funded from current revenue.

However, net public debt is only a partial picture of government liabilities and assets. Net financial liability provides a fuller picture. **Net financial liability** equals total government liabilities less financial assets including equity assets. Equivalently, financial assets less total liabilities are called **net financial worth**.

Net public worth
Total assets less total liabilities

The third and most comprehensive measure of government's financial position is **net public worth**, which equals total assets less total liabilities. This is the most important measure. If government has high net public worth, it should be able to reconfigure the components of the balance sheet to provide safe levels of net public debt and net financial liability.

Table 2.6 shows recent trends in the net debt and net public worth of Australian governments. Over the decade before the global financial crisis (GFC) in 2007-08, net debt fell substantially as Commonwealth and state governments sold PTEs and often adopted overall budget surpluses. Thus, at end 2008-09, general government net debt was negative (relevant assets exceeded relevant liabilities). Following the GFC, net public debt rose to around 20% of GDP. By international standards these are still very low amounts of public

debt. On the other hand, with the rise in debt, the net public worth fell from around 75% of GDP before the GFC to around 50% of GDP at end 2016-17.

These totals conceal major differences between different levels of government. The states collectively have high net public wealth (in land, fixed assets and PTEs) and very low debt. On the other hand, at end 2015-16, the Commonwealth had a negative net worth of $403 billion, including a net debt of $300 billion due to recent budget deficits as well as large unfunded superannuation liabilities. Thus, the Commonwealth net debt was over 90% of all government debt and equivalent to 18.1% of GDP. Some policy implications are discussed in Chapter 29.

International Comparisons

Table 2.7 shows general government expenditure (GGE) as a percentage of GDP in 13 OECD economies from 1960 to 2014. Excluding Australia, GGE rose from an average of 28.4 per cent of GDP in 1960 to 44.2 per cent in 1980, reflecting large increases in most countries over this period. GGE then remained quite stable in relation to GDP, albeit with ups and down in various economies, over the next 25 years to 2005. However, government deficits to bailout major financial and other institutions and to avert a global financial and economic meltdown led to substantial increases in GGE. Average GGE as a percentage of GDP rose to 47.3 per cent in 2010. It fell slightly to 46.3% in 2014 mainly due to declines in the UK and United States.

Since 1980, there has been minor convergence among OECD countries. The GGE / GDP ratio has risen substantially in some countries, notably Japan, Spain and the United States. On the other hand, high GGE countries like, the Netherlands and Sweden have reduced the GGE GDP ratio. However, there remain large differences between these and other OECD counties.

A major consequence of the recent rise in public expenditure and the associated deficits has been the rise in public debt. In countries such as Japan, Greece and Italy, net public debt rose to around 100 per cent of GDP in 2009 and gross public debt was an even higher proportion of their GDP (see Table 29.3)

Table 2.7 General government expenditure as a percentage of GDP

Country	*1960*	*1980*	*1990*	*2000*	*2010*	*2014*
Australia	21.3	32.6	33.3	33.5	34.5	36.0
Austria	35.7	48.1	38.6	52.1	52.8	52.3
Belgium	30.3	58.6	54.8	49.1	53.3	55.2
France	34.6	46.1	49.8	51.6	56.4	57.1
Germany	32.4	47.9	45.1	45.1	47.3	44.3
Italy	30.1	42.1	53.4	46.2	49.9	50.9
Japan	17.5	32.0	31.3	37.3	39.6	40.3
Netherlands	33.7	55.8	54.1	44.2	48.2	46.2
Spain	18.8	32.2	42.0	39.1	45.6	44.9
Sweden	31.0	60.1	59.1	52.7	51.2	51.5
Switzerland	17.2	32.8	33.5	33.7	32.8	33.5
United Kingdom	32.2	43.0	39.9	36.6	47.8	43.3
United States	27.0	31.4	33.3	32.8	42.9	38.0
Average (exc. Australia)	28.4	44.2	44.6	43.4	47.3	46.3

Source: *OECD 2016.*

Table 2.8 Structure of general government expenditures (as percentages), 2015

Country	*General public services*	*Defence*	*Public order and safety*	*Economic affairs*	*Health*	*Education*	*Social security and welfare*	*Other*[a]
Australia	12.5	4.4	4.9	10.0	19.4	14.6	28.2	6.1
Denmark	13.5	2.0	1.8	6.7	15.6	12.8	43.0	4.5
France	11.0	3.1	2.9	10.0	14.3	9.6	43.1	6.0
Germany	13.5	2.3	3.6	7.1	16.3	9.6	43.1	4.6
Italy	16.6	2.4	3.7	8.1	14.1	7.9	42.6	4.6
Japan	10.4	2.3	3.2	9.5	19.4	8.7	40.7	5.6
Spain	14.9	2.2	4.6	10.0	14.2	9.3	39.1	5.7
Sweden	14.1	2.3	2.6	8.4	13.8	13.0	41.6	4.2
United Kingdom	10.6	5.0	4.7	7.1	17.8	12.0	38.4	4.5
United States	13.8	8.8	5.4	8.7	24.2	16.2	20.8	2.1
OECD	13.2	5.1	4.3	9.3	18.7	12.6	32.6	4.2

(a) Includes environment protection, housing and community amenities, and recreation, culture and religion.

(b) Based on 2015–16 expenditure.

Source: OECD Statistics, *General Government Accounts*, *Government Expenditure by Function*, 2015–16.

Table 2.8 shows the shares of the major components of general government expenditure in various OECD countries in 2015. Because classifications may vary between countries, too much should not be inferred from minor differences. However, clearly, inter-country differences in GGE result more from differences in income transfers (social security and welfare payments) than from differences in expenditure on goods and services. Australian expenditure on goods and services was broadly in line with international practice, but expenditure on social security and welfare was significantly lower than in most other countries except for the United States.

The reasons for the low level of social security payments in Australia include smaller proportions of elderly and unemployed people than in Europe, more reliance on the private sector for retirement incomes and workers' compensation, and a more means-tested approach to social assistance than is common in many other countries. These issues are discussed in more detail in Chapters 22 and 23 below.

Determinants of Government Expenditure

What drives the levels of general government expenditure that we have observed? As we have noted, the major components of government expenditure are expenditure on consumption goods and services and welfare (transfer) payments. We focus mainly on these below but also discuss briefly transfer payments other than welfare payments and capital expenditure.

But, to start, we make two preliminary observations. First, expenditure is a product of the quantity of goods or services supplied and their unit costs. Therefore, we need to consider what drives both quantities and costs. Second, quantities may be driven by normal demand factors, broadly by individual preferences, or by social preferences (views on equity and social justice). Individual preferences are easier to measure and forecast and we can draw on conventional economic studies and explanations. Social preferences can be explained and understood but are harder to measure and forecast.

On the other hand, a more political approach to explaining public expenditure emphasises the role of politicians, bureaucrats and special interest groups in influencing government expenditure. These factors are examined in Chapter 10.

Government consumption expenditure

To explain government spending on consumption goods, we need therefore to consider the quantity and unit cost of government consumption goods. Holding other public expenditure constant, government expenditure rises in relation to GDP if either (1) the quantity of publicly provided goods rises faster than the output of market goods or (2) the cost of publicly provided goods rises relative to the cost of other goods.

As discussed in Chapter 4 (see Table 4.1), governments supply a wide range of goods, including pure public goods that are necessary for the functioning of the state, a range of goods associated with market failures of some kind, including various health, education and infrastructure services, and many goods for social purposes, again often health and education services, but also housing and family support services of various kinds, to name just a few examples. Clearly the demand for many of these services depends on both individual and social preferences.

Individual preferences are primarily a function of income, demographics, external circumstances and relative prices. Demand for public goods may rise faster than GDP if the income elasticity of demand for public goods is greater than 1.0, demographic changes increase the demand for major public services like health or education or increased population density and urban congestion increase demand for environmental and urban services. As we see below, there is no clear answer whether this demand does grow faster then GDP.

In a review of the evidence from cross-sectional studies of state and local government data in the United States, Mueller (2003) found that most estimates of the income elasticity of demand for public goods are less than 1.0 and that few estimates of the income elasticity are substantially greater than 1.0.

On the other hand, several studies (e.g., OECD, 2003; Villanyi *et al*., 2010; Australian Treasury, 2010) have shown that an ageing population is likely to increase the GGE/GDP ratio. The OECD (2003) estimated that the increasing proportion of over 65s in the population will increase public expenditure on health care and long-term care by between 1.5 and 4.0 percentage points of GDP up to 2050. An ageing population also reduces the workforce participation rate which raises the GGE/GDP ratio.

Circumstances, such as increased urban densities, congestion and environmental pressures, may also raise the demand for some government services, for example public transport and waste disposal services. Some, but not all, such pressures may be dealt with by regulation rather than by increased public services.

Turning to social preferences, similar factors drive the provision of welfare services as drive the provision of cash transfers that are discussed below. Fundamental to this is society's perception and acceptance of social needs. However, there may well be social preferences to provide services, including health, education and social services of various kinds for those who cannot afford them and to do so in preference to income transfers. These issues are discussed at many points in this book.

It may be supposed as a default assumption that the costs of public production would rise approximately in a similar fashion to the costs of private production. In that case, relative quantities of production rather than relative prices would drive the contribution of public production to GDP.

In a classic article, Baumol (1967) argued that costs of public production would rise faster than costs of private production because government mostly provides labour intensive services. He argued that opportunities for productivity improvements in government are low because of the small role of technology and the relatively few opportunities for substituting capital for labour. If wages in government employment rise with market wages, as they are likely to, unit costs would rise faster in government than in private firms. Mueller (2003) cites various studies that found a relative rise in the price of government-provided goods and that this explained part of the rise in general government expenditure.

However, it is not clear that this phenomenon is due to technical constraints or that productivity must rise more slowly in the public sector. Computers and other innovations provide many opportunities for increased productivity in provision of services, which is now the dominant part of the private sector in most OECD countries including Australia. If productivity does rise by less, and unit costs by more, in the public sector than in the private, this may be due to bureaucratic inertia and constraints rather than to the inherent technology of supplying public services. Tunny (2000) found that, in so far as the cost effect exists, the causes of rising costs appear to be political rather than technical.

Personal transfer payments

Personal transfer payments depend on (1) the number of beneficiaries and (2) the levels of benefits provided. Clearly, government determines both eligibility for benefits and benefit levels. Ideally these decisions would be a result of social preferences taking account of both the ability to pay of the better-off in society and the needs of the less well-off, as well as to any disincentives arising from income transfers. Of course, in some cases, the decisions are arbitrary or self-promoting political decisions rather than an outcome of social preferences.

However, given the eligibility criteria, the number of beneficiaries depends on economic, social and demographic circumstances. Government transfer payments rise with increased numbers of age-based (young and old) dependants in society, families and single parents, unemployment and earnings inequalities. Widespread demographic and social changes have increased the proportions of elderly persons and single-parent households in many societies. Market forces along with globalisation have greatly increased economic inequality within countries (see Chapters 20 and 34). In most OECD economies, transfer payments have risen in proportion to GDP because governments have extended eligibility for personal benefits, for example to single parents and for maternity leave, and because they have increased personal benefit levels. However, the proportions of unemployed and elderly persons have also risen, reflecting economic and social factors.

In Australia, in the 1960s under 1 million people (about 10 per cent of the population) received benefits from the Commonwealth government. Today, out of a population of 24.3 million, some 2.5 million receive public pensions, 2 million families receive family support payments, over 0.7 million receive disability allowances and another 0.7 million receive unemployment benefits. Ignoring other smaller categories of recipients and counting families as one individual, there are some 6 million people in these major benefit categories or about 24 per cent of the population. This reflects social or economic changes as well as social preferences and political determinations.

Other transfer payments

As we observed above, other transfer payments have fallen significantly relative to GDP in Australia in recent years due principally to declines in government borrowing and lower nominal interest rates. Government subsidies to business have also fallen. However, government borrowing and interest rates may of course rise.

Government capital expenditure

Early theories suggested that government capital expenditure would rise as a proportion of GDP with increased demand for the transport, communications and energy infrastructure necessary for economic development.[13] In so far as these are public goods (associated with some form of market failure) the government would have a major role in the supply of these goods. In developed economies it is sometimes argued that public investment is required to cope with high-density urban systems and environmental degradation. While there is casual

[13] This is sometimes called Wagner's law of expanding state activity (named after the 19th century economist).

observation that this investment is required and is being undertaken in Australia's largest cities (Sydney and Melbourne), we are not aware of empirical studies of the interaction between government capital expenditure and urbanisation. Indeed, in Australia it is often argued that low population density creates higher demands on public expenditure.

Summary

Economic factors explain some changes in government consumption expenditure and transfer payments. The demand for public consumption goods rises with income, but it does not appear to be income elastic. Demand also changes with changes in demographic structure. Government expenditure may also rise relatively fast because of low increases in productivity in the public sector, but this may reflect bureaucratic and political inefficiency rather than technical factors of production.

Economic and demographic factors can also explain some increases in transfer payments associated with increased dependants and income inequality. Moreover, some commentators have observed that technical changes, such as increased cash transactions in developing countries and electronic innovations in developed economies, have facilitated the means by which governments can collect tax and fund desired expenditures.

In a study of the determinants of the growth in public expenditure, Borcherding (1985) estimated that economic factors (income and cost changes) explained 38 per cent of the growth in US public budgets. Most of the balance would be explained by an increase in transfer payments. This is, of course, a very dated study and based on the US economy. But it seems likely to-day that public expenditure, especially the large component involved in transfer payments, reflects to a large extent social preferences or political factors.

Summary

- The *Commonwealth of Australia Constitution Act 1900* (with amendments) establishes the main powers of the Commonwealth and state governments, including their economic powers.
- The most common measure of government size is the ratio of general government expenditure to GDP.
- Government also regulates the private sector in many ways, including by coerced private expenditures which may substitute for public expenditure.
- General government expenditure in Australia is about 35 per cent of GDP. The largest components are government consumption expenditure and personal benefit transfers.
- Australian governments raise nearly 80 per cent of current revenue via taxation, with most of the rest raised by sale of goods and services, receipts of interest and dividend payments, fees and fines.
- Commonwealth taxes account for over 80 per cent of all taxes collected but the Commonwealth accounts for only half of own purpose expenditures.
- The overall budget balance equals current revenue less operating expenses and capital expenditure.
- The net operating balance equals current revenue less current expenses.
- The two key balance sheet concepts are net public debt and net public worth. Net public debt is broadly financial liabilities less financial assets (excluding equity holdings). Net public worth is total assets less total liabilities.
- Compared with other developed economies, Australian government expenditure is low. This reflects mainly a lower level of income transfers due to means testing of benefits. It also reflects a relatively high level of mandated private expenditure for welfare purposes.
- Government expenditure on goods and services can be explained partly by market failures and by socioeconomic factors such rising income, the relative unit costs of government production and demographic factors. Socio-economic factors, such as demographic changes, also explain some of the growth in transfer payments.
- However, social preferences relating to household needs and political decisions are the main drivers of transfer payments and also impact on government capital and consumption expenditures.

Questions

1. If government size is measured by the ratio of general government expenditure to GDP, what happens to the size of government if:
   i. government increases welfare payments to unemployed persons, when such transfer payments are not recorded as part of GDP?
   ii. unemployment rises?
   iii. interest rates fall?
   iv. government increases income tax rates?
   v. government requires all households with more than a certain level of income to hold private health insurance?
   vi. government subsidises private health insurance?
   vii. the terms of trade improve?
2. Explain why tax expenditures can have the same real effects as actual government expenditures. Give an example.
3. Explain the difference between cash and accrual accounting.
4. What is the distinction between current and capital expenditure? What are the arguments for and against treating capital expenditure differently in the government's accounts?
5. The headline cash budget balance is a common measure of budget outcomes. What are the main weaknesses of this measure?
6. Are student university fees included in Commonwealth government revenue? And are HECS repayment obligations included in the net worth of the Commonwealth?
7. The Australian government has established a Future Fund with revenues from past budget surpluses and asset sales quarantined in the Fund to pay for unfunded superannuation liabilities. What are the main arguments for and against such a fund?
8. How can the income elasticity of demand for public goods be estimated from cross-sectional data for different jurisdictions, such as US states and local governments? What difficulties might arise in estimating this income elasticity?
9. Does Baumol's theory of low public sector productivity growth partly explain the increase in government expenditure as a percentage of GDP?
10. Personal benefit payments as a percentage of GDP have doubled since the 1960s. What factors explain the increase in personal benefit payments? Is this trend likely to continue in the future?
11. What are
    i. the major government expenditure benefits in kind to Australian households?
    ii. the major business subsidies provided by the Australian government?
12. In presenting the 2011–12 budget, the Australian government announced provision of $3.1 billion to the National Broadband Network Company which it owns, but it did not include this amount in the estimated expenditure or in the budget deficit of $22.6 billion. How could the government justify excluding this payment from the budget? Is this exclusion appropriate?

Further Reading

Australian Bureau of Statistics (2005) *Australian System of Government Finance Statistics: Concepts, Sources and Methods*, ABS, Cat. No. 5514.0, Canberra.

Australian Bureau of Statistics (2017) *Government Finance Statistics: Australia ABS*, Cat. No. 5512.0, Canberra.

Borcherding, T.E. (1985) 'The causes of government expenditure growth: a survey of the U.S. evidence', *Journal of Public Economics*, 28, 359–382.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapter 4, Worth Publishers, New York.

Mueller, D.C. (2003) *Public Choice III*, Chapter 21, Cambridge University Press, Cambridge.

Robinson, M. (2002) 'Accrual accounting and Australian fiscal policy', *Fiscal Studies*, 23, 287–300.

Rosen H. and T. Gayer (2014), *Public Finance*, 10th ed., Chapter 1, McGraw-Hill, New York.

Treasurer (2010) *Australia to 2050*: *Future Challenges*, Treasury, Canberra.

Treasurer (2011) *Budget Strategy and Outlook 2011–12* (Australian Government budget <http://www.budget.gov.au>).

Part 2

Markets and Government

Chapter

3

Competitive Markets: Efficiency and Welfare

Although there is a sense in which property ought to be common, it should in general be private. When everyone has his own separate sphere of interest, there will not be the same ground for quarrels; and they will make more effort, because each man will feel that he is applying himself to what is his own.

Aristotle, *The Politics*

Economic Efficiency ♦ Efficiency in a Single Market ♦ Conditions of Efficiency in all Markets ♦ Efficiency, Equity and Social Welfare ♦ Competitive Markets and Efficiency ♦ Competitive Markets and Equity ♦ Conclusions

In this chapter we discuss the benefits of competitive markets. The case for competitive markets is based on three main propositions. First, market trades are freely chosen actions from which all trading parties expect to benefit. By contrast, government regulations of trades restrict the choices that individuals can make and the benefits from free choices. Second, competitive markets allocate resources efficiently. Price signals reflect scarcities. No central planner is required to calculate scarcities. In response to these prices, it is argued that firms in competitive markets satisfy the wants of consumers better than any alternative system would do. Third, markets promote economic growth by rewarding innovation and risk taking. In this chapter, we examine the second proposition. The other arguments are discussed elsewhere. Specifically, Chapter 5 discusses how markets promote economic growth.

The idea that competitive markets can produce an efficient and equitable allocation of resources is formalised in the First and Second Theorems of Welfare Economics respectively. The theorems are quite technical but understanding them provides important insights into the meaning of economic efficiency and the relationship between efficiency and welfare.

We start the chapter by introducing the concept of economic efficiency. The following sections describe the economic conditions for production and consumption that provide an efficient allocation of resources in a single market and in all markets simultaneously. These conditions are achieved in a system of perfectly competitive markets (the First Welfare Theorem). However, efficient allocations of resources are often inequitable. The Second Welfare Theorem shows that if government can redistribute resource endowments in an equitable way without distorting economic behaviour, a perfectly competitive economy will produce an outcome that is both efficient and fair. However, it turns out that resources cannot be redistributed without distorting behaviour and markets are rarely perfectly competitive. There is therefore an ongoing tension between the model of a perfectly competitive economy and what an actual economy can achieve.

Economic Efficiency

An economy is efficient if it provides the maximum amount of goods that individuals want from the resources available. In an efficient economy all potential gains are exploited. Efficiency maximises the welfare of individuals in the community given their productive endowments and the resources and technology available.

Overall economic efficiency requires three specific kinds of efficiency: production, consumption and product mix efficiency. To describe these terms, we draw on the concept of the **production possibilities frontier** (PPF), which is shown in Figure 3.1. A PPF shows the maximum quantity of goods, in this case food and clothing, which can be produced in any period, given resources and technology. Both goods are represented in equivalent physical units, such as a loaf of bread or a shirt. Once an economy is on the production possibility frontier, food output can increase only if clothing output falls, and vice versa. The PPF also shows how much of each good is given up for a unit increase in the other one. The schedule is usually drawn concave to the origin because the marginal output of a good declines as more resources are applied to its production.

Production possibilities frontier
Shows the maximum quantity of goods that can be produced from given resources and technology

Productive (or technical) efficiency means producing the maximum output of goods from given resources. It means producing each good in the most efficient way (with minimum use of resources). Producing at any point on the PPF is a necessary and sufficient condition for productive efficiency. An allocation of resources is technically efficient if it is impossible to increase the output of one good without decreasing the output of another good. If an economy produces at a point within the PPF envelope, such as at *F*, some resources are employed inefficiently or are unemployed.

Consumption (or exchange) efficiency means that goods are allocated to the individuals who want them. For any given output, at say point *H* in Figure 3.1, consumers will receive the bundle of food and clothing that maximises their satisfaction (utility), given their income and preferences. The consumption (exchange) of goods is efficient if it is impossible to increase the utility of one person without reducing the utility of another person.

Product mix (or overall) efficiency means that firms produce the goods that people want given available production technologies. An economy could produce point *H* output and be technically and exchange efficient but not produce the mix of goods that individuals most

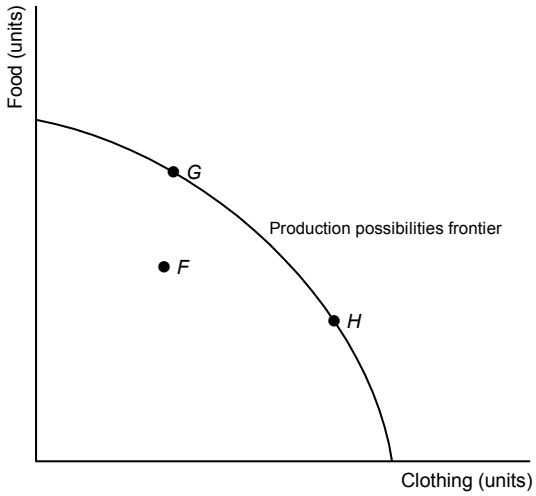


Figure 3.1 Production possibilities frontier

want. A fully efficient economy must produce the desired product mix. This mix depends on the distribution of income and preferences. The preferred mix could be anywhere along the PPF. We describe below how this preferred mix is defined and achieved.

Pareto efficient
A situation such that no one can be made better off without making someone else worse off

When all three efficiency conditions are achieved, the outcome is described as **Pareto efficient**. This means that resources cannot be reallocated so as to make someone better off without making someone else worse off.[1] Conversely, resources are used inefficiently if a reallocation could increase the welfare of one person without reducing the welfare of anyone else.

A reallocation of resources is Pareto efficient (or a Pareto improvement) if it raises the welfare of at least one person and does not reduce the welfare of anyone else. A reallocation is **potentially Pareto efficient** if there are losers but the benefits of a change exceed the costs (the net benefit is positive). In this case it is possible via compensation for some individual(s) to gain from the change and for no one to lose from it. In practice, changes with a positive net benefit are commonly described as efficient. However, if there are any losers the change is not Pareto efficient.

Efficiency in a Single Market

Perfect competition
A market for a homogeneous good with many informed buyers and sellers and free market entry

A basic proposition of economics is that a competitive equilibrium is Pareto efficient. A **competitive equilibrium** exists when a market contains many informed buyers and sellers, there is free entry and supply equals demand. This ensures that all trades that are valued by consumers and producers are made. Any good that consumers value above or equal to its marginal cost of production will be supplied. Goods with a value below the marginal cost of production will not be produced. In this section, we show that a single competitive market achieves an efficient outcome. In the next section, we examine the conditions required for economy-wide efficiency.

Figure 3.2 shows a competitive market with a standard downward-sloping demand curve and upward-sloping supply curve. Note first that the demand curve can be interpreted as a marginal benefit curve. The maximum price that someone is willing to pay for an extra unit of a good reflects the marginal benefit that he or she expects to receive from it. Second, the benefit that consumers obtain from trades (their consumer surplus) is the area between the demand curve and the horizontal price line (P_1), given by areas (A) + (B). Turning to production, the supply curve shows the quantity of goods that firms will supply at given prices. In a competitive market the supply curve is also a marginal cost schedule (it shows the cost of the last unit produced) because each firm maximises operating profit by producing up to the point where marginal cost equals price. There are no fixed costs in the supply schedule in this figure. The producer operating surplus is given by the area between the price line (P_1) and the marginal cost curve. Equilibrium market occurs where demand equals supply at quantity Q_1 and price P_1. Thus, at P_1, Q_1, the sum of benefits to consumers and producers is maximised.

This equilibrium point is Pareto efficient. When supply equals demand, all trades that consumers and producers value are made. The marginal benefit (MB) of an additional unit of output equals the marginal cost (MC). If MB = MC, it is not possible to change the quantity supplied and make someone better off without making another person worse off. On the other hand, suppose that the price is regulated at P_2 and supply reduced to Q_2. At this point, MB > MC. Output can be increased and consumers or producers, or both, can benefit from an increase in output. The loss of consumer and producer surplus is given by the sum of areas (B) + (D). Conversely, if MC > MB, output should be reduced and producers would benefit.

[1] This concept was named after the Italian economist Vilfredo Pareto (1909).

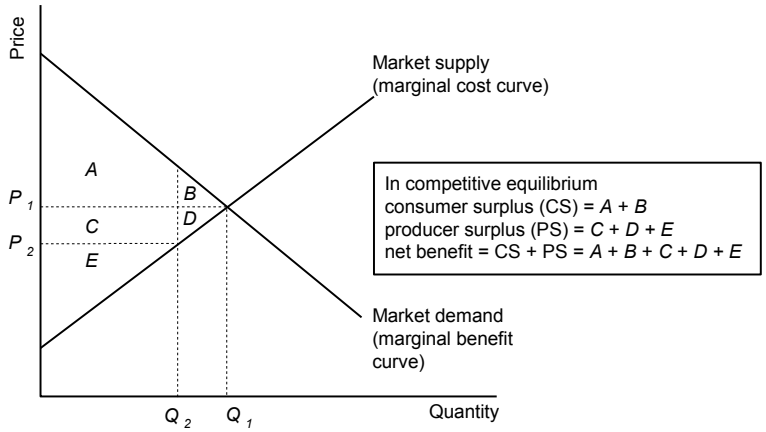


Figure 3.2 Efficiency in a competitive market

Given informed consumers, the critical condition for ensuring that MB = MC is that price equals marginal cost (P = MC). This is a key feature of perfectly competitive markets. Given a downward-sloping demand curve, informed consumers increase their purchases until their marginal benefit equals the market price. On the supply side, in a perfectly competitive market, firms take the market price as given. Production costs are minimised because inefficient firms do not survive. Firms maximise profits by increasing output until the marginal cost of production equals the market price.

It follows that, in perfectly competitive markets, MB = MC, consumer and producer surpluses are maximised and the outcome is Pareto efficient. No buyer or seller can be made better off by a move from Q_1 to another point without making someone else worse off.

However, the conclusion that a perfectly competitive market produces a Pareto-efficient outcome assumes that prices equal marginal cost in all related markets producing substitute or complementary goods. If this condition does not hold, marginal cost pricing in a competitive market may not produce a Pareto-efficient outcome. This important issue, known as the second-best problem, is discussed at several points below (e.g. in the discussion of pricing in Chapter 17).

Conditions for Efficiency in all Markets

We now examine how an efficient economy can achieve the three main efficiency conditions (efficient production, consumption and overall product mix) across all markets. We assume an economy with two individuals (Amy and Ben), two factors of production (labour and capital, e.g. machines), and two goods (clothing and food). Analysis of all markets concurrently is known as general equilibrium analysis.

Efficient production

Efficient production requires efficient use of the factors of production. Labour and capital must be used in such a way that the output of one good cannot be increased without a fall in output of another good. Figure 3.3 overleaf (an Edgeworth–Bowley box diagram) illustrates how this can be achieved.[2]

[2] This box diagram technique is named after two 19th century economists.

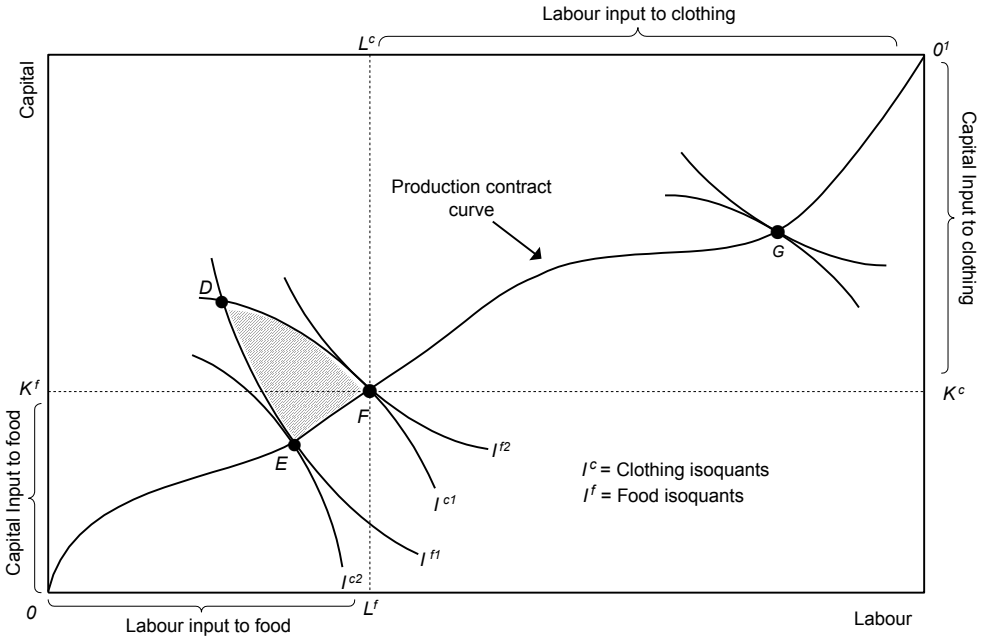


Figure 3.3 Production efficiency

The supply of labour and capital are shown on the horizontal and vertical axes respectively. Resources allocated to food are measured from the bottom left corner of the diagram (with origin *O*). Resources allocated to clothing are measured from the top right corner (with origin O^1). Thus, at point *F*, OL^f units of labour and OK^f units of capital produce food. All other labour and capital resources are employed to produce clothing.

Figure 3.3 also shows production isoquants for food (I^f) and clothing (I^c). An isoquant shows the combinations of inputs that produce a given output. The isoquants for food production are convex to the bottom left corner. Isoquants for clothing are convex to the top right corner. The slope of an isoquant at any point is the marginal rate of technical substitution (MRTS) of capital for labour — this is the marginal trade-off between two factors of production holding output constant. It shows the extra capital needed to maintain output at the same level when there is a marginal fall in labour. Convex isoquants imply a diminishing MRTS. As fewer units of capital (or labour) are employed, increasingly more units of labour (or capital) are required to achieve the given level of output.

Productive efficiency requires that, for any given output of food, output of clothing is maximised. Given convex isoquants, if food is produced at the level corresponding to isoquant I^{f1}, output of clothing is maximised by finding the clothing isoquant that is tangent to I^{f1}. Thus, at point *D* productive efficiency is not achieved. If we move to point *E*, for the same level of food (I^{f1}) more clothing can be produced (I^{c2} is higher than I^{c1}). This is a Pareto improvement. More generally, a move from *D* to anywhere between points *E* and *F* is a Pareto improvement because production of either or both goods increases, with no fall in the output of the other good. At any tangency point, the slopes of the isoquants are the same. This critical condition for production efficiency implies that the MRTS of capital (*K*) for labour (*L*) is the same for production of food (*f*) as for clothing (*c*).

$$MRTS^f{}_{KL} = MRTS^c{}_{KL} \tag{3.1}$$

Equation 3.1 can be generalised to multiple inputs, producers and goods. Efficient production requires that the marginal rate of technical substitution between any two inputs is the same for all producers who use both inputs in any market.

If the marginal rates of technical substitutions are not equal, production is inefficient. Suppose that, at the margin, machines are relatively efficient at producing food. For example, suppose that in food production the marginal output of one machine equals the output of three workers, but that, in production of clothing, the marginal output of one machine equals the output of only two workers. Food output could be increased, with no reduction in output of clothing, by allocating more machines to produce food and more labour to produce clothes.

Note the significance of the curve (*EFG*) that joins the isoquant tangency points. This curve shows all efficient combinations of food and clothing output that can be produced from the labour and capital available. The *EFG* line is known as the production contract curve. The output combinations on this curve define all points along the PPF.

Efficient consumption

Efficient consumption requires that, for any given incomes and preferences of consumers, goods are exchanged so as to maximise their satisfaction. Suppose that some point on the PPF is achieved and that Amy prefers clothes and Ben likes food. What shares of food and clothes would be efficient and how would markets achieve these shares?

Figure 3.4 presents an Edgeworth–Bowley box diagram for consumption. Here, the output of food and clothing are shown along the horizontal and vertical axes respectively. Note that this box can be viewed as sitting within a PPF, with the top right-hand corner touching the PPF. Amy's consumption of food and clothing is measured from this corner. Ben's consumption is measured from the left bottom corner. The diagram also shows the preference (indifference) curves of Amy and Ben for food and clothing.

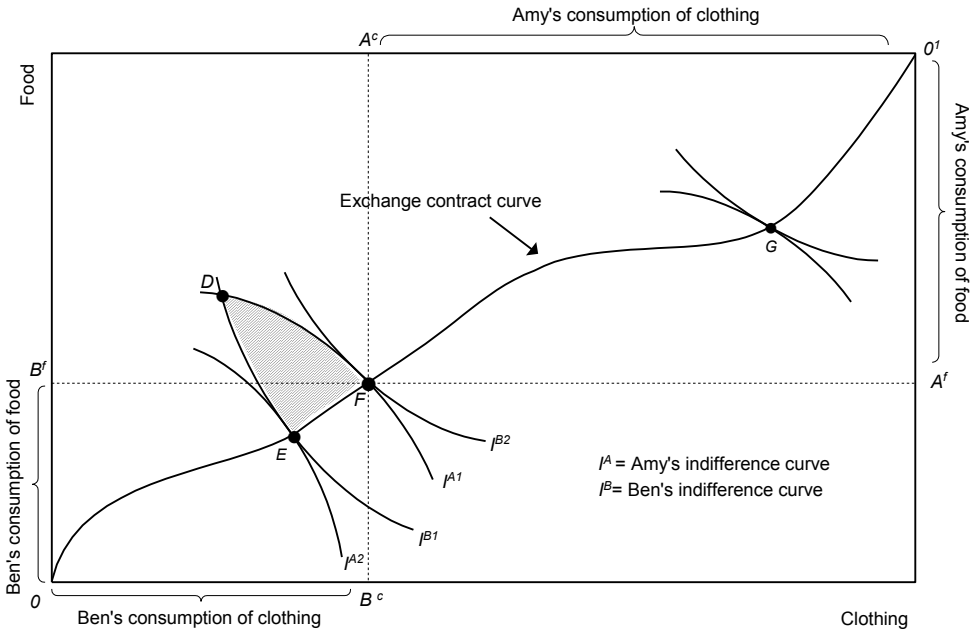


Figure 3.4 Consumption (exchange) efficiency

An indifference curve shows the combinations of food and clothing that provide a constant level of utility for each individual. For Ben, the curves (I^B) are convex to the bottom left corner; for Amy, they (I^A) are convex to the top right corner. Both are presumed to prefer a balanced set of goods to an unbalanced one. The slope of an indifference curve at any point is the marginal rate of substitution (MRS) of food for clothes — this is the marginal trade-off between two goods holding utility constant. It shows the extra amount of food needed to maintain utility at the same level when there is a marginal fall in amount of clothes. Convex indifference curves imply a diminishing MRS. As fewer units of food (or clothes) are consumed, increasingly more units of clothes (or food) are required to maintain the same level of utility.

Pareto efficiency requires that for any level of utility achieved by one individual, the utility of the other one must be maximised. Say, initially they are exchanging at point *D*. For Ben's given level of utility here (I^{B1}) it is possible to increase Amy's utility from I^{A1} to I^{A2} by moving to point *E*, where Amy obtains her highest utility (given Ben's utility) because her indifference curve is at a tangent to Ben's.

This move is a Pareto improvement. Indeed, a move from *D* to anywhere between points *E* and *F* is a Pareto improvement because the utility of either or both persons would increase. At *E*, the slope of Amy's and Ben's indifference curve is the same. At this point, Amy's marginal rate of substitution of food for clothes is the same as Ben's:

$$MRS^A_{fc} = MRS^B_{fc} \tag{3.2}$$

where *A* stands for Amy and *B* for Ben.

A move from *D* to *G* is not Pareto efficient. Ben is better off at *G* than at *D* but Amy is less well off. However, at *G,* the MRSs are equal. Thus, a move from *D* to *G* is a potential Pareto improvement.

Consider the implications if this marginal equality did not apply. Suppose that Amy is willing to exchange three units of food for an extra shirt, but that Ben would require only two units of food in return for giving up one shirt. Both would be better off if Ben gave shirts to Amy in exchange for food. The exchange should continue until Amy and Ben accept the same marginal rate of exchange.

The exchange conditions in Equation 3.2 can be generalised across all consumers and all goods. Pareto efficiency requires that the marginal rates of substitution between any pair of goods must be the same for each individual who consumes the goods on offer. If this condition is not met, exchange could make at least one person better off without making someone else worse off.

Two more points should be made. First, the efficient MRS may change if the distribution of income changes. In Figure 3.4, the MRS at point *F* may be different from the MRS at point *E*. Second, the various efficient exchange points shown in Figure 3.4 make up the Pareto-efficient exchange contract curve (*OEFGO*[1] in Figure 3.4).

General product mix or overall efficiency

The third efficiency requirement is that firms must produce the mix of goods that individuals want given production constraints. The point on the PPF must be the preferred point given individuals' incomes and preferences. To satisfy this condition the value of a marginal unit of a good must be the same for consumers as for producers. Therefore, the marginal rate at which consumers wish to exchange food for clothing must equal the marginal rate at which producers can transform food into clothing. Formally,

$$MRS^A_{fc} = MRS^B_{fc} = MRT_{fc} \tag{3.3}$$

where MRT_{fc} is the marginal rate of transformation of food into clothing, that is, the rate at which the economy transforms clothes into food (the negative of the slope of the PPF).

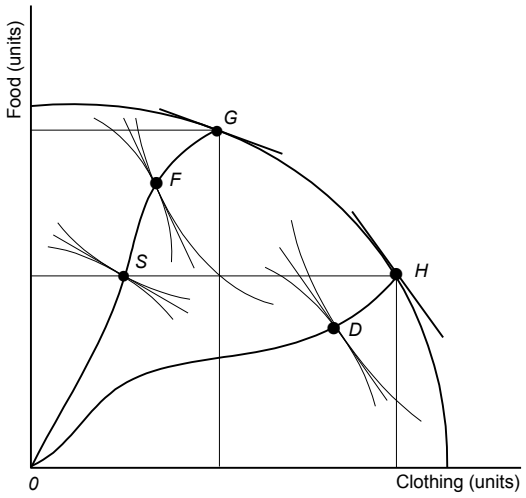


Figure 3.5 General (product mix) efficiency

Of course, clothing is not transformed directly into food. Rather the resources used can be employed for another purpose. Thus, an MRT_{fc} of 3:1 implies that three units of food can be produced by forgoing one unit of clothing. Suppose that MRS_{fc} = 4:1, which states that consumers are willing to give up four units of food for each unit of clothing forgone, and that MRT_{fc} = 3:1. This implies that more clothing should be produced until equilibrium is achieved at some ratio between 4:1 and 3:1.

The product-mix (overall) efficiency condition of Equation 3.3 is illustrated in Figure 3.5. Graphically the slope of the indifference curves must equal the slope of the PPF. Suppose that the economy is on the PPF at point G. The implied consumption Edgeworth–Bowley box for this output has origins at *O* and at *G* (corresponding to *O* and O^1 respectively in Figure 3.4). Assume also that the allocation *F* of the total output between consumers is efficient so that *F* is on the efficient exchange contract curve *OG* (which is the locus of points of tangency of Amy and Ben's indifference curves). Now suppose that the common slope of their indifference curves at *F* is not equal to the slope of the PPF at *G*. This means that $MRS^A_{fc} = MRS^B_{fc} > MRT_{fc}$.

Because the indifference curves at *F* have a steeper slope than the PPF at *G*, consider shifting inputs from food to clothes production, moving down PPF to *H*. This creates a new consumption Edgeworth–Bowley box with the same *O* origin but *G* shifted to *H*. We give Amy the same consumption bundle as before so that her consumption allocation is now at *D* where the distance *DH* = *FG*. Amy is therefore on the same indifference curve: her indifference curves drawn at *F* and *D* are the same measured from the origins *G* and *H* respectively. So, Amy's utility is not affected by this change in output mix. Ben's consumption is still measured from *O*. However, he is better off at *D* than at *F* because *D* lies above his indifference curve through *F*. Thus, the change in the allocation from *F* to *D* is a Pareto improvement. At *D*, Equation 3.3 is satisfied. Firms are producing the mix of goods that consumes want.

Work-leisure efficiency

The analysis above assumes that individuals can choose an efficient mix of work and leisure. Work–leisure efficiency requires that the marginal rate at which someone is willing to substitute leisure for income (market goods) should equal the marginal rate at which he or she can transform leisure into income.

$$MRS_{ly}^{A} = MRT_{ly}^{A} \tag{3.4}$$

where l is leisure and y is income. If the value of Amy's marginal output exceeds the value that she places on her leisure time, Amy is under-employed.

Efficiency, Equity and Social Welfare

So far, we have taken the distribution of income as given and determined efficient outcomes. However, many of these outcomes would be inequitable. To examine the relationship between efficiency, equity and welfare we introduce the related concepts of a point utility possibilities curve and the utility possibilities frontier.

A **point utility possibilities curve** (PUPC) shows the maximum utilities that individuals can obtain from different distributions of a given output of goods. Equivalently, it shows the utilities that Amy and Ben derive from points along the exchange contract curve. Take a point on the PPF in Figure 3.5, such as *G*, and distribute this output in all possible proportions from all output going to Amy to all to Ben. This replicates the contract curve *OG* in Figure 3.5. This would produce a PUPC such as *GG* in Figure 3.6. Now if we adopt the same process for output mix *H* we may get PUPC *HH*. In this way PUPCs may be drawn for any output along the PPF and for any division of this output. Note that PUPCs may cross if individuals have different preferences, for example if Amy prefers clothes and Ben prefers food.

Note also that points *F* and *S* in Figure 3.6 represent the same allocations as *F* and *S* in Figure 3.5. As we have seen, point *F* does not satisfy the overall product-mix efficiency condition because the MRSs at this point are greater than MRT. Given Amy and Ben's incomes and preferences, there will be another output combination at which more clothes and less food are produced which can make one or both of Amy and Ben better off than at *F*. There will be a point such as *D* on *HH*, which is Pareto superior to *F*.

Utility possibilities frontier
Shows the maximum utility one person can achieve given the utility of the other person

The **utilities possibilities frontier** (UPF) shows the maximum utility that Amy or Ben can achieve, given the level of utility obtained by the other party for any set of goods. Graphically it is the outer envelope of all points of Pareto efficiency. Figure 3.6 shows the outer envelope for just two points on the PPF. The full UPF is derived from a complete set of product mixes of food and clothing and a complete set of income distributions (for given tastes). If the PUPCs cross, the UPF is a jagged curve as in Figure 3.6 or Figure 3.7.

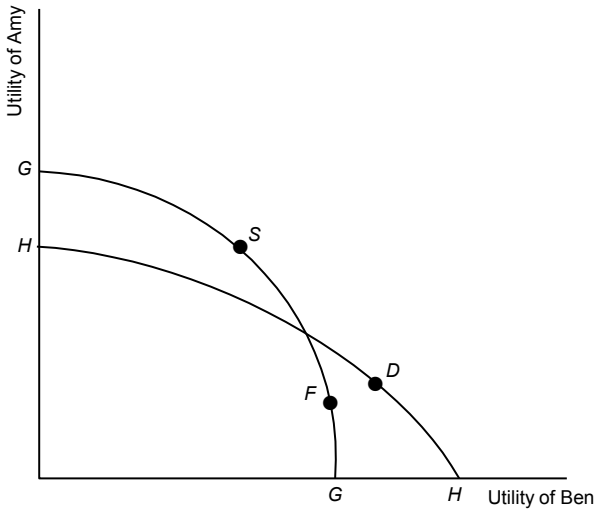


Figure 3.6 Point utility possibilities curves and the utility possibilities frontier

By describing all points on the UPF as (Pareto) efficient and by not choosing between these points, economists are attempting to make statements about the nature of markets and efficiency without making judgements about the social value of alternative distributions of welfare.

Value judgements and social welfare

However, value judgements cannot be avoided.[3] Even the concept of Pareto efficiency, when interpreted as a welfare criterion, embodies three value judgements. First, the welfare of society depends only on some defined welfare of individuals and there is no social interest beyond this. Second, each individual is the best judge of their own welfare. Third, social welfare is an increasing function of the welfare of all individuals. If Amy's utility increases when Ben's stays constant, the change is Pareto efficient and society is said to be better off. Amy and Ben's utilities are assumed to be independent. But this may not be so.

More fundamentally all positions on the UPF cannot be regarded as equally desirable. Pareto efficient points include positions where Amy or Ben has minimal welfare. Such positions are usually unacceptable. Other points on the UPF are more equitable and usually more desirable. Indeed some points off the UPF may be more desirable than some points on it. This highlights a critical difference between efficiency and optimality. An optimal outcome takes into account efficiency and equity.

To take both efficiency and equity into account, we need a **social welfare function** that expresses social welfare as a function of the level and distribution of utilities of individuals. Thus suppose that social welfare (W) is:

Social welfare function
A function that relates the overall welfare of society to the welfare of its members

$$W = f(u_1, u_2, \ldots u_n) \tag{3.5}$$

where u represents utility and there are $1 \ldots n$ citizens.[4] We can allow for distributional concerns by giving differential weights to u_i (the utility of each individual). A social welfare function enables us to rank social states and choose between Pareto-efficient outcomes.[5] However, the choice of social welfare function requires ethical judgement(s).

Similar problems arise when considering any use of resources. Most uses affect someone adversely, so that limiting decisions to Pareto-efficient improvements would be highly restrictive. As we have noted, economists often describe a use of resources as efficient if the net benefit is positive (there is a potential Pareto improvement). If the gainers compensate the losers, the resource usage would result, after compensation, in at least one person gaining and no one losing. This approach underlies much policy analysis (see discussions relating to cost-benefit analysis in Chapters 7 and 8). However, if compensation is not achieved, as it often is not, there is again a trade-off between efficiency and equity.

Figure 3.7 illustrates the idea of a social welfare function. In the figure, W_1, W_2, and W_3 represent the social indifference curves from Equation 3.5 for a two-person economy. A social indifference curve is a locus of points which provide equal social welfare for various combinations of utilities between Amy and Ben.

---

[3] Value judgements involve a judgement as to what is good. This in turn involves a subjective and unprovable ethical judgement

[4] This individualistic social welfare function (SWF) is known as a Bergson–Samuelson SWF. Various forms of SWF are discussed in Chapter 7.

[5] Formally, the public policy aim may be represented as maximising social welfare ($W = f(u_1, u_2)$) subject to constraints. The constraints are the preferences of consumers ($u_1 = f(x_1, y_1)$ and $u_2 = f(x_2, y_2)$), production technology ($x = f(K_x, L_x)$, $Y = f(K_y, L_y)$) and resource constraints ($K = K_x + K_y$ and $L = L_x + L_y$), where x and y are goods and K and L are capital and labour respectively.

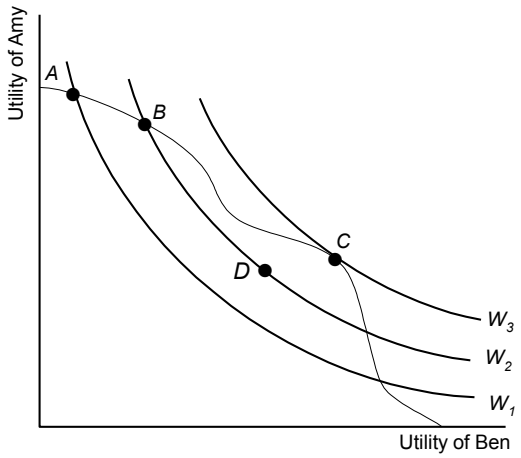


Figure 3.7 Utility possibilities and social welfare

In Figure 3.7, the higher the value of the suffix, the higher the level of welfare. Given the UPF, social welfare is maximised at *C*. Starting from *A* or *B*, a redistribution of income from Amy to Ben increases social welfare. Further, given the social welfare function as shown by the social indifference curves, a movement from *A* to *D* would also increase welfare, even though *D* is not on the UPF, with the gain in equity more than compensating for the loss of efficiency. However, in the absence of knowledge of a specific social welfare function (and it is not possible to achieve this by economic reasoning alone), one cannot reach the preferred decision.

Competitive Markets and Efficiency

What kind of economy achieves production, consumption and product mix efficiency? The answer is a perfectly competitive economy.[6] The First Theorem of Welfare Economics states that if there are markets for all goods and all markets are perfectly competitive, an economy achieves a Pareto-efficient outcome.[7]

The key features of a perfectly competitive market are that there are many buyers and sellers, the sellers are supplying similar goods and firms can freely enter or exit the market. There are no excess profits. Also buyers and sellers are assumed to be well informed about prices and the quality of goods. Given these conditions, the market collectively determines all factor and product prices. No one firm or individual can affect a price. All buyers pay the same price for a similar good. Prices adjust to clear all excess demand and supply, equilibrium quantities and prices are achieved, and all Pareto-efficient trades are made.

Consider first productive efficiency. In a perfectly competitive market inefficient firms will not survive. Each firm employs factors until the value of their marginal product (VMP) equals their price.[8] Suppose that the rental price (r) of a machine is twice the price of a unit of labour (w). Each firm will purchase machines until their marginal product (MP) falls to two times the marginal product of labour. The marginal rate at which capital is substituted for labour equals

[6] For a more formal (mathematical) proof of the First Theorem of Welfare Economics, see most intermediate microeconomic texts, for example Varian (2006).

[7] In theory, an omniscient government managing a complete command and control economy could also achieve a Pareto-efficient allocation of resources

[8] VMP is the change in revenue due to the sale of the additional output contributed by the hiring of one more unit of a factor of production

the ratio of the marginal product of labour to the marginal product of capital, which in turn equals the price ratio of labour to capital.

$$MRTS^{f}_{KL} = MRTS^{c}_{KL} = \frac{MP_L}{MP_K} = \frac{w}{r} \tag{3.6}$$

Because all firms face the same input prices, all firms employ the same marginal rate of technical substitution between inputs, as in Equation 3.1. This ensures productive efficiency.

Turning to consumption, following demand theory (see Chapter 6), all consumers are assumed to behave in accordance with a systematic set of preferences and to aim to maximise utility given market prices and income constraints. Individuals maximise their utility when the marginal rate at which they wish to substitute one good for another is in inverse proportion to the relative prices of the two goods. For example, if Amy is willing to substitute three units of food for one unit of clothing, the price of food must be one-third of the price of clothing. Thus, we have

$$MRS^{A}_{fc} = MRS^{B}_{fc} = \frac{P_c}{P_f} \tag{3.7}$$

Because all individuals face the same relative prices in competitive markets, all individuals have the same marginal rates of substitution. This ensures efficient consumption.

Third, consider product mix efficiency. To show that a perfectly competitive market produces an efficient mix of products, it is useful to express the MRT in terms of marginal cost (MC). MC is the incremental cost of one more unit of output. Recall that MRT is the slope of the PPF. The slope also represents the ratio of the marginal costs of producing clothing and food. The marginal cost of producing clothes is the food forgone. Conversely, the marginal cost of food is clothing forgone. Thus,

$$MRT_{fc} = \frac{MC_c}{MC_f} \tag{3.8}$$[9]

In perfect competition, the market determines the price of a product and producers expand output until marginal cost equals price. Therefore,

$$\frac{MC_c}{MC_f} = \frac{P_c}{P_f} \tag{3.9}$$

Combining Equations 3.8 and 3.9, the marginal rate at which clothes are transformed into food must equal the price ratio of food to clothing. Equation 3.7 shows that the MRSs of all consumers equal the same price ratio. Therefore a perfectly competitive market satisfies the necessary condition for product mix efficiency, Equation 3.3, that the MRTs by all producers for all goods equal the MRSs of all consumers for those same goods.[10]

Finally we should note that a perfectly competitive market also produces an efficient amount of work and leisure. In a competitive labour market the wage equals the value of the marginal product of labour. Employment expands so long as the value of the marginal product exceeds the value of leisure. In equilibrium, the wage and the marginal product equal the opportunity cost of leisure forgone.

---

[9] $MC_f = \Delta TC_f/\Delta f$ and $MC_f = \Delta TC_c/\Delta c$. Suppose that at some point on the PPF, a small amount of resources valued at ΔTC is transferred from producing food to producing clothing. Now, the change in resource costs in food (i.e. ΔTC_f) and that in clothes (i.e. ΔTC_c) both equal ΔTC. Thus $MC_c/MC_f = (\Delta TC_c/\Delta c)/(\Delta TC_f/\Delta f) = \Delta f/\Delta c$, which is the slope of the PPF ($=MRT_{fc}$)

[10] Equation 3.9 is another way to represent the efficiency condition. It shows that the marginal cost of each commodity must be reflected in its price. Thus if the opportunity cost of a commodity is relatively high, efficiency requires that its price be relatively high—this is because the high price signals the consumers to economise its use.

The work–leisure efficiency condition is not met in uncompetitive labour markets, where employers or employees restrict entry into the labour force in some way. Government regulations and taxation also distort the conditions for work–leisure efficiency.

Implications of the First Theorem of Welfare Economics

The First Welfare Theorem provides the formal basis for the efficiency of markets. By showing that a complete set of competitive markets produces a Pareto-efficient outcome, it formalises Adam Smith's famous argument that individuals in the pursuit of their own interests are led by the invisible hand of a competitive market to work in the general interest. Government's role in ensuring that resources are allocated efficiently would be limited to ensuring the effective operation of competition in all factor and goods markets.

This would not be a small task. Government would establish and enforce the rules for the economy, the interactions between private parties and the interactions between private agents and government. Government would establish and protect property rights and the commercial system of contracts and exchange. Fulfilment of contracts is essential for an effective market system.

However, the First Theorem has three major limitations. First, few markets are perfectly competitive. Most economies contain various imperfectly competitive markets, in which one or more of the requirements of a perfectly competitive market do not exist. In some sectors, for example for some environmental goods, there are no markets at all. When the conditions for perfect competition do not exist, there is said to be market failure and a potential role for government in the allocation of resources is established. Second, in the model of the economy described above, human capital and technology are taken as given. As we will see in Chapter 5, a perfectly competitive economy can produce an efficient inter-temporal allocation of resources. However, the perfect competition model of the economy does not explain technological change or economic growth. Third, the First Theorem says nothing about the equitable distribution of income. Pareto efficiency ensures only that some point on the UPF is achieved. But many positions on this frontier are socially unacceptable. The socially preferred outcome depends on equity as well as efficiency criteria. We now address this issue.

Competitive Markets and Equity

As we have seen, at many points on a UPF the distribution of welfare can be highly unequal. For example, in Figure 3.7 points *A* and *B* represent high levels of utility for Amy and low levels for Ben. To determine an optimal outcome, we need a method for choosing between points on the UPF and possibly also between some states of welfare that are not on the frontier. For this we need a social welfare function (SWF) that will enable us to rank economic states. Needless to say the choice of SWF depends critically on views about equity. These issues are taken up in Chapter 7. For the discussion here, we suppose that there is an agreed SWF and that social states can be ranked and we consider the implications for competitive markets and the role of government.

The Second Theorem of Welfare Economics. This is where the Second Theorem steps in. This theorem states that any Pareto-efficient allocation (any point on the UPF) can be achieved by perfectly competitive markets if society starts with the appropriate distribution of resources or if resources can be so redistributed without cost. Suppose that, with the existing distribution of wealth, a competitive market would produce an outcome at point *B* in Figure 3.7, but that society prefers point *C*. The Second Theorem shows that any position such as *C* can be achieved by a lump sum redistribution of *initial individual endowments* followed by the operations of perfectly competitive markets (see, for example, Varian 2006). A lump sum transfer is a fixed amount that does not change with a change in circumstance of the taxed

entity. It is difficult to achieve a move from *B* to *C* after *B* has been achieved because this implies that a specific set of goods has been produced. A redistribution of these goods might not produce a move from *B* to *C*.

The Second Theorem has important implications. It shows, in principle, that decentralised competitive markets combined with individualised lump sum transfers can achieve any desired distribution of welfare subject to production constraints. Moreover, government could achieve this result without intervening subsequently in the allocation of resources in markets. Equity and efficiency would thus be separated. If this were feasible, government would be responsible for ensuring an appropriate set of initial endowments. The competitive economy would then ensure an efficient and optimal allocation of resources.

The Second Theorem assumes that endowments can be redistributed without distorting the use of resources. This means that any redistribution must not disturb the marginal conditions necessary for Pareto efficiency (Equations 3.1 to 3.3) or the critical relevant price relativities (Equations 3.6 to 3.9). Nor must government distort the critical condition for an efficient work–leisure split, namely that workers receive the full value of their marginal output. If income taxation or any other method of redistributing income distorts any of these marginal relationships, there will be a loss of output and/or utility.

Endowments can be redistributed without cost if individualised lump sum transfers are possible. A transfer (a tax or a grant) is a **lump sum transfer** when the amount of the transfer is not affected by the taxpayer's or recipient's actions. In this case the transfer would not affect an individual's incentive to work or consume or undertake any other form of economic activity. Prices would still equate to marginal cost. If lump sum transfers are possible, government could redistribute income to achieve any desired Pareto-efficient outcome.

Lump sum transfer
A transfer that is not affected by the behaviour of the taxpayer or the recipient

However, individualised lump sum transfers based on an individual's capacity are virtually impossible to achieve. Taxes or grants that are based on behaviour are liable to change behaviour and are therefore not lump sum taxes. Lump sum taxes based on fixed personal attributes, such as a person's height, sex or IQ, are likely to be arbitrary or unfair, or both. Moreover, virtually all transfers relate to behaviour in one way or another and are likely to change behaviour. This means that they almost always have some efficiency effects and result in some loss of real income as a trade-off for a gain in equity.

Conclusions

An efficient economy produces the goods that people most want at least cost. In itself, this is desirable. However, an efficient economy does not guarantee equity and indeed may be quite inequitable.

The conditions for an efficient allocation of resources provide a valuable guide to how resources should be allocated. Efficient production requires that the marginal rate of technical substitution between any two factors of production must be the same for all firms who use both factors. Efficient consumption requires that the marginal rates of substitution between any pair of goods must be the same for each individual who consumes the goods on offer. Overall product mix efficiency requires that the economy must produce goods in combinations that match people's willingness to pay for them. This occurs when the marginal rate at which firms can transform any two goods equals the marginal rate at which consumers wish to exchange the goods. Work–leisure efficiency requires that individuals receive the value of their marginal product.

The unifying requirement for all these conditions of efficiency is that the prices for all factors of production and for all goods should equal their marginal cost. This is achieved in a perfectly competitive economy. Accordingly, a perfectly competitive economy achieves a Pareto-efficient allocation of resources (an economy where no one can be made better off without making someone worse off). If individualised lump sum transfers can then effect a

desired redistribution of endowments, a perfectly competitive economy can produce an equitable as well as an efficient outcome.

The general equilibrium economy portrayed here is based on standardised (homogeneous) goods in a static environment. In more complex models, goods are characterised by four main features—their location, time and uncertainty, as well as their homogeneous quality. Advanced texts, for example Hindricks and Myles (2006), show that, if markets exist for all goods with all these characteristics, perfectly competitive economies can achieve an efficient allocation of resources over time. However, even these more complex models of the economy often take technology as given. As we see in Chapter 5, technical progress (and economic growth) is facilitated in a competitive economy. This is an important additional advantage of competitive markets.

However, no economy is perfectly competitive. In all economies market failures occur because of non-existent, incomplete and imperfectly competitive markets. When market failures exist, markets do not produce an efficient allocation of resources. Moreover, if there are several market failures and two or more conditions for efficiency are not satisfied, achievement of the other conditions is not necessarily beneficial. This is known as the second-best problem. In such cases an examination of policy options is required.

Finally, individualised lump sum transfers as a function of individuals' capacities are both unfeasible and would change incentives and behaviours. This means that society can achieve a more equitable distribution of income only by sacrificing some efficiency in use of resources.

Summary

- An efficient economy produces the maximum amount of goods that people want given their preferences and the productive resources and technology available.
- An economy is described as Pareto efficient if no one can be made better off without someone else being made worse off. There are no unexploited economic gains.
- Overall economic (Pareto) efficiency requires that production and consumption (exchange) are efficient and that the optimum mix of goods is produced.
- These efficiency conditions are achieved in a perfectly competitive economy. A key reason is that in such an economy the prices of all factors of production and of all goods equal their marginal costs.
- However, an efficient economy is not necessarily equitable because some people have low earning capacity.
- If government could redistribute resources using individualised lump sum transfers, then a perfectly competitive market could produce efficient and fair outcomes.
- In practice, markets are far from complete or perfectly competitive. Also, government cannot redistribute resources without affecting individual behaviour and distorting markets.
- Therefore, government involves regulation of markets, redistribution of income and trade-offs between efficient and equitable outcomes.

Questions

1. What is the relationship between achieving Pareto efficiency and maximising net social benefit defined as the difference between total benefits and total costs?
2. When can win–win outcomes occur rather than outcomes where there are gainers and losers?
3. How are the utility possibilities frontier and the production possibilities frontier related? Why does the utility possibilities frontier matter?
4. The First Welfare Theorem claims that competitive markets result in a Pareto-efficient outcome. Explain briefly why. What is the relevance of the assumptions that producers maximise profits and consumers maximise utility?
5. Determine the equilibrium quantity and price in a competitive market for shirts, assuming the supply of shirts is given by $Q^S = 20 + 4P$ and the demand for shirts is $Q^D = 65 - 5P$. Will this equilibrium outcome be Pareto efficient?
6. Amy and Ben have different tastes such that Amy wants a large number of shirts and Ben wants large quantities of beer. But in market equilibrium they are prepared to exchange the same amount of beer for a shirt. Explain this apparent paradox.
7. Suppose that a manufacturing company requires either three units of capital or two units of labour to maintain a given level of output, while a service company would require two units of labour for each unit of capital for its level of output. What is the marginal rate of technical substitution of inputs for each firm? Does this represent a situation of efficient production? Why/why not?
8. Suppose that an economy produces two goods, clothes and food. At current margins, consumers are willing to exchange four units of food for one unit of clothing and firms can produce one unit of clothing at the expense of two units of food. Is this an efficient allocation? To be efficient, should the economy produce more clothes or more food?
9. Why is it inefficient to charge two consumers different prices for the same good? Airlines often charge passengers different prices for similar seats on the same flight. Is this consistent with efficient consumption?
10. Consider an economy that produces clothes and bread. Explain why the economy will not produce an efficient mix of products if a tax is imposed on the consumption of shirts but not on bread. What kind of deadweight loss will be incurred?
11. Discuss the following propositions.
    i. A move from a point within the utility possibilities frontier to a point on the frontier is always a Pareto improvement.
    ii. A Pareto improvement is a necessary and sufficient condition for an increase in social welfare.
12. If all outcomes from a competitive economy are efficient, why do we need a social welfare function?

Further Reading

Friedman, L. (2002) *The Microeconomics of Public Policy Analysis*, Chapter 12, Princeton University Press, Princeton, New Jersey.

Harford, T. (2006) *The Under Cover Economist*, Chapter 3, Little, Brown, London.

Hindriks, J. and Myles, G.D. (2006) *Intermediate Public Economics*, Chapter 2, MIT Press, Cambridge, MA.

Lange, O. (1942) 'The foundations of welfare economics', *Econometrica*, 10, 215–228.

Pindyck, R.S. and Rubinfield, D.L. (2005) *Microeconomics*, Chapter 16, Prentice Hall, New York.

Rodrik, D. (2015), *Economics Rules: The Rights and Wrongs of the Dismal Science*, Chapter 2, W.W.Norton, New York.

Rosen, H. and T. Gayer (2014), *Public Finance*, 10th ed., Chapter 2, McGraw-Hill, New York.

Chapter

4

Market Failures, Equity and Government

The important thing for Government is not to do things which individuals are doing already, and to do a little better or worse; but to do those things which at present are not done at all.

Maynard Keynes, *The End of Laissez-Faire*

Public Goods ♦ Externalities ♦ Imperfect Competition ♦ Information Failures ♦ Other Market Failures ♦ Equity Issues ♦ Government Functions ♦ Government Instruments

In Chapter 1, we saw that government has four main economic functions. These are establishing a legal infrastructure that enables markets to work, resource allocation functions when there are market failures, provision of social welfare including a fair distribution of income, and macroeconomic management.

As we have seen, an efficient market economy requires a complete set of competitive markets and fully informed buyers and sellers. However, markets may not exist or be incomplete or uncompetitive and economic agents may be imperfectly informed. And there are numerous reasons why markets may not produce equitable, or socially acceptable, income or welfare outcomes.

In this chapter, we discuss the four major forms of market failure: public goods, externalities, imperfect competition and information failures. We also discuss the role of government in providing equity and social welfare. Finally, we bring this together by discuss the main functions of government that arise from the analysis of market failures and equity. We also describe the main instruments available to government to carry out these functions.

Public Goods

In economics, the term "public good" is a technical term. It refers to a good which is **non-excludable** or **non-rival**, or in some cases both non-excludable and non-rival. When these attributes occur, there is market failure. Markets are unlikely to produce an efficient amount of the good. Hence, government has a responsibility to ensure that an efficient amount is produced. This good may then be produced by the public or private sector. Also, the term 'public good' does not describe all goods supplied by government. For welfare reasons, government supplies many goods that provide only private benefits to the recipients, for example some health care services.

A public good
A good that is non-excludable or non-rival in consumption

A good is **non-excludable** if individuals or firms cannot be excluded from consuming it or, more broadly, cannot be excluded from obtaining any benefit from it. National defence and

Table 4.1 Public and private goods

	Exclusion not feasible (or is costly)	*Exclusion feasible*
Consumption non-rival	National defence	Some economic infrastructure services, such as
	Law and order services	national highways and ports
	Basic financial services	Fire protection
	Benefits of basic education	Outputs of research and development
	Public health services	Cable TV broadcasts
	Urban planning and design services	Parks and beaches (until congested)
Consumption rival	Crowded urban roads	Food and clothing
		Seats at football matches
		Heart bypass operations

provision of internal security are classic non-excludable public goods. Education may also be viewed as a public good (as well as a private one) because firms and society more generally benefit from the supply of more educated workers although they do not receive the educational services directly. Non-excludability occurs when the supplier cannot charge for a good or benefit. Note that these benefits imply significant positive externalities. A positive externality occurs when an external party to the market transaction benefits from an economic activity but does not pay for it. When the scale of the positive externality is large, economists generally describe the good as a public good.

Competitive markets provide goods efficiently when firms can identify the individuals who benefit from the goods they provide and can charge for consumption of the goods. When consumption is non-excludable, individuals can receive the benefits of the good without paying for it. This is known as **free riding**. A free rider is a person or firm that benefits from a public good but does not contribute to its provision. In such cases, firms cannot receive full or sometimes any payment for the goods provided and *so will under-supply the goods*.

Free riding
Consuming a good without contributing to it

A good is described as **non-rival** when someone's consumption of a good does not prevent someone else from consuming the same good, for example listening to a broadcast weather report. The additional use in this case has zero marginal cost. When consumption of a good is non-rival, the efficient charge is the marginal cost of use, which may be low or even zero. However, if the good is excludable, firms may charge above marginal cost and *there will be under-consumption of the good*.

Public goods are most commonly goods that provide collective benefits to large numbers of people or firms rather than private benefits to individual consumers. As well as national defence, internal security and education, examples include surveying services as the basis for property rights, regulation of the financial system, flood management, public health services, waste collection services (garbage and wastewater) and various forms of economic infrastructure.

Table 4.1 shows examples of public and private goods, classified by their excludability and rivalness characteristics. Note that these are guides, not precise concepts. Excludability depends on technology and can often be achieved at a cost. For example, electronic sensors can be used to collect tolls for most roads. The consumption of some goods, such as roads or parks, is non-rival when demand is low but rival when use and congestion increase.

Externalities

Externality
An impact on an individual or business not associated with carrying on the activity that causes the impact

An **externality** occurs when a cost or benefit arises from an activity that does not accrue to the firm or person carrying on the activity. These effects typically occur because of a lack of enforceable property rights. External impacts may be positive or negative.

When the external impact is positive and widespread, the activity generating it may be viewed as a public good. But smaller positive externalities are not generally classed as public

goods. For example, a household's purchase of waste services or a farmer's property improvements may benefit neighbouring households or properties respectively. However, when the agency creating the positive externality cannot charge for it, the improvements may not be made. More generally, markets tend to under-supply goods with positive externalities.

A negative externality occurs when an activity has an adverse impact on a third party. Examples are industrial air or water pollution, disposal of waste in public places, and traffic congestion. Negative externalities are common and of special concern. In the absence of government regulations of one or another kind, agents have no incentive to prevent negative externalities. Rather, a firm may save expenses while causing a negative externality, as when it discharges untreated wastes. This markets over-supply goods with negative externalities.

Note, however, an important distinction between physical and pecuniary externalities. **Physical externalities** affect the production of firms or the welfare of individuals and are not transmitted by market prices. Such externalities usually involve inefficient use of resources.

On the other hand, **pecuniary externalities** are transmitted through changes in prices. Any change in demand or supply may alter prices and the welfare of a third party. For example, a natural gas discovery may reduce the demand for coal and consequently reduce prices and wages in the coal industry. When prices change, for every loss there is an offsetting gain. Price changes have distributional effects and may be a major social or political concern. But they are not in themselves an inefficient use of resources.

Effects of negative externalities. A negative externality implies that the total cost of an activity exceeds the private cost. Consider production of electricity from a coal-fired generator that emits carbon dioxides that cause global warming and sulphur dioxides that harm health. Figure 4.1 shows demand and supply for electricity for a firm in a competitive market. In this market, the price is given and the demand curve for each firm is horizontal. There are two cost schedules. Schedule S_P is the marginal private cost curve, which shows the firm's marginal cost as a function of output. Given the market price P_M, a profit-maximising firm will produce Q_M units of electricity. Schedule S_S is the marginal social cost curve, which includes the marginal damage costs of the negative externalities (which are assumed here to rise with output). Output is efficient when the benefit of a marginal unit of power equals the marginal social cost. Thus, Q_E is the efficient level of output. The firm, which ignores the damage costs produces excessive quantities of electricity and air pollution.

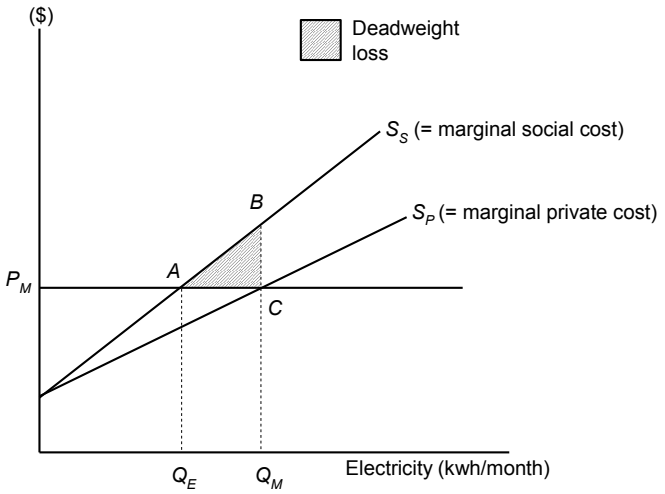


Figure 4.1 Inefficiency with negative externality in a competitive market

The measure of economic inefficiency is known as the **deadweight loss** (DWL). This is a loss for which there is no offsetting benefit. Identifying and quantifying DWLs due to inefficient use of resources is fundamental to economic analysis. The DWL in this case is the excess of social marginal cost over marginal benefit, i.e. the shaded area *ABC* in Figure 4.1.

Deadweight loss
A loss for which there is no offsetting benefit

In general, because a firm producing a negative externality does not bear the cost, it has no incentive to reduce the damage. Consequently, the product mix is inefficient because the marginal rate at which firms can transform one good into another is inconsistent with the marginal rate at which consumers wish to exchange one good for another. Therefore, the conditions for the First Welfare Theorem are not met. As shown in Figure 4.1, some external cost may be consistent with economic efficiency (at output of Q_E). However, unregulated markets generally produce too many goods that cause external damage and deadweight losses.

Common property resources

Common property resources are another source of inefficient resource use with externality features. **Common property resources** are resources that are open for free use to all. Examples include local public commons, air and water, some fossil fuel deposits and animal resources such as wildlife and fisheries. Many environmental goods are essentially common property resources. Because use of the common property resource is free and does not reflect the (rival) value of the resource to other users, the resource is likely to be over-utilised.

Common property resource
A resource that is open to all for use

A classic example is overgrazing of the village commons (Hardin, 1968). When land is plentiful a herder can graze all his or her sheep on it without affecting the productivity of the land for other herders. As the number of sheep on the land increases, the amount of grass per sheep declines and the fertility of the land declines. Each extra sheep on the commons reduces the value of the commons to other herders. Collectively the herders would gain by maintaining the sheep flock at an efficient size, where the marginal benefit of an extra sheep on the commons equals the marginal social cost. But herders do not pay the marginal damage cost of grazing extra sheep and have no incentive to reduce their flock on the commons because the benefits would accrue to other herders. Thus, the commons are overgrazed.

The problem can be illustrated by a prisoner's dilemma game. Table 4.2 shows the benefits, using an arbitrary numeraire, to two herders of two strategies (restricting sheep numbers or placing a large number on the commons). The outcomes depend on both their decision and the decision of the other herder. For each herder, the dominant strategy is to place a large number of sheep on the commons. This maximises their gain whatever strategy the other herder adopts. However, if both herders follow this strategy, both finish up with fewer benefits from the commons than they would with a cooperative strategy! Both herders would gain from an agreement to restrict sheep numbers on the commons.

Similar issues arise with other common property resources. The modern problem of open access fisheries is discussed in Chapter 12. Without appropriate controls on access to the resources, agents have little incentive to use the resources efficiently. Common property resources are likely to be over-exploited. Cooperation is likely to be more efficient than competition. However, the greater the number of agents accessing the common property, the less likely is the emergence of an efficient cooperative strategy.

Table 4.2 The tragedy of the commons as a gaming problem (benefits of alternative strategies)

		Strategy of herder 2	
		Restricts sheep numbers	*Places large no. on commons*
Strategy of herder 1	*Restricts sheep numbers*	100, 100	20, 130
	Places large no. on commons	130, 20	40, 40

Imperfect Competition

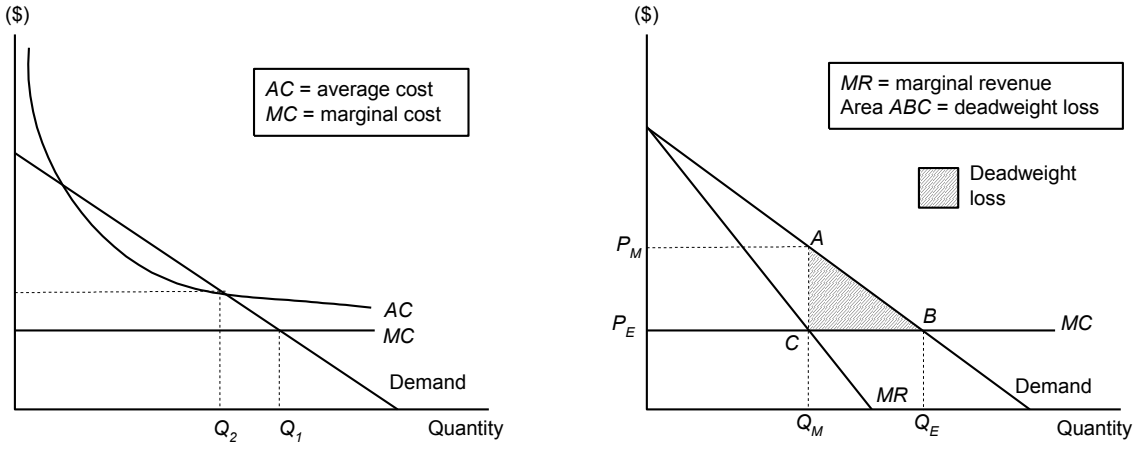
In a competitive market, many firms produce similar goods and all are price takers. Firms charge the market price for their goods and pay the market price for inputs. Each firm faces a horizontal demand curve for their goods and a horizontal supply curve for factors of production. A firm charging an above-market price would obtain no sales. On the other hand, in an imperfectly competitive market, where there are few buyers or suppliers, firms may set prices for their goods above marginal costs and negotiate the prices of inputs.

Imperfectly competitive markets may occur in product or factor markets. The former includes monopolies, oligopolies and monopolistically competitive markets. Imperfect competition occurs for various reasons including economies of scale or scope, differentiated products and competitive advantage due to innovation. When there are economies of scale (due to fixed costs or division of labour), unit costs fall over all or most of the market range of output. Economies of scope occur when one supplier can produce two products more cheaply than can two suppliers. Competition is also limited when firms collude. In all these markets, firms have some control over the price of their products.

Natural monopolies have the greatest potential to control markets and misallocate resources. A natural monopoly exists when a single firm can supply a good to an entire market at a lower cost than can two or more firms. Typically, fixed costs are high and the level of output that minimises long-run average cost is high relative to market demand. Figure 4.2a shows such a market.[1]

Evidently a limited number of large suppliers, even a monopoly, may be more cost-efficient than several smaller suppliers. Moreover, to reward innovation some monopoly profits must be allowed. If market forces lead to the formation of large firms because they are more efficient than small firms, this should not necessarily be discouraged.

However, when firms can set prices, prices are usually set above marginal cost and the quantity of goods supplied is inefficiently low.



(a) Natural monopoly

(b) Inefficiency of monopoly (Q_M supplied instead of Q_E)

Figure 4.2 Monopoly markets

[1] Strictly, the necessary condition for natural monopoly is cost sub-additivity, not falling long-run average cost. Cost sub-additivity occurs when the costs incurred by one firm at a given level of output are lower than the sum of the costs of a set of competing firms whose combined output is the same as that of the single firm.

As shown in Figure 4.2b, a feature of this market is that, for any quantity supplied, marginal revenue from sale of an extra unit is generally less than the price received. Unless the firm can practise price discrimination, it must lower the price it charges for at least some other units. Thus the marginal revenue curve sits below the downward-sloping demand curve.[2] A firm maximises its profit by supplying output up to the point at which marginal revenue equals marginal cost. In this market, it sets the price at P_M rather than at the efficient level (P_E), where price equals marginal cost.[3] Instead of supplying the efficient amount of output (Q_E), the firm supplies only Q_M. The DWL is given by the area ABC.

This market power is inconsistent with the condition for overall product mix efficiency. This condition requires that the marginal rate at which two goods can be transformed into each other (MRT) is equal to the marginal rate at which consumers wish to substitute one good for the other (MRS). However, MRT reflects the marginal cost of production while the MRS reflects the relative prices of two goods. When the price of a good exceeds the marginal cost of production, the MRT does not equal the MRS.

Market power occurs in factor markets as well as in product markets. For example, in resource markets a few firms may control the supply of mineral or energy resources or land for urban development. In labour markets, unions may control the conditions under which labour is supplied. Professionals such as doctors or lawyers may combine to restrict entry into their profession and limit the supply of services. Any restriction on the supply of a factor of production tends to raise its price above its marginal opportunity cost and to result in an inefficient under-supply of goods or services.

Information Failures

Among other conditions, economic efficiency requires a complete set of markets, including markets for future goods and services, and well-informed economic agents.[4] Completeness and information are related. Markets are incomplete when certain goods or services cannot be traded because there is no organised market (Black, 2002). This often reflects poor information. Second, either a buyer or seller may have information that is not available to the other party. This is known as **asymmetric information** and tends to create inefficient outcomes. A third set of issues arises when individuals fail to act in their own best interests because of poor information, perverse preferences or irrational behaviour. In such cases, the assumption that each consumer is the best judge of his or her own interests (consumer sovereignty) is challenged. In this section we discuss incomplete markets, asymmetric information and consumer sovereignty.

Asymmetric information
Occurs when a seller or buyer has information that is not available to the other party

Before examining these issues, we should note a related issue, namely that markets tend to under-supply knowledge. Knowledge has strong positive externalities and is often non-excludable or non-rival. Firms under-invest in research when the benefits cannot be fully appropriated. And, once produced, knowledge is non-rival in that consumption by one agent does not reduce the amount of knowledge available to any other agent. Thus, knowledge is often viewed as a public good.

---

[2] Given a linear demand curve, marginal revenue $MR = P(1 + 1/\eta_d)$ where η_d is the price elasticity of demand and is negative.

[3] With a linear demand curve, a monopolist maximises profits with a percentage mark-up on marginal cost equal to $-1/\eta_d$ where η_d is the (negative) demand price elasticity. Thus the mark-up is inversely related to the price elasticity.

[4] Advanced texts, for example Jha (1998), show the conditions required for efficiency in future as well as present markets.

Incomplete markets

Two main kinds of markets are sometimes said to be incomplete: capital markets and futures or insurance markets. Cited examples of capital market failures are small businesses that cannot raise capital at commercial rates for product development, women who cannot borrow for house purchases and students who cannot borrow for tertiary education against their future earnings. For future goods, markets are required in which the supply of goods is contingent on the state of the world. For example, the ticket price for an open-air concert may be refundable if it rains. Ideally securities would be available for every possible future state of nature. Although markets provide prices for a large range of goods, including many future goods and contingencies such as rain for many events, there are perhaps inevitably many states of nature for which securities are not available.

However, a lack of market transactions is not in itself evidence of market failure. For a market for a service to exist there need to be informed buyers who are willing to pay the marginal cost of supply. Without evidence of consumer demand and costs, it cannot be concluded that a market ought to exist and is incomplete.

Certainly, a lack of information that should readily be available may constitute a market failure. However, markets are inevitably constrained by uncertainty. Insurers cannot assess future risks for all individuals. Other factors that are not necessarily market failures may also contribute to incompleteness. For example, transaction and enforcement costs may limit the reach of capital markets for small loan transactions with a high risk and limit insurance contracts for future contingencies. Thus, incomplete markets do not necessarily indicate market failures.

Markets with asymmetric information

Asymmetric information occurs in many markets. Sellers often have information about their products that buyers do not have. Unless firms inform consumers accurately about the characteristics of their products, consumers may purchase goods that they do not want. Food products may contain unhealthy ingredients, drugs may not deliver improved health, motor vehicles may be unsafe and so on. In a classic article Akerloff (1970) showed that this asymmetry of information has two sets of costs. Not only are too many poor products consumed, but also good products are driven out of the market. Taking the second-hand car market as his main example, he showed that where buyers cannot tell the difference in quality between a 'lemon' and a good car, they will offer (at best) an average price for a second-hand car. This means that the owners of good cars cannot get true value and there are fewer sales of good cars than would occur in an informed market.

In competitive markets, firms have an incentive to protect their reputation with branding, warranties for their products and other strategies. Nevertheless, without regulations that protect consumers, consumers may be poorly informed, or worse misinformed, and make unwise purchases. Unregulated markets may produce poor or unsafe goods.

However, firms are not always more informed than consumers, especially when the conditions for adverse selection or moral hazard exist. Both phenomena lead to an under supply of some goods, notably of insurance. **Adverse selection** occurs when a firm offers an insurance policy to the market and collects consumers who are most likely to benefit from it. Consumers know more about their risks (e.g. health status) than do insurers, but hide this information because this may increase the price of insurance. If insurers cannot distinguish between low- and high-risk individuals, they provide insurance at one price to all. They provide insurance at unprofitable rates to high-risk individuals and fail to provide insurance at an actuarially attractive rate to low-risk individuals. Consequently, insurance markets offer too few services to low-risk individuals.

Moral hazard occurs when a contract promises people payments on certain conditions and people change their conduct in ways that make those conditions more likely to occur. For

example, house and contents insurance may reduce the care that a homeowner takes to prevent burglary or bush fire damages. Insurers must allow for this increased risk even though some people would be willing to take extra care and pay lower premiums. Insurance discourages an efficient amount of risk-avoidance behaviour and insurers provide too few services to careful individuals.

Asymmetric information also occurs in factor markets. Misinformation and inefficient use of resources occur when firms fail to inform capital markets about the true state of their business. Labour resources are employed inefficiently when workers are unaware of the health risks of the work environment and not compensated appropriately. Major issues of adverse selection and moral hazard arise with workers' compensation for workplace injury. Because insurers cannot readily tell the state and cause of an injury and because the availability of income compensation for injury may change a worker's behaviour, markets rarely provide fully efficient insurance for workplace injury.

Consumer sovereignty and behavioural economics

The idea that individuals are the best judge of their own welfare (a concept described as **consumer sovereignty**) is a critical requirement for the view that competitive markets are efficient. As we saw in Chapter 1, economists have long recognised that there may be exceptions to the principle of consumer sovereignty. These exceptions are traditionally described as merit or demerit goods. **Merit goods** are goods that government considers individuals should consume even if they do not demand them. Examples include compulsory education, safety belts and some major cultural goods such as built heritage. Government may consider that it should provide these goods or compel individuals to consume them. On the other hand, **demerit goods** are goods that the government considers individuals should consume in smaller amounts than they do or possibly not consume at all. Well-known examples are alcohol, tobacco, drugs and gambling.

In recent years there has been a surge of interest and research into these ideas under the label of behavioural economics. As Diamond and Vartiainen (2007) point out, there is accumulating evidence that the standard model of rational and informed consumer decision making is inadequate because it fails to account fully for perverse preferences, notably for incomplete self-control and for bounded rationality. Incomplete self-control occurs when agents make decisions that are contrary to their own best interests. This occurs most prominently with various addictions but also with under-saving and procrastination. Bounded rationality reflects poor information processing ability. There are many features of bounded rationality including use of overly simple decision rules, failure to understand probabilities, over-confidence, the use of irrelevant information and loss aversion. As Kahnemann and Tversky (1979) famously pointed out, with uncertain prospects and especially when probabilities are low, individuals often fail to maximise expected utility.

Technically, incomplete self-control and bounded rationality are not market failures in the sense that the market creates, or is responsible, for these problems. However, in so far as markets do not produce welfare enhancing outcomes, government has a potential role in responding to these issues. In such cases, it may be argued that government knows better than the individual what is good for him or her. This argument may be used to promote compulsory education, to require members of society to save for retirement, to prevent households living in flood-prone areas or to subsidise high arts such as opera and ballet. It may also be used to discourage demerit goods such as drugs and gambling. These policies may be contentious because they over-ride individual liberties. Moreover, if they are applied generally, the benefits of correcting the behaviours of the target group need to be weighed against any costs to the non-target group. However, as we will see in Chapter 15 and elsewhere, many public policies reflect merit or demerit good assumptions.

Other Market Failures

As we have seen, public goods, externalities, imperfect competition and information failures are potential major causes of market failure. Are there other significant market failures? We review below issues relating broadly to transaction costs and the assumption of market equilibrium and the special case of second-best conditions. some possible candidates.

Transaction costs. The efficient operation of markets requires that factors of production are mobile, transaction costs are low and enforcement costs are low. If factors are not mobile, factor prices vary by location and do not equalise for similar levels of productivity. If transaction costs are high, exchange becomes costly and marginal rates of substitution vary by location and opportunity. If contract enforcement costs are high, property rights are devalued and exchange is diminished. All this is true, but it does not necessarily make markets inefficient compared with other means of resource allocation. Zero mobility, transaction and enforcement costs are unrealistic. As a practical matter, markets can be described as inefficient only if mobility, transaction and enforcement costs are significantly higher than they should be given the state of technology in the economy.

Equilibrium and disequilibrium. Efficient markets also require that markets are in a state of equilibrium, where all willing buyers and sellers can conduct their trades and the number of willing trades are maximised. There are no excess unsold stocks and there is no unfulfilled demand with long queues for a good.

In practice, markets are often in disequilibrium, with excess demand or supply because the price fails to clear the market. Thus, the labour market is sometimes described as being in disequilibrium because wages do not fall to equate the demand and supply of labour, so that there is unemployment. A market may also be described as in disequilibrium if the current price diverges markedly from the expected long-run equilibrium price. For example, the current price of housing may be considered high relative to some modelled equilibrium price because of speculative demand or slow supply response to a real increase in demand. Indeed, due to speculative demand, unconstrained lending and borrowing and deceitful securitisation practices in the United States, house prices around 2005 to 2008 rose wildly out of line with real long-term values and were a major cause of the global financial crisis (GFC).

Similar disequilibria can occur in commodity markets. A rise in demand for a commodity may cause prices to rise and farmers to respond with such an increase in output that prices then fall below the initial price! But this lower price is not sustainable. At this low price, output is now cut back so hard that the price shoots even higher than in the initial case! When supply curves are more elastic than demand curves, and firms have unrealistic expectations, a shock to the system can lead to unstable outcomes rather than a return to equilibrium. These kinds of market reactions are known as the Cobweb Theorem.

Disequilibria result from both imperfect foresights of buyers and sellers and lags in responses to exogenous events. Current prices then fail to reflect underlying demand and supply conditions and there is excess demand or supply along with associated deadweight costs. Arguably these failures may be described as information failures. However they are labelled, recent experience indicates that such disequilibria are not always self-correcting, at least in the short term.

In conclusion, immobility of labour, transaction and enforcement costs, faulty expectations and slow or excessive responses to exogenous events may cause markets to work less efficiently than is desirable. They may also contribute to incomplete markets.

To some extent these failures may be viewed as information failures of markets rather than as separate issues. Moreover, markets do provide incentives for economic agents to reduce

transaction and enforcement costs, to invest in profitable transactions and to correct over-reactions. Nevertheless, the global financial crisis provided strong lessons about the frailty of unregulated markets to self-correct mistakes and the need for government regulation.

The special case of second best

As we saw in Chapter 3, a Pareto-efficient economy is often described as a 'first best' economy and the efficiency conditions required in each industry for achievement of such an economy are 'first-best' conditions. It would be natural therefore to suppose that, the closer that an economy is to meeting these conditions, the more efficient it will be. Conversely, the more market failures there are, the greater would be the inefficiency and the deadweight loss. But this is not necessarily the case. Suppose that smoking creates negative externalities and that oligopolistic tobacco firms exploit their market power (as they do) to sell cigarettes at prices well above marginal cost. Tobacco consumption and the consequential negative externalities will be less than would occur in a more competitive industry. In such cases one market failure may partly offset another.

Second-Best Theorem
When there is more than one market distortion, correcting one distortion may make the allocation of resources more inefficient

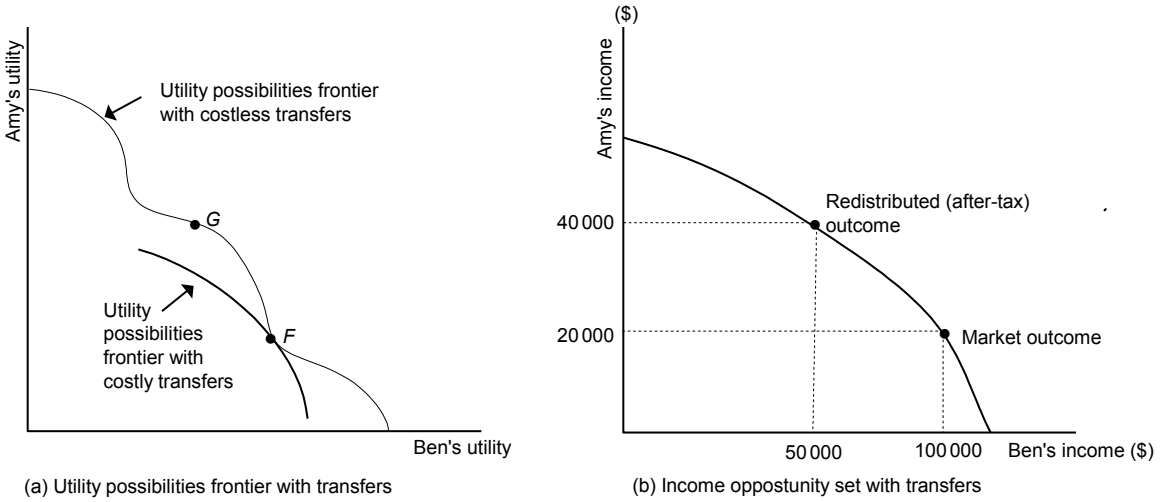
The Second-Best Theorem (Lipsey and Lancaster, 1957) states that, in the presence of any departure from the first-best conditions, implementing first-best conditions in all other markets is not necessarily efficient. The most efficient outcome may be achieved by a departure from the first-best conditions in another market. Suppose that a tax is imposed on a good such as beer which creates a wedge between the price of the good and its marginal cost and hence a deadweight loss. Then a tax on a substitute good, say wine, which increases the consumption of beer, may have a net positive economic impact rather than a deadweight loss. Given the prevalence of market failures, and indeed the departures from competitive conditions in many publicly controlled sectors, this theorem has important implications for public policy (see further discussion in Chapter 17). While it is generally sensible to aim to achieve the first-best conditions, it is often necessary rather to determine the second-best policy. This is the set of departures from perfectly competitive conditions that collectively results in the least misallocation of resources.

Equity Issues

Of fundamental concern, efficient markets often produce highly unequal outcomes. In competitive labour markets workers are paid the value of their marginal product. However, this product and the related wage depend individual productivity. Such a wage may be inadequate to support an individual or household according to the norms of the society. Moreover, productivity may be a function of numerous factors, including genetics, upbringing and social context, which have no immediate connection to market conditions.

As discussed in Chapter 3, in principle the Second Welfare Theorem shows that any desired distribution of outcomes can be achieved given an appropriate initial distribution of endowments brought about by individualised lump sum transfers. This is a tax that is not affected by behaviour. Given this new set of endowments, a perfectly competitive economy with no market failures would then produce efficient and equitable outcomes.

However, the Second Welfare Theorem has little practical relevance. Ability to earn is not itself readily observable. Thus, it is impossible to identify the economic capacity of individuals without observing behaviour. But once taxes are levied on behaviour, behaviour may change. Suppose that half the members of a community can produce output valued at $100 000 a year and the other half can produce output worth only $20 000. If government knew ex-ante the income earning ability of persons in each group, it could achieve equality by a lump transfer of $40 000 from each member of the first group to each member of the second group. However, if government redistributes income after it is earned, the tax depends on income and everyone has an incentive to join the low-output group.



(a) Utility possibilities frontier with transfers

(b) Income oppostunity set with transfers

Figure 4.3 Opportunity sets in utility and income space

Figure 4.3a shows a utility possibilities frontier (UPF) for Amy and Ben. If transfers are costless, any point on the UPF can be obtained. Suppose that point *F* on the UPF represents the initial distribution of welfare, then any redistribution of income away from the competitive market output is likely to reduce the quantity of value of output. This restricts the actual social choice to any point along the feasible UPF. The feasible frontier shows the actual set of utility outcomes between which society can choose, given the costs of redistributing resources.

The feasible outcomes can be drawn more practically in income space. Figure 4.3b shows an income opportunity set—how income might be distributed between Amy and Ben. In this example, the market provides Amy and Ben with a total income of $120 000, of which Ben receives $100 000 and Amy gets $20 000. If government transfers $20 000 from Ben to Amy, Ben obtains $50 000 (after reducing his work hours) and Amy gets $40 000 (assuming the transfer does not change Amy's work hours). Total income falls from $120 000 to $90 000.

Given that market incomes are almost always inequitable, government must aim to distribute the benefits of economic activity fairly in accordance with some ethical view of welfare (which may be miserly or generous). If individualised lump sum taxes were possible, the outcome could be efficient and fair. Because they are not feasible, governments must use a combination of distortionary taxes and social grants to achieve distributional objectives. Because taxes and grants generally change behaviour, redistribution almost always reduces the income available for redistribution. The trade-off between efficient and socially fair allocations of resources is at the heart of many public policy decisions.

Government Functions

Each of the major economic functions of government (establishing a legal infrastructure, addressing market failures, providing social welfare and macroeconomic policy) represents a substantial task. Government would be responsible for establishing and protecting property rights and the commercial system of contracts and exchange. It would provide a wide range of public goods, develop policies for externalities and common resources, regulate uncompetitive markets, develop information and regulate misinformation. Also, it may be responsible for a substantial redistribution of income.

However, conclusions about the roles of markets and government should be drawn cautiously for two main reasons. First, the efficiency of markets does not depend on complete fulfilment of the first-best conditions in a static model. The real world is too complex to be fully represented by a mathematical model, even such a complex model as the Arrow–Debreu general equilibrium model which underlies the First and Second Welfare Theorems. In practice, economists often rely on the concept of workable rather than perfect competition. Workable competition is the practical amount of competition between firms required for efficient exploitation of product and process technologies. Admittedly this is a loose definition, but so long as markets are generally competitive and new entrants to markets are feasible, competition encourages firms to produce what people want with efficient low-cost methods. Also, as discussed in the next chapter, competitive economies are usually more innovative and generally produce higher rates of economic growth than less competitive economies.

Second, market failures indicate only a *prima facie* role for government. Government should determine how resources are used only if it can do so better than can markets. There can be no general presumption that this is the case. An advantage of markets is that prices convey a great deal of information about the value of the goods to consumers. The idea that government is better informed about the needs of individuals than are private firms or individuals themselves, and should therefore over-ride market transactions, should be treated with caution. In the words of Pigou:

> The case (for intervention), however, cannot become more than a *prima facie* one until we have considered the qualifications which government may be expected to possess for intervening advantageously. It is not sufficient to contrast the imperfect adjustments of unfettered private enterprise with the best adjustment that economists in their studies can imagine. For we cannot expect that any public authority will attain, or will even wholeheartedly seek, that ideal. Such authorities are liable alike to ignorance, to sectional pressures and to personal corruption by private interests. A loud voiced part of their constituents, if organised for votes, may easily outweigh the whole.[5]

Summarising government functions. Table 4.3 overleaf summarises the main resource allocation and social welfare functions of government—along with some examples. In addition, activities are classified as basic, intermediate and activist government functions. The economic and social arguments for undertaking the basic and intermediate functions are generally strong. However, there is much discussion about where precisely the lines should be drawn, especially for the more activist activities. It should also be noted that some goods are provided for both efficiency and equity reasons. Thus health and education services appear under both addressing market failure and improving equity functions.

Government Instruments

So far we have focused mainly on *when* government should intervene in the economy. We now briefly discuss *how* government should intervene. Government can carry out policies in six main ways by:

1. Low intervention methods: creating markets, facilitation and provision of information.
2. Regulating markets: by establishing the main rules for markets and in some cases by regulating the quantities that firms should produce or the prices they may charge.

---

[5] Pigou (1920) p. 332.

Table 4.3 Functions of government with examples

Functions	*Creating markets and addressing market failure*				*Social welfare*
Basic	*Providing pure public goods*:				*Protecting the poor:*
	Defence		Law and order		Poverty relief programs
	Provision of property rights		Land and property surveys		Disaster relief
	Provision of currency		Macroeconomic management		Alleviating major inequities
	Primary education		Public health		
	Basic economic infrastructure		Transport, power, water		
	Basic communications		Flood management		
Intermediate	*Providing other public goods:*	*Addressing externalities:*	*Regulating market power:*	*Overcoming imperfect information:*	*Providing social assistance and equity:*
	Secondary and tertiary education	Environmental protection	Competition policy	Consumer protection	Unemployment benefits
	Health services	Regulating common property resources	Utility regulation	Financial regulation	Family allowances
	Other transport and communication services			Occupational health and safety	Pensions for retirees
					Progressive taxation
					Health services
					Education services
Activist	*Coordinating private activity:*				*Redistribution:*
	Fostering markets		Supporting industries or firms		Wealth redistribution
	Providing merit goods				

Source: Adapted from World Bank (1997).

3. Influencing markets with fiscal instruments: employing taxes to raise prices and discourage consumption or subsidies to do the reverse.
4. Funding the provision of goods and services by private firms.
5. Producing goods, either by government agencies or through public trading enterprises.
6. Redistribution: income transfers or transfers-in-kind.

Table 4.4 shows these general strategies and various other instruments available to government. It also distinguishes between low, medium and high-intervention instruments. Low intervention includes market creation where no market exists, moral exhortation, for example, to encourage public health practices and facilitation of development projects. More active intervention includes financial incentives to firms or households and payments for goods provided by private firms. High intervention instruments include binding regulations on firms, direct provision of services and redistributive activities.

Critically, the effectiveness of government policies depends not only on the choice of policy *but also on how policy is implemented*. In this book there are many examples of the importance of instrument choice, for instance in the discussions of environmental policy, the delivery of publicly funded services, the provision of education, transport and health services and the delivery of income transfers and social assistance policies.

Table 4.4 Main instruments of government

General approach	*Possible Instrument*	*Comments*
Low intervention methods		
Do nothing		This is often an option Allows markets to determine outcome
Create markets	Auctions for common resources	Has environmental applications
Provision of information and moral exhortation	Public announcements	For example information on public health Exhortation to conserve water
Facilitation	Non-financial assistance to industry	Assistance with project development
Medium intervention methods		
Provision of incentives to markets or individuals	Financial instruments (taxes or subsidies)	May encourage or discourage activities For example, subsidies for health insurance Has environmental applications
Provision of public funding for services	Public payment for services provided by private sector	For example, payments for doctors May include user payments
High intervention methods		
Regulations of firms and households	Legislation Subordinate regulations	Includes general administration under statutes
Provision of services through public trading enterprises	Public production of goods with user charges	PTEs have some operating independence from government, but not policy independence May include some subsidies to consumers
General government supply of services	Public funding and production of services	Usually essential services May include user payments
Redistributive activities	Taxation and cash benefits Provision of benefits in kind	Income transfers are a high proportion of GDP Benefits may include food, housing, public health services
Institute fundamental change	Constitutional changes	Rare in Australia

Summary

- The main market failures are public goods, externalities, imperfect competition and information failures.
- Public goods are typically characterised by large positive externalities. Formally they are non-excludable and non-rival goods. Private firms under-supply non-excludable goods and over-charge for non-rival goods.
- Markets can produce negative externalities, especially environmental costs.
- Private firms are also likely to over-exploit common property resources.
- In imperfectly competitive product markets, firms tend to charge prices above marginal cost and to supply too few goods. Uncompetitive factor markets may also lead to an under-supply of goods and services.
- Many economic agents are poorly informed. There is over-consumption of poor products and failures in insurance markets. The global financial crisis was an extreme manifestation of information failures.
- Also markets, including efficient ones, can produce large income inequalities.
- The allocative role of government is to establish the conditions for efficient markets, provide public goods and develop policies for externalities, imperfectly competitive markets and poor information.
- However, government should act only if it can improve on market outcomes. This depends on the extent of market failure and the efficiency or otherwise of government.
- Government also has a major responsibility for income distribution and for social welfare generally.
- However, income redistribution in any form nearly always changes behaviour and causes some inefficient outcomes. Hence there is invariably a policy trade-off between efficiency and provision of social welfare grants and programs.
- Policy instruments include regulating markets, use of taxes and subsidies, funding the private supply of goods, public production of goods and income redistribution.
- Government effectiveness depends on the choice of instrument as well as on choice of policy.

Questions

1. Design a two-person gaming model as in Table 4.2 in which the strategies are to steal or not to steal. Show how both parties can gain from rules protecting private property and that laws against stealing are a public good.
2. Major sporting events like the Olympic Games often have physical and pecuniary external effects. Give examples of both kinds of externality and show why physical externalities may have efficiency and equity effects whereas pecuniary externalities have only equity effects.
3. Figure 4.2 shows the misallocation of resources with a monopoly supplier. Is a similar misallocation likely to occur in an oligopolistic market or with monopolistic competition?
4. Suppose that a monopolist's total cost is $C = 12 + 6Q$ and that the demand curve is given by $Q^d = 25 - P$.
   i. What is the price set and quantity sold by a profit-maximising monopolist?
   ii. What price should be set, and what quantity sold, to maximise the benefits to society?
   iii. What is the deadweight loss in this example?
5. Suppose a vaccination program reduces the spread of infectious disease and contains a positive externality equivalent to $50 per vaccination. The marginal cost of supplying vaccinations is constant at $60. The annual demand for vaccinations in the market is given by $Q^d = 1\,000\,000 - 10\,000\,P$. Determine the following:
   i. The number of vaccinations that would be provided by a competitive market.
   ii. The socially efficient number of vaccinations given the positive externality
6. Is the Cobweb Theorem of market prices and quantities an example of market failure?
7. Does heavy smoking exhibit both lack of self-control and bounded rationality?
8. What are incomplete markets? Can they be described as market failures?
9. Are insurance markets likely to provide inefficient insurance for unemployment or disability?
10. Do university student services have public good characteristics? Can compulsory student fees to fund these services be justified?

11. Why is an unequal distribution of income not necessarily described as a market failure?
12. Why does a failure to maximise expected utility reflect bounded rationality?
13. What economic costs occur when government redistributes income? Is it possible to redistribute income without economic cost?
14. Give an example of a second-best policy with an explanation.

Further Reading

Akerloff, G. (1970) 'The market for lemons: Quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, 80, 488–500.

Arrow, K.J. (1963) 'Uncertainty and the welfare economics of medical care', *American Economic Review*, 53, 942–73.

Coase, R.H. (1974) 'The lighthouse in economics', *Journal of Law and Economics*, October, 357–376.

Dixit, A. (2009) 'Governance institutions and economic activity', *American Economic Review*, 99, 5–24.

Hardin, G. (1968) 'The tragedy of the commons', *Science*, 162.

Kahneman, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decisions under risk', *Econometrica*, 47, 263–291.

Lipsey, R. and Lancaster, K. (1957) 'The general theorem of second best', *Review of Economic Studies*, 24, 11–32.

Chapter

5

Economic Growth and Government

Understanding the causes of the wealth and poverty of nations is 'the grand object of all enquiries in Political Economy'.

Thomas Malthus, letter to David Ricardo

The Causes of Economic Growth ♦ Optimal and Efficient Economic Growth ♦ Markets and Economic Growth ♦ Government and Economic Growth

Globally, there are great differences in living standards. What causes these differences? Why have North American economies grown faster than the European ones or East Asian economies grown faster than Latin American ones? Are resource endowments, markets or governments responsible? Answering such questions has long been a central concern of economics. And, as we will see, government has a major role to play.

In this chapter we examine the causes of economic growth and the policy implications. Of course, economic growth depends on efficient use of resources, as discussed in Chapter 3. However, that discussion took technology and the supply of capital and labour as given in any period. Formally, it assumed a given production possibility frontier. Economic growth is about expanding output principally by improving productivity in various ways, including technological advances and increasing capital inputs.

We should note that the welfare objective is usually to increase GDP per capita, not simply to increase GDP as population grows. Some other important caveats should be noted. High growth rates may not enhance overall welfare if it entails large sacrifices in current standards of living, significant inequalities in society or long-term degradation of the environment.

The chapter starts by examining the main causes of economic growth. The second section discusses optimal and efficient rates of economic growth. The third and fourth sections examine the roles of markets and government respectively in creating economic growth.

The Causes of Economic Growth

Economic output depends on the supply of factors of production and their productivity. An aggregate production function links output to inputs. Thus, a country's aggregate production function can be expressed generally as:

$$Y = A f(K, L, N) \tag{5.1}$$

where Y is the total output of the economy, K is the stock of physical capital, L is labour units, N is natural resources and A is a measure of productivity or technology (also known as multi-

factor productivity). As technology improves, the value of A increases and more output is produced from a given combination of inputs.

In this general formulation, capital and labour are treated as standard (homogeneous) units. Technical innovation is a catch-all variable that allows both for improvements in the ways in which inputs are used and for improvements in labour skills (human capital) and capital productivity. However, labour can be estimated in skilled labour unit equivalents, which allows for the acquisition of workforce skills by investment in human capital.[1]

Equation 5.1 relates the level of output to the levels of inputs and technology. Economic growth depends on changes in inputs and technology. Growth models focus especially on changes in capital, labour and technology. Natural resources are generally regarded as fixed or, alternatively, as a function of capital investment and technology. The quantum of resources is also very hard to measure. Accordingly, natural resources are omitted from most growth models.

In a competitive economy in which factors of production are paid the value of their marginal product, Equation 5.1 can be transformed into a specific growth accounting equation.

$$\Delta Y/Y = [(1-\theta) \times \Delta L/L] + (\theta \times \Delta K/K) + \Delta A/A \qquad (5.2)$$

where Δ represents the change in the respective variable, thus $\Delta Y/Y$ is the change in output, $\Delta L/L$ is the change in labour and so on. $(1-\theta)$ and θ are weights equal to the proportion of total output (GDP) received by labour and owners of capital respectively. To see how this works, suppose that labour and capital's share of GDP are 0.75 and 0.25 respectively and that, over a year, labour grows by 1.0 per cent, capital by 2.0 per cent and productivity by 1.5 per cent. It follows that GDP would grow by $(0.75 \times 1.0) + (0.25 \times 2.0) + 1.5 = 2.75$ per cent in a year. Box 5.1 shows a common production function of this nature.

Economists traditionally viewed investment in physical capital as the main determinant of economic growth, especially of growth per capita. Investment was in turn regarded as depending mainly on savings. Changes in technology were regarded as determined exogenously. In so far as technical progress was explained, it was embodied in new investment and depended also on savings.[2]

Box 5.1 The generalised Cobb–Douglas production function

The generalised Cobb–Douglas production function is a practical and commonly used aggregate production function. This is expressed as a power function:

$$Y_t = Ae^{rt}K_t^{a_1}L_t^{a_2} \qquad (5.3)$$

where Y, K and L are as described in the text, e equals 2.718, A, r, a_1 and a_2 are constants, and t refers to the year. Given data for Y, K and L, the values of the constants can be estimated.

Assuming constant returns to scale, which is often observed, the a_i exponents sum to one ($a_1 + a_2 = 1$). This means that if K and L are both multiplied by λ, then output will be multiplied by λ. Typically, a_1 is about 0.25 and a_2 is about 0.75.

This implies diminishing returns to extra inputs of capital or labour, holding the other input and technical progress constant. The effects of diminishing returns are avoided by simultaneously increasing the other input or by technical progress.

Equation 5.3 can be converted to a growth equation as follows:

$$\dot{Y} = r + a_1\dot{K} + a_2\dot{L} \qquad (5.4)$$

where a dot over a variable represents change per unit of time and r is the rate of technical progress. The rate of growth of output is the weighted sum of the rates of input growth plus r. If $r > 0$, growth can occur without any growth in inputs. This is known as disembodied technical progress.

[1] For excellent discussions of growth models, see Jones (2002) or Aghion and Howitt (2009).
[2] This traditional view is sometimes formalised in the neo-classical or Solow–Swan model

Table 5.1 Output growth (%) attributable to capital, labour and productivity, 1960–89

Source of growth	*North America*[a]	*Europe*[b]
Capital stock per capita	39	43
Capital quality	7	6
Labour hours per capita	13	-16
Labour quality	17	10
Productivity	24	57
Total output	100	100

(a) Average for United States and Canada.

(b) Average for France, Germany, Italy and United Kingdom.

Source: Dougherty and Jorgenson (1996).

However, econometric studies have found that changes in capital and labour explain only a small part of economic growth. A large part is explained by the residual element (technical progress). For example, Aghion and Howitt (2009) estimated that changes in total factor productivity accounted for two-thirds of the growth of OECD economies between 1960 and 2000, leaving one-third unexplained.

Some growth accounting models now include estimates of the quality of physical capital and labour as well as the quantity. Table 5.1 shows estimated contributions of the quantity and quality of capital and labour, and residual productivity growth, to the growth of GDP in North America and Europe between 1960 and 1989. According to these estimates, changes in the stocks of capital and labour inclusive of quality changes explained 76 per cent of economic growth in North America but only 43 per cent of growth in the four large European countries (where labour hours fell significantly), with residual productivity explaining the balance in both cases. Almost certainly, in this century, the extraordinary digital technological advances are responsible for a large part of recent economic growth.

Thus, there remain large unexplained productivity changes broadly associated with changes in technology. Recent discussions suggest several possible drivers of productivity growth: a competitive economy open to international trade, investment in human capital through education, government support for innovation, a democratic and inclusive system of government which protects property rights and a culture of trust (sometimes described as social capital). We examine below the roles of these various drivers of economic growth.

Economic growth factors

Natural resources (natural capital). The contribution of natural resources to growth is complex. Some resources are finite and non-renewable and decline with exploitation. However, resources may be discovered or economic ways to exploit previously uneconomic resources may be found. Other resources are renewable and their productivity may depend on how they are managed. Also, environmental assets depend on the management of wastes. While the quantity of wastes usually rises with GDP, management usually improves so that the impact of growth on environmental quality may be negative or positive. Thus, it may not be clear at any point in time whether the effective supply of natural capital is rising or falling. Moreover, if the supply is rising, this is generally the result of capital investment or technical innovation. Partly because of these conceptual and measurement difficulties, natural capital generally gets less attention than other forms of capital in the economic growth literature.

This was not always the case. In the 19th century, economists such as Malthus and Jevons were deeply concerned about the impact of population pressures on scarce natural resources,

notably on agriculture and coal resources. At the start of the 20th century, ten of the 12 largest companies in the United States were resource-based.[3]

However, commodity prices fell through most of the 20th century. The major growth industries such as microelectronics, biotechnology, telecommunications, machine tools and robots, and computers depended more on human capital than on natural resources. To-day, the top seven most highly valued stocks on the US stock market are technology companies.

However, in this century the emergence of China and India as economic powers, with relatively few natural resources for their size of populations, has greatly increased the demand for resources and commodity prices rose to hundred-year highs. In Australia, investment in energy and mineral resources became major sources of economic growth.

On the other hand, soil degradation and shortage of water are constraints on growth. Globally there are deep concerns about the impacts of climate warming on resources. Overall, resources are likely to continue to be major factors in economic growth. Protection of natural capital will enhance long-run economic growth.

Physical capital. This has four main components: private fixed capital (plant, equipment and commercial property), residential investment, inventories and public capital investment. Of these, private fixed investment and public capital investment, including many elements of economic infrastructure, are the most important determinants of economic growth. The latter is of course under government control. What are the main drivers of private fixed capital?

Private investment was explained traditionally by a country's propensity to save. East Asian countries were cited as examples of cultures that encouraged saving and hence investment. But investment depends on domestic savings only in a closed economy. Most economies have been open for a long time. Between 1870 and 1929, savings were a far higher proportion of GDP in the UK than in the United States, but investment was higher in the United States, funded largely by British savings. Investment opportunities were greater in the United States because scale effects encouraged the manufacture of specialised machinery, standardised goods and interchangeable parts, and the low price of materials relative to labour encouraged the use of machinery.

Today, there is a world capital market. Foreign capital may cost more than local capital to allow for regulation and exchange rate risks. Thus, local savings are still important, but they are not a major constraint on development. Investment depends on the relationship between the marginal return on capital and the borrowing rate. Investors must be able to obtain an acceptable after-tax return on capital, with due allowance for risks. Government factors that influence the return on capital include provision of reliable economic infrastructure, access to markets, competitive tax rates, security of property and freedom from regulatory risk. Government may also encourage private investment by subsidies for research and development (R&D) or by tax breaks for capital investment. However, the Productivity Commission (2007) found R&D subsidies create limited additional investment.

Labour and human capital. As we have observed, an increase in labour may increase GDP but result in a fall in GDP per capita. On the other hand, an increase in foreign labour may raise average GDP per capita but reduce GDP per capita for existing residents. A critical factor here, though not the only factor, is the role of human capital.

Human capital refers to "the stock of knowledge, skills and abilities that determine the labour productivity of an individual", (Black et al., 2013). This is significantly determined by the education and training in society. As discussed in Chapter 12, many studies find high returns to investment in education and training, with an extra year of schooling raising wages by about 10 per cent. However, as noted there, these results should be viewed with care as

[3] Thurow (1996) p. 66.

some of the estimated gains may reflect unmeasured ability or non-education inputs. Studies of international rates of economic growth also strongly support the role of education in growth. Barro (1997) estimated that an extra year of male upper-level schooling raised the average growth rate in 100 countries by 1.2 percentage points per year. The OECD (2001a) concluded that, for OECD countries as a whole, "each extra year of full-time education (corresponding to a rise in human capital by about 10 per cent) is associated with an increase in output per capita of about 6 per cent". Hanushek and Woessmann (2010) estimated that, in OECD countries, school improvements falling within observed existing performance levels could raise GDP by between $90 and $275 trillion over 80 years to 2090.

Institutional capital. This refers to the capacity of a country's institutions to create the conditions for economic growth. Two critical conditions are necessary to encourage innovations, entrepreneurial activity and investment in the application of technology.

One is a secure and lawful environment, including established property rights, rule certainty and tax stability. Lawlessness, crime and violence, along with an unpredictable judiciary, are major impediments to economic growth. Drawing on the results of a survey of 3600 firms in 69 countries, the World Bank (1997) concluded that a sound and secure institutional capacity is a major determinant of investment and economic growth. Aghion *et al.* (2008) found a significant correlation between economic growth and democracy. In their very readable book, Acemoglu and Robinson (2013) describe how government institutions, including the rule of law and property rights, dramatically affect economic outcomes across countries.

The second critical condition is competitive and open markets. While we discuss below arguments for and against market competition as an incubator of innovation, we find that, overall, competitive markets are much more innovative than monopolistic or protected markets. Trade openness encourages specialisation, increases market size and opportunities for scale economies, technology transfers and knowledge spillovers. American citizens have long benefited from economies of scale and low prices of the large free trade area of the United States. By contrast the populations of South American countries suffered from the small markets and high prices in their closed economies.

Social capital. Whereas institutional capacity is about rules and systems, social capital is about social norms and relationships and especially about the amount of trust and cooperation in society. Fukuyama (1999) defines social capital as 'an instantiated set of informal values or norms shared among members of a group that permits them to cooperate with one another'. Putnam (2000) defines the concept as 'social networks and the norms of reciprocity and trustworthiness that arise from them'.

Bowles and Gintis (2002) argue social capital reduces transaction costs that impede markets, allows informal institutions for credit and finance, encourages recognition of the social costs of private actions, encourages effective political participation and gives credibility to policy announcements of government. A high level of trust allows governments to reduce the amount of regulations and encourages citizens to invest over longer terms. The role of the mafia in crippling economic development in Southern Italy for centuries is a classic example of the destructive impacts of poor social capital.

Aghion and Howitt (2009, pp. 422–425) provide an excellent summary of the empirical issues in testing these hypotheses and of recent empirical studies. A prime issue is how to measure social capital. This is typically done by measures of club membership or by surveys that report the amount of trust that citizens have in each other. Then, if economic growth is correlated with this measure of social capital, there is the issue of causality. Does trust create economic growth or does growth create trust? The econometric response to this is to try to find an instrument that is correlated with trust, for example genetic resemblance, that cannot be a function of economic growth. OECD (2001a) reviewed several studies and found that

social capital tends to reduce poverty and anti-social behaviour, improve health and increase longevity. Aghion and Howitt (2009) also cite several studies showing a relationship between economic growth and trust in the society, but few have shown the nature of the causality. Guiso *et al.* (2005) is cited as one of the few that suggest trust is a positive factor for growth.

If social capital is important, what determines the amount of it in a society? History and culture are important. So is education. Secondary education allows people to participate fully in society, increases employability and reduces anti-social behaviour (see Chapter 12). As Glaeser *et al.* (2001) observed, many individuals invest in membership of social networks because they see the economic as well as social benefits.

Government also has a major role. An inclusive and compassionate government can significantly enhance social capital. On the other hand, social capital declines with social polarisation, discrimination, inequality, an absence of social security and arbitrary government actions.

Optimal and Efficient Economic Growth

The optimal rate of economic growth is the rate that maximises the welfare of citizens over time. This is rarely the highest achievable growth rate. The Soviet Union under Stalin in the 1930s and the People's Republic of China under Mao Tse-Tung in the 1950s and 1960s tried to increase economic growth by shifting resources into capital formation and away from rural production. Millions starved in the process.

Although productivity improvements may provide some painless economic growth, growth in future consumption generally requires some sacrifice of present consumption (unless capital is wholly externally financed). Economic growth may also be maximised in a low tax and benefit environment that rewards the successful but leaves behind the less productive. And it may be maximised in the short run by ruthless exploitation of natural resources. To determine an optimal rate of economic growth we need a social welfare function that weights consumption across generations.

We focus here on the more modest objective of achieving an efficient rate of economic growth. Drawing on the concept of Pareto efficiency, the rate of economic growth is efficient when resources cannot be reallocated across time periods to make one person better off without making someone else worse off. The analysis of efficiency over time is similar to the analysis of efficiency at a point in time. In both cases, the necessary conditions for overall efficiency are: efficient consumption, efficient production, and product mix efficiency.

Consumption efficiency. In Figure 5.1a overleaf the line *AB* shows the consumption available to an individual, Amy, in two periods. The slope of *AB* is the marginal rate at which consumption in one period can be exchanged for consumption in the next. The slope equals $-(1 + r)$ where r is the real return on savings. Suppose that r equals 5 per cent: if Amy forgoes \$100 of consumption in the first period, she can obtain \$105 of consumption in the second one.

The indifference curves (I_i) show combinations of consumption in the two periods between which Amy is indifferent. These curves are usually convex because the lower the level of consumption in any period, the greater the compensation required for a further sacrifice of consumption in that period. The slope of the indifference curve at any point shows the marginal rate at which Amy is willing to exchange consumption now for consumption later.

To maximise her utility, Amy aims for the highest feasible indifference curve. She achieves this by borrowing or lending so that the marginal rate at which she is willing to exchange present for future consumption equals the rate of exchange between present and future consumption in the market (point *S*). At any other point on the *AB* line she could increase her utility by changing her consumption pattern. The same principle applies to all consumers.

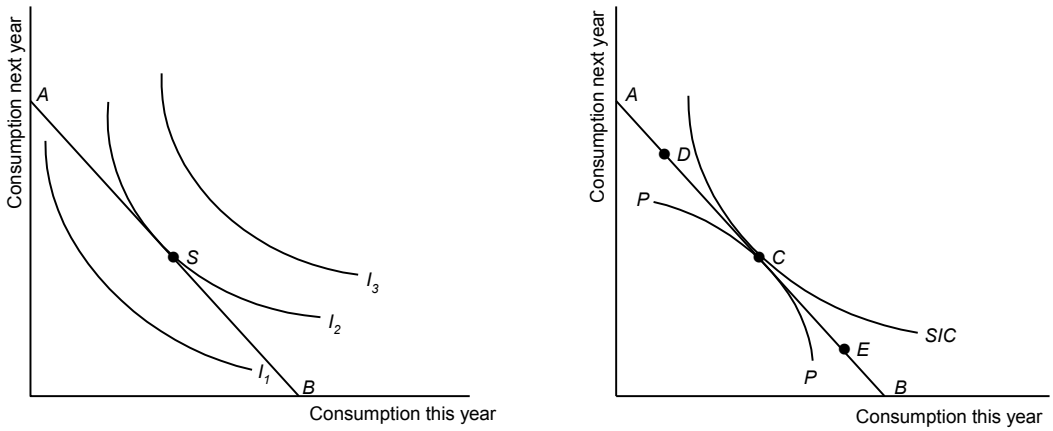


Figure 5.1 Efficient inter-temporal markets

Although individual budget lines vary with income, in a competitive market there is a market-set rate of exchange between present and future consumption and the slope of all such budget lines equals this exchange rate $-(1+r)$. This implies the condition for inter-temporal consumption efficiency—*the marginal rate of substitution between future and present consumption (MRS) of all individuals* $= -(1 + r)$ *and is the same for all individuals*.

Production efficiency. In Figure 5.1b the curve *PP* shows the combinations of consumer goods that an economy can produce in each period. Firms can increase output of consumer goods in period 2 by producing more capital goods and less consumer goods in period 1. The slope of the *PP* curve reflects the marginal rate at which present goods can be transformed into future goods. With declining returns to capital applied to fixed resources, the *PP* curve is concave. Firms invest so long as the return on investment is sufficient to pay for the investment. The marginal rate at which firms use resources to transform current goods into future goods (MRT) must equal the price of transformation, the market rate of interest. Faced with a real interest *rate r, all firms maximise the present value of their profits by choosing a level of output, such that MRT* $= -(1 + r)$.

Overall inter-temporal efficiency requires that the *marginal rate of substitution of future for present consumption must equal the marginal rate of transformation of present into future goods* with both equal to $-(1 + r)$. This is also illustrated in Figure 5.1b. The indifference curve here represents a notional social indifference curve (SIC) between present and future consumption. At the margin, SIC must be tangent to *AB* (because MRS is equal for all individuals). The point *C* is a Pareto-efficient outcome. At *C*, the marginal rate at which present goods can be transformed into future goods equals the marginal rate at which all consumers are willing to exchange future for present consumption. The desired savings by households are just sufficient to finance the desired borrowing by firms. Individuals borrow to equate their MRS between period 1 and 2 to $-(1 + r)$. At point *D*, there would be too little consumption this year; at point *E*, there would be too little consumption next year.

Conclusion. This analysis shows the necessary conditions for efficiency over time. However, it does not consider inter-generational or equity issues more generally or environmental outcomes. Nor does it explain what determines the key drivers of economic growth: capital investment and technical innovation. We also need to discuss whether markets or government produce efficient and equitable economic growth. We turn to these issues below.

Markets and Economic Growth

In this section we examine first how markets establish the quantity of physical capital stock and its price (the rate of interest). We then discuss the drivers of innovation. Finally, we discuss whether markets produce an efficient rate of economic growth.

The quantity and price of physical capital

The amount of capital stock in a country is a function of the demand for, and supply of, capital. Investment, a flow variable, refers to the change in the capital stock. The rate of interest is the price of capital.

Figure 5.2a shows how the quantity and price of capital (the rate of interest payable for capital) are determined in a closed economy. The demand for capital curve slopes downward because firms can profitably employ more capital when the price is low. The locus of the curve depends on profit opportunities and technology. The supply of capital curve represents the savings decisions of households. At higher interest rates, each dollar of consumption deferred yields more consumption in the following period. Thus, households are usually willing to save more at higher interest rates.[4] The locus of the supply curve is a function, *inter alia*, of household incomes. In period 1, the demand for capital and the supply of savings are in equilibrium at quantity K_1 and interest rate r_1. If the supply of saving increases in period 2 (with no shift in demand), there is a new equilibrium quantity of capital stock K_2 at a lower rate of interest r_2. Capital investment is the difference between K_2 and K_1 plus any investment required to offset depreciation of K_1 stock.

Figure 5.2b includes international savings. For small open economies, such as Australia, an elastic supply of international funds is available at the international lending rate allowing for local exchange rate and other risks (r_F). Thus, the supply of international capital lowers the rate of interest from r_D to r_F. The total capital stock rises from K_{D1} to K_F. However, capital stock financed from local savings falls from K_{D1} to K_{D2}. The fall in the price of capital provides a benefit to firms equal to area $r_D ADr_F$.

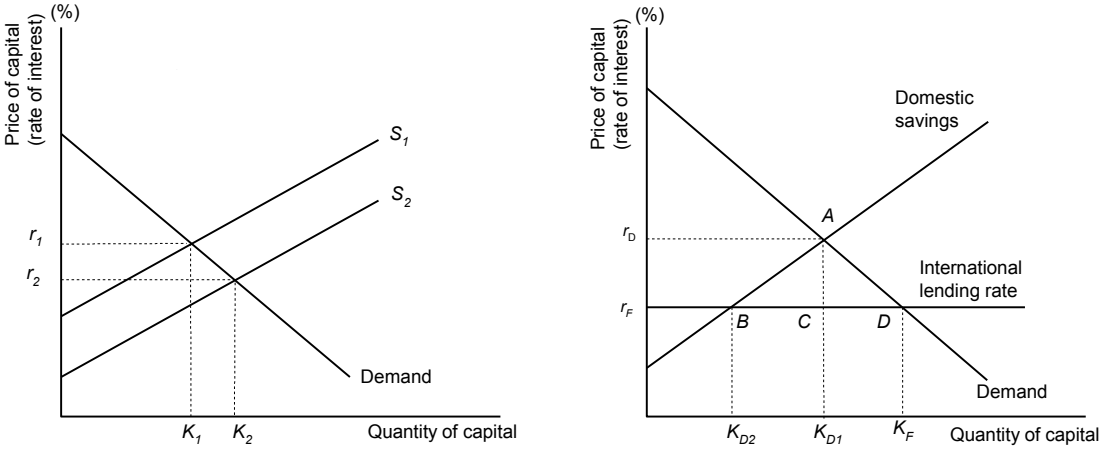


Figure 5.2 Quantity and price of capital (the rate of interest)

[4] This statement is based on the substitution effect of higher interest rates. High interest rates also have income effects. For given savings, higher interest rates increase income in the next period. This may have the reverse effect of discouraging saving.

On the other hand, domestic savers lose a surplus equal to area $r_D ABr_F$. The net efficiency gain to the local economy is area ABD.[5] For small economies the rate at which international capital can be attracted to a country is often a more important determinant of capital investment than the level of domestic savings.

Markets and innovation

The capitalist system is remarkably good at creating innovation. In competitive markets, prices constantly transmit new information about consumer demands and new incentives to producers. Competition pushes resources from low productivity or value activities to higher ones. Unlike in the static model, technology is not a given. Firms face continuous change and continually seek innovation in the form of new products or processes. Under competition, innovation is mandatory for survival of the firm. Competitive firms create new products, improved quality and product differentiation. Also, new technology is likely spread faster under capitalism than under other systems because innovators have an incentive to commercialise their knowledge in some way. Entrepreneurs bring the new concepts and processes into use.

A key feature of the innovative process is 'creative destruction'—a term first employed by Schumpeter (1943). Creative destruction refers to the entry and exit of firms into markets. Quality-improving innovations lead to new products displacing existing products and new firms replacing old ones. Schumpeter's concept of competitiveness has several features of the competitive model (many participants, low barriers to entry and limited long-run excess profits). However in Schumpeter's world, most products are differentiated and production techniques are rarely identical. Indeed, imperfectly competitive markets are an inevitable result of technical innovations. As Schumpeter (1943) observed:

> Entrepreneurial profits are the prizes offered by capitalist society to the successful innovator … The introduction of new methods of production and new commodities is hardly conceivable with perfect and perfectly prompt competition from the start. And this means that the bulk of what we call economic progress is incompatible with it. As a matter of fact, perfect competition is and always has been temporarily suspended whenever anything new is being introduced … even in otherwise perfectly competitive conditions.[6]

Porter (1990) also argued strongly for the role of innovation in creating economic growth. Firms must possess a competitive advantage in the form of lower costs or differentiated and preferred products. To achieve this advantage, an economy must continually upgrade capabilities and technology. In Porter's words, national competitive advantage:

> grows out of the capacity of a nation's firms to improve and innovate relentlessly … A nation's firms must relentlessly improve productivity in existing industries by raising product quality, adding desirable features, improving product technology, or boosting production efficiency … At the same time, an upgrading economy is one which has the capability of competing successfully in entirely new and sophisticated industries.[7]

However, in some contrast to Schumpeter, Porter stressed the importance of competitive product markets to innovation: 'among the strongest empirical findings from our research is the association between vigorous domestic rivalry and the creation and persistence of competitive advantage in an industry'.[8] Domestic rivalry stimulates new rivals through spin-

---

[5] To translate investment into economic growth, suppose that net investment (gross investment less depreciation) is 15 per cent of GDP and that the average real rate of return on capital is 12 per cent. With a constant labour supply, GDP would grow by 1.8 per cent per annum.

[6] Schumpeter (1943) pp. 104, 107.

[7] Porter (1990) p. 621.

[8] *Ibid*., p. 117.

offs, creates and attracts factors, upgrades home demand, encourages and rewards related industries. As examples of successes of the competitive market model, Porter cited the pharmaceutical industry in Switzerland, chemicals in Germany, computers and software in the United States and electronics in Japan. It might be added that arguably, Australia's most competitive domestic industry—the sports industry—is also one of its most successful international businesses. On the other hand, industry cartels reduce productivity and international market share. Examples are the Swiss watch industry, the German camera industry and the Australian coastal shipping and steel industries.

Research summarised by Aghion and Howitt (2009, Chapter 12) generally supported the case for product market competition, for example Nickell (1996). However, the results depend on the circumstances with better results in more technically advanced sectors. Aghion, Blundell *et al*. (2006) estimated the impact of industry entry rates on total firm productivity using panel data with 32 000 annual observations of firms in 166 different four-digit industries from 1980 to 1993 in the United Kingdom. They found that high entry rates increase the productivity of firms near the technology frontier but had little or negative impact on productivity of firms further from the frontier. Aghion, Burgess *et al*. (2006) report on the effects of de-licensing entry in India based on an annual panel of 24 000 observations in 85 industries over 18 years. They found that de-licensing increased the dispersion of output but did not actually increase entry.

In summary, competition is a strong force for innovation, investment and economic growth. Sustained productivity increases depend on the competitiveness of the economy. The incentives range from entrepreneurial profits to sheer survival. Protected monopolies are generally less creative than more competitive industries. But the competition that creates economic growth is not perfect competition. Although some rewards for innovation can be viewed as economic payments to entrepreneurs for time, ingenuity and risk, successful innovation often produces supernormal profits and an imperfectly competitive market.

These findings have some significant policy implications. Government may have to provide some protection for innovation by way of patents. However, generally, it should not only police anti-competitive behaviour among incumbent firms, it should also actively ensure that there are minimal barriers to entry and exit. And as Aghion and Howitt (2009) observe, complementary policies should help labour and capital move from laggard sectors to advanced sectors. In addition, they observe, as do most economists, that policies of supporting 'champion' firms to lead national innovation are unlikely to be successful.

Do markets produce efficient economic growth?

So competitive markets (though not perfectly competitive markets) drive growth, but do they produce an efficient rate of economic growth? It can be shown that perfectly competitive markets would achieve the three key conditions (consumption, production and overall economic efficiency) necessary for a Pareto-efficient inter-temporal allocation of resources just as they would for an efficient static allocation of resources. Where there are perfectly competitive markets for all goods, including future goods, individuals can borrow or lend at the inter-temporal price (the interest rate) and no firm or individual has independent power over any price. All agents face the same set of prices, including the same interest rate. To maximise utility, all consumers substitute marginal future consumption for present consumption at the unique relative price provided by the interest rate. To maximise profits, all firms use resources to invest in future output up to the point at which the marginal rate of return equals this same rate of interest. Because households and firms make their marginal inter-temporal decisions (for consumption and production) in accordance with the same relative price of present and future goods, the marginal rate of substitution of future for present consumption equals the marginal rate of transformation of present into future goods.

Market failures. However, all the market failures that occur in a one-period world occur in a multi-period world. Indeed, some failures are more significant in a long-run world. Firms under-invest in goods with substantial positive externalities. Thus, they tend to under-invest in areas of economic infrastructure such as power, water, transport and telecommunication networks. Firms are also likely to over-exploit common property resources and to create pollution and other negative externalities, thus reducing the productivity of natural resources.

Another important long-run market failure is under-investment in human capital and technology. Firms under-invest when investment produces benefits to third parties, which is typically the case with education and training. Baumol (2002) estimates that on average less than 20 per cent of the total economic benefit of innovations accrues to those who invest in making them happen. The rest of the benefit spills over to the society at large. Public investment in knowledge is particularly important because of the non-rival nature of knowledge (Dowrick, 2003). Protecting new ideas by patents creates monopoly powers for the holders and constrains the use of a non-rival good that can be used as an input to generate further knowledge and products.

We have also observed that innovation requires some product differentiation and produces supernormal profits. These innovations create imperfectly competitive markets. The optimal degree of departure from perfect competition is a difficult policy issue. Baumol (2002) argues that oligopoly is the industrial structure that best fosters productive innovation. On the other hand, Porter (1990) and Aghion and Howitt (2009) contend, with significant evidence, that innovation is maximised in a competitive industrial structure.

A classic inter-temporal issue is whether people make well-informed and efficient savings decisions. Pigou (1920) traditionally and behavioural economists Diamond and Vartiainen (2007) recently have conjectured that many individuals are short-sighted, undervalue future consumption and save too little. This would raise interest rates and reduce capital formation and economic growth. On the other hand, many people save as a precaution against adverse events, not just to optimise inter-temporal consumption. Precautionary saving lowers interest rates and raises investment and economic growth. Which distortion has the larger impact on savings and flow-on consequences for economic growth is hard to tell.[9]

In summary, in competitive markets firms innovate and invest in new products so long as the rate of return satisfies savers. In principle, there is equilibrium in the capital market as investors and savers make all desired trades. This would produce an efficient rate of economic growth. However, the supernormal profits and imperfect competition that result from innovation may lead to inefficiency in the allocation of resources. More importantly, all the market failures that occur in the static model also occur over time. In addition, firms under-invest in economic infrastructure and education, which are basic drivers of economic growth. Further, in the absence of regulations markets would almost certainly over-exploit natural resources.

Government and Economic Growth

We now discuss how government can contribute to the main forms of capital that, along with technical innovation, drive economic growth. We also consider how government may adversely impact on growth. Finally we examine empirical studies of the relationship between growth and government.

Institutional capital, the provision of a safe and lawful trading environment and open and competitive economies, is pre-eminently the responsibility of government. As we have seen

[9] Another concern is that market interest rates reflect the views of current participants in the market and not the interests of future generations. This may lead to higher interest rates and encourage current consumption of resources at the expense of future consumption.

so dramatically in recent years, confidence in the security of the financial sector is crucial to the mobilisation of savings and the efficient deployment of scarce capital resources. Economies need a national financial system and prudential supervision that will ensure the safety and soundness of financial institutions, while also allowing for competition and new entrants into the market. In addition, an open and competitive trading environment and liberal trade, capital markets and investment regimes are essential for economic growth. But markets are not automatically competitive. As Porter (1990, p.62) observed, 'Few roles of government are more important to the upgrading of an economy than ensuring vigorous domestic rivalry'.

Investment in **physical capital** in the form of economic infrastructure (such as energy and water supply, transport and communication facilities) is another major responsibility of government. As we have observed, firms under-supply such goods or, if they provide them, may try to create monopolies or exploit market power in one or other way. Note, however, that government responsibility for the provision of economic infrastructure does not necessarily imply public ownership of infrastructure enterprises. As we see in Chapters 18 and 19, public infrastructure can be supplied in many ways.

Some economists, for example Thurow (1996), argue that the provision of public infrastructure is strongly correlated with private productivity growth in many countries, and that this is an additional argument for government involvement. However, despite many macroeconometric studies of the possible external productivity benefits of public infrastructure, the size of these benefits remains questionable.[10] Cost–benefit analysis remains the best method for evaluating optimal investment in infrastructure.

Government also has a major role to play in the development and maintenance of **human capital** through education and health services. Public investment in education is required because markets undersupply education for three main reasons—the public good/positive externality nature of education, capital market imperfections and inequities in market provision. However private agents also gain substantially from investing in education, so the public gain needs to be identified (see Chapter 12). Some publicly financed health services are also required to ensure the basic health and fitness of the workforce.

In addition, firms under-invest in research and development because they cannot capture all the positive spin-offs from research findings. Also, technical knowledge has some non-rival features of a public good. Its use by one agent does not preclude its use by another agent, although this may reduce its profitability. Thus there is a strong public interest in research and development. Granting private property rights (patents) in discoveries is needed to encourage research. Also, governments may invest directly in research and development or subsidise privately funded research and development. These expenditures have produced large benefits in agriculture and telecommunications.[11] However determining the optimal amount of public funding of research is extremely difficult given the uncertainties of the outcomes.

Government must also protect **natural capital**. An economy may increase output in the short run while depleting its natural resources. However, standard national income accounts do not record depreciation of natural resources. In the long run, depletion of natural resources

[10] Using macroeconomic time series data, Aschauer (1989) estimated that the elasticity of GDP with respect to public investment was about 0.4 in the United States, which implied a very high rate of return to public investment. Others, for example the World Bank (1994a), have questioned the methodology used and the results gained. Lau and Sin (1997) estimate that the elasticity is about a quarter of Aschauer's estimate, but conclude that the externality benefits from investment in public infrastructure are positive. The practical conclusion is that the supply of economic infrastructure should be determined by microeconomic (cost–benefit) studies rather than by macroeconomic assumptions.

[11] In the 19th century, governments built the world's first telegraph lines. Recently, the US Department of Defense largely funded the Internet for the first 20 years of its life. Also in recent times, publicly funded research contributed greatly to increases in agricultural productivity, especially in less developed economies.

may reduce living standards. Given market failures in environmental resource use, efficient use of environmental resources is a major function of government (see Chapter 13).

Finally, government must maintain **social capital** through provision of social security services, basic health and education services and an income transfer system that reduces social inequalities. It also has a cultural leadership role. An inclusive and tolerant government supports social capital; a divisive and intolerant government can destroy it.

Government actions that reduce economic growth

Government can also reduce economic growth. Obviously, it may fail to perform efficiently the functions just described. Also, government expenditure, taxes or public policies may be inefficient. Government may allocate too few, or too many, resources to public goods. Government regulation of trade and industry, anti-competitive policies, output and price controls and other regulations may undermine investment and innovation. And more fundamentally, an extractive and corrupt government undermines private trade and investment (Acemoglu and Robinson, 2013). Here we highlight four issues.

The first issue is potentially excessive public expenditure. Public expenditure on goods and services nearly always crowds out some private expenditure either by direct competition for resources or via a reallocation of expenditure from the private to the public sphere. This is true of tax-financed or deficit-financed expenditure. Government borrowing tends to raise interest rates. This reduces private investment unless there is a strong multiplier effect of public expenditure. Saunders and Klau (1985) reviewed several econometric models and concluded that, although public expenditure generally stimulates the economy in the short run, borrowing usually crowds out some private investment. However, if the public projects are efficient (obtain high social rates of return), the crowding out of some private expenditure is **not** necessarily inefficient.

Second is the related issue of taxation. Taxation generally reduces the supply of capital and labour. In so doing, it often creates economic (deadweight) losses (see Chapter 27). From an economic growth perspective, taxation of income may discourage entrepreneurship, savings and investment, and labour supply. Taxation of capital and labour reduces the net return to capital and labour respectively and encourages factors of production to move from heavily taxed sectors to less heavily taxed (and less productive) sectors, including leisure.[12]

The third issue is income transfers. High personal benefits, for example for unemployment, disability and retirement, may reduce the labour supply. Following a wide-ranging survey of the impacts of benefits in Australia, Gruen (1982) concluded that 'an improvement in welfare benefit provisions can have fairly substantial effects both on the number claiming benefits and on the economic behaviour of potential claimants'.

Fourth, government is often a relatively inefficient producer of goods and services (see Chapters 16 and 18). The reasons include the lower level of competition, weaker incentives for efficient allocation of capital, greater operational constraints (e.g. audit responsibilities and wage controls) and slow responses to market changes.

However, an important caveat is in order. Many government policies or actions that decrease growth of GDP may on balance enhance social welfare. This is especially the case for some income redistribution policies. This is another version of the efficiency–equity trade-off that is pervasive in public economics.

[12] These generalisations must be treated cautiously because the income effects of taxes may offset the substitution effects and cause labour to work longer hours (see Chapter 26)!

Economic growth and government: evidence

There are many kinds of evidence on the relationship between economic growth and government. Much is relatively casual. For example, some observers contrast the high growth rates of East Asian countries with the slower rates of African countries and attribute this, among other factors, to the smaller role of government in East Asia. On the other hand, Stiglitz (1996) observes that governments have played a major role in some success stories and cites Japan and South Korea.

More formally, the World Bank (1997) drew its conclusions on the importance of institutional capacity from a wide-ranging survey of businesses operating in 69 countries. Porter (1990) likewise drew his conclusions about markets and economic growth from survey data. In Chapters 16 and 18, we report on various studies of the efficiency or otherwise of government as a provider of services. Aghion and Howitt (2009) report on industry-wide studies of the factors determining total factor productivity. Such sources can provide important insights into the economic growth and the role of government.

However, the dominant form of analysis has been cross-country econometric studies of the determinants of economic growth. These studies typically attempt to explain differences in the average rate of growth in GDP per capita in a large number of countries over a sample period using cross-sectional regression analysis. The independent variables may include government expenditure or tax revenue, the openness of the economy, levels of education, investment as a percentage of GDP, and geographic and even religious variables.

Sachs and Warner (1997) is an example of such a study. As shown in Table 5.2. Sachs and Warner found that 11 variables explained 84 per cent of the variance of growth per capita in 83 countries between 1965 and 1990. All variables have a plausible sign and are significant, or nearly so, at the 95 per cent level of significance (the *t*-statistic is greater than 1.95).

In relation to government, Sachs and Warner found that economic growth was correlated positively with openness of the economy to trade, government saving (current revenues less current expenditures) and an index of institutional quality (an average of sub-indexes for the rule of law, bureaucratic quality, corruption and the like).[13]

Table 5.2 Cross-country model of economic growth per capita 1965–1990[a]

Independent variable	*Estimated oefficient*	*t-statistic*
LnGDP per economically active person in 1965	–1.5	(–6.5)
Share of years open, 1965–90	10.9	(3.7)
GDP in 1965 times share of years open	–1.1	(–3.0)
Growth of economically active population – population growth	0.7	(1.9)
Central government budget balance (saving) 1970–90	0.11	(5.2)
Institutional quality index (1980)	0.32	(3.8)
Tropics	–0.8	(–3.0)
Landlocked	–0.6	(–2.3)
Share of natural-resource exports in GDP, 1970	–3.9	(–4.0)
Life expectancy	0.3	(2.8)
Life expectancy squared	–0.0026	(–2.3)
Adjusted R^2	0.84	

(a) Based on 83 countries with a mean growth of 0.33 per capita.

Source: Sachs and Warner (1997).

[13] The data for these indices were drawn from the *International Country Risk Guide*.

Sachs and Warner also found evidence for the conditional convergence theory of growth—holding constant other factors, such as government policy and investment, less developed countries grow faster than more developed ones. On the other hand, growth was lower in tropical countries (with poor soils and poor health), in landlocked countries (with low trade) and in resource-based economies. Curiously, the authors do not explain why they excluded two variables that they earlier found to be significantly correlated with economic growth. Sachs and Warner (1995) reported that economic growth was correlated positively with the share of total investment in GDP and negatively with the share of government consumption.

This kind of analysis requires careful interpretation, especially when an independent variable may be determined jointly with the dependent variable. Examples are the two-way relationships between economic growth and investment or between growth and human capital. A correlation does not prove that investment or human capital causes economic growth. The causal relationship might be in the reverse direction. Alternatively, both correlated factors may be caused by a third variable.[14] Another problem is multicollinearity — a correlation among explanatory variables. For example, high rates of investment tend to be correlated with education and openness. If independent variables are correlated, the precision of the estimated coefficients is reduced. If they are omitted, the model may be specified imperfectly and an estimated coefficient may pick up some of the impact of the omitted variables.

Other statistical problems include the sample set (which countries and years to include), the measurement of some variables such as openness, functional form because some relationships may be non-linear, parameter variability and limited degrees of freedom due to the small number of observations relative to potential explanatory variables. For example, analysing economic growth in 100 countries from 1960 to 1990, Krueger and Lindahl (2001) found that education and human capital stocks are an important factor in economic growth. However, this finding was driven mainly by the role of education in less developed countries. In their analysis, the relationship between education and economic growth was less clear in OECD countries. Aghion and Howitt (2009) found that the role of education depends on the stock of human capital and the proximity of the economy to the technological frontier. Primary and secondary education is important for countries that can import technology. Investment in higher education is important for countries close to the technological frontier.

Table 5.3 summarises the results of other cross-country growth studies with a focus on the role of government. Durlauf *et al.* (2005) provide a comprehensive list of such studies.

For the reasons just cited, conclusions must be drawn cautiously. However, the following conclusions appear valid:

1. There is general support for the theory of conditional convergence. However, when low-income countries do not achieve the necessary conditions (adequate physical and human capital) for economic growth, they fail to converge on higher-income countries.
2. Economic growth is usually positively associated with the quality of local institutions, the rule of law, open economies and the share of output allocated to investment generally and to investment in education (human capital improvement) specifically.
3. There is some evidence that economic growth is inversely related to the shares of government consumption expenditure and taxation in GDP, but the evidence is not conclusive. Microeconomic studies of the impacts of taxation on investment and labour supply provide stronger evidence of the negative impact of government on economic output (Temple, 1999).

[14] Technically, the simultaneous problem can be resolved if we can find exogenous variables to use as instruments. However, there are few such variables in cross-country data sets.

Table 5.3 Selected studies of growth rates of per capita income and role of government

Study	*Sample*	*Main results for the role of government*
Barro (1991)	98 countries, 1960–85	Growth positively related to initial human capital and measures of political stability. Inversely related to initial per capita GDP, proxy for market distortions, share of government consumption in GDP
Dowrick (1992)	Sample of OECD countries, 1960–85	Output growth is inversely related to the share of personal income tax in GDP
Engen and Skinner (1992)	107 countries, 1970–85	An increase of 2.5 percentage points in the share of tax reduced long-term output by 0.18 percentage points
Tanzi and Schuknecht (1995)	OECD and newly industrialised countries, 1975–90	Economies with low increases in public spending have lower unemployment, are more innovative (have more patents), and have smaller black economies
Cashin (1995)	23 developed countries, 1971–88	Productive public spending, including public investment and transfer payments that raise private investment, increases economic growth. Distortionary taxes that reduce the marginal return to private capital dampen economic growth
Dowrick (1996)	116 countries, 1950–90	Economic growth is positively related to government consumption up to a level of around 12 per cent of GDP and negatively related at levels above about 18 per cent of GDP
Sala-I-Martin (1997)	About 100 countries, 1960–90	Growth is positively correlated with the rule of law, the openness of the economy and the degree of capitalism. No measure of government spending was found to affect growth in a significant way
Barro (1997)	100 countries, 1960–90	Growth is related positively to schooling, rule of law, a democracy index, life expectancy and terms of trade. It is related negatively to initial GDP, fertility rates and public consumption expenditure

Box 5.2 Economic growth and government in Australia

Over the last 60 years the Australian economy has experienced relatively high growth rates. From 1960 to 1974, GDP grew at a high average rate of nearly 5 per cent per annum. Over the next 20 years to the mid-1990s, the average rate of growth fell to around 3 per cent per annum. Since the mid-1990s, the growth rate has averaged over 3 per cent per annum. The economy was not significantly affected by the Asian meltdown in 1998, the slowdown in OECD economies around the turn of the century or the global financial crisis which started in 2007. These generally high rates of growth reflect, among other factors, a rich resource endowment, high immigration and a high level of institutional capital.

In the 1960s, general government expenditure was below 30 per cent of GDP. It rose to around 35 per cent of GDP in the early 1970s and has fluctuated around this level since then. Some critics argue that this reduced growth rates in the late 1970's and 1980's, However, by OECD standards, government expenditure is modest (Chapter 2). Recent growth rates have been achieved without a significant decline in government expenditure as a proportion of GDP.

Starting in the early 1980s, the Australian economy has been substantially deregulated. Major features were the floating of the exchange rate in 1983, large-scale privatisation programs and the introduction of national competition policy in 1996 (see Chapter 14). Several reports by the Productivity Commission have argued plausibly that high growth rates since the mid-1990s are attributable to the more competitive environment.

In recent years, economic growth has been greatly facilitated by the extraordinary rates of growth in East Asia, most notably in China, and by related growth in the resources, education and tourism sectors.

The role of government in this has been more as facilitator than driver. Australia is seen as having stable institutions, laws and property rights which encourage private capital investment. Public investment in education and health provide a capable and healthy productive workforce. On the other hand, there is a view that government should have invested more on public infrastructure.

Conclusions

The major factors that enhance economic growth are sound legal and financial structures that encourage capital investment, a high level of human capital and its corollary investment in education (and in health), bountiful natural resources and its corollary preservation of the environment, a domestically competitive economy and an open economy with liberalised trade. There is also some evidence that democracy and a high level of social capital enhance economic growth.

Markets are a prime source of capital investment and innovation. However, government has the major responsibility for ensuring that these key conditions for economic growth are achieved. But government must also be mindful of other social objectives, notably creating and maintaining a socially just society and a sustainable economy in the long run.

Summary

- Economic growth depends principally on investment in physical and human capital and on innovation. Private investment and innovation depend on sound legal and financial structures and an open and competitive economy that encourages trade and entry into, and exit from, industries.
- The rate of economic growth is efficient when the rate at which consumers want to substitute future for present goods equals the rate at which the economy can transform present into future goods.
- However, this concept of efficiency is limited because innovation is taken as given.
- Competitive markets are a dynamic force for investment and innovation. However, the process of innovation creates supernormal profits and the risk of imperfect competition.
- Also, private firms may under-invest in public goods, including economic infrastructure, technology and education. They may also over-exploit resources.
- Government has an important role in providing the institutional, social and human capital necessary for economic growth. Government functions include providing a safe and lawful environment, ensuring markets work fairly, promoting domestic competition and international trade, addressing market failures, and protecting natural capital and investment in economic infrastructure, education, science and technology, and the health of the population.
- Government may also slow down economic growth though high government expenditure and taxes. Poor regulation may restrict competition, stifle innovation and add to industry costs. Growth is usually lower in highly regulated economies, in closed economies and in countries that fail to protect private property.
- However, government also has other social objectives, notably creating and maintaining a socially just society and a sustainable economy in the long run which may involve a trade-off with short-term economic growth.

Questions

1. Can a country have a high level of income and a low rate of economic growth or vice versa? If so, give examples and explain why.
2. What are the technical economic conditions necessary for an efficient rate of economic growth? What are the limitations of this concept?
3. Is an efficient rate of economic growth always the same as an optimal rate of economic growth? Are high rates of economic growth always better than low rates?
4. Do high levels of domestic savings create high rates of economic growth?
5. What kinds of markets contribute most to economic growth? What is the evidence? What policy conclusions can be drawn?
6. What role does government have in facilitating economic growth?
7. How may government have an adverse effect on economic growth?
8. Does higher government expenditure increase both economic output and growth?
9. Do lower interest rates increase growth by increasing investment?
10. Given that investment is a function of profit, does a tax on profit reduce investment and economic growth?
11. What problems arise in trying to estimate the effect of government activity on economic growth from econometric cross-country studies of economic growth?

Further Reading

Acemoglu, D. and Robinson, A. (2013), *Why Nations Fail: The Origins of Power, Prosperity and Poverty*, Random House.

Aghion, P. and Howitt, P. (2009) *The Economics of Growth*, MIT Press, Cambridge, MA.

Bouis, R., Duval, R. and Murtin, F. (2011) *The Policy and Institutional Drivers of Economic Growth across OECD and Non-OECD Economies: New Evidence from Growth Regressions*, OECD Economics Department Working Papers, No. 843, OECD, Paris.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W. (2005) 'Growth Econometrics', in *Handbook of Economic Growth*, P. Aghion and S.N Durlauf, (eds), Elsevier, Amsterdam, pp. 555–677.

Hindriks, J. and Myles, G.D. (2006) *Intermediate Public Economics*, Chapters 19 and 21, MIT Press, Cambridge, MA.

Jones, C.I. (2002) *Introduction to Economic Growth*, Chapters 1, 2 and 10, W.W. Norton, New York.

Productivity Commission (2016), *Increasing Australia's Future Prosperity*, Discussion Paper, Canberra.

Productivity Commission (2017), *Shifting the Dial, 5 Year Productivity Review*, Report No. 84, Canberra.

Rogers, M. (2003) 'A survey of economic growth', *Economic Record*, 79, 112–135.

Sachs, J.D. and Warner, A. (1997) 'Fundamental sources of long-run growth', *American Economic Association, Papers and Proceedings*, 97, 184–188.

Part

3

Economic Evaluation and Public Policy

Chapter

6

Valuing Individual Preferences

Utility and its measurement lie at the heart of political economy.

Jules Dupuit

Individual Preferences and Utility ♦ Deriving Demand Curves from Preferences and Budget Constraints ♦ Valuation Principles Further Considered ♦ From Valuation Principles to Practice

Respect for individual values lies at the heart of democracy. And as Dupuit (1844), a French engineer, observed over 170 years ago, the valuation of individual preferences lies at the heart of the economic approach to public policy. What do people want? How are their preferences valued? In this chapter we discuss the concepts of preferences and how they can be valued in monetary units. It should be noted that monetary units are adopted as a practical measure. We could measure preferences in bottles of rum like the early British settlers, but dollars are much more practical. These measures of preference provide the fundamental elements from which public policy can be constructed.

As a starting point, we may suppose that the value of a good to someone is the price that he or she is willing to pay for it. More precisely it is the maximum price. This represents the value of other goods that he or she is willing to give up for the new good. This is generally viewed as a sensible approach. Nevertheless, complications arise and must be dealt with. First, the value of something is not unique; it depends on the quantity consumed. The value of the marginal item consumed may be a lot less than the average value of the good. Second, the price that someone is willing to pay for a good often depends on their income. If a person's real income changes, he or she may place a different value on a good. Third, there may be no observable price for some goods, notably for non-market goods.

In this chapter we discuss mainly conceptual issues in valuing goods. The first section describes the nature of preferences and the utility or wellbeing of individuals. We then discuss how to derive demand curves from preferences. The third main section discusses valuation principles in more detail. Finally, we discuss how to put these valuation principles into practice focussing on standard issues in estimating individual demand in markets. In Chapter 11 we discuss various practical issues of estimating individual preferences using other revealed or stated preference methods.

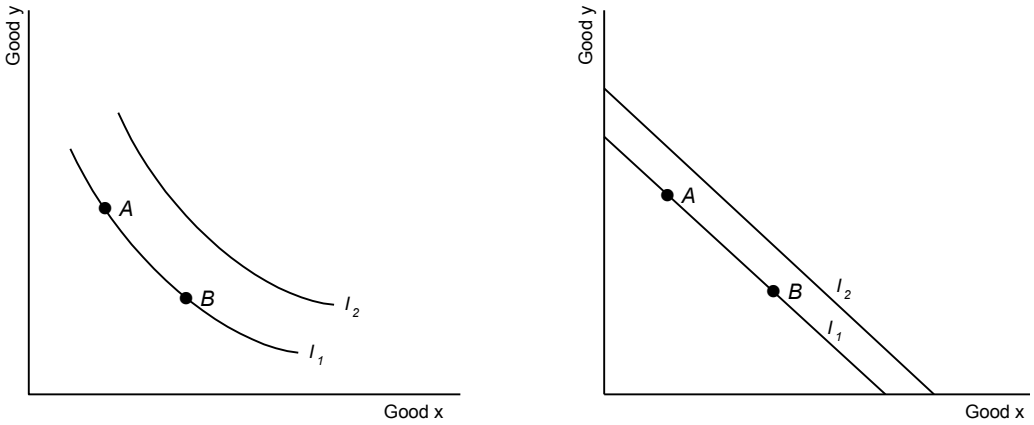
Individual Preferences and Utility

The standard economic theory of valuation starts from the premise that inferences about individual preferences can be drawn from observations of their choices in various contexts, but especially in markets. This is broadly known as revealed preferences. This assumes that individuals make rational (consistent and informed) choices and have sensible preferences. In Chapter 4 we saw that behavioural economists challenge these assumptions and have good reasons for doing so. However, this does not mean that the standard theory of revealed preference has no practical use. Bernheim and Rangel (2007) provide a comprehensive review of behavioural public economics and conclude that preferences may still be discovered by selective application of the revealed preference principle and that practising behavioural economics requires modifying rather abandoning the standard theory of revealed preference. In so far as preferences cannot be inferred from observed (revealed) choices, for whatever reason, other ways to value individual preferences may need to be adopted. Of these the most common way is the use of stated preference methods, which we discuss in Chapter 11.

To illustrate the nature of preferences it is useful to work with the choice between two goods, x and y. These goods can be consumed in various quantities, which make up a consumption bundle. This bundle can be represented by (q_x, q_y) where q_x and q_y denote the quantity of goods x and y respectively. Suppose that there are three consumption bundles (A, B and C) which are represented by (3, 4), (2, 5) and (4, 3) units of goods x and y respectively.

The theory of preferences assumes that individuals can rank all such bundles. This means that they either regard any two bundles as equal value (i.e. they are indifferent between them) or prefer one bundle to the other. Second, preferences are assumed to be transitive. Transitivity means that, given any three consumption bundles (such as A, B and C), if A is preferred to B and B is preferred to C, then A is preferred to C. Thirdly, if someone is indifferent between A and B and between B and C, she is indifferent between A and C. These assumptions, taken together, ensure that an individual has a complete preference ordering.

In addition, individuals are assumed to prefer more of any good. If bundle A contains more of one good and no less of the other than does bundle B, then bundle A is preferred. Finally, individuals are assumed to attempt to attain the highest level of satisfaction possible, which is the most preferred bundle of goods consistent with their budget. In other words, in their choices they attempt to maximise their utility subject to a budget constraint.



(a) x and y are normal goods

(b) x and y are perfect substitutes

Figure 6.1 Alternative forms of indifference curves

Mapping preferences and trade-offs

These preference concepts can be illustrated with the aid of indifference curves (see Figure 6.1). An indifference curve shows combinations of goods that give an individual equal satisfaction (utility). Indifference curves imply trade-offs between goods and implicit values of goods. An extra unit of good x is worth the quantity of good y that an individual is willing to give up for it. A higher indifference curve implies a higher level of utility.

Indifference curves generally slope downwards and are usually convex. The downward slope shows that a person is willing to sacrifice some amount of one good in order to obtain more of the other. The slope of the curve at any point shows the marginal rate of substitution of good y for good x (MRS_{yx}). MRS_{yx} is the marginal amount of good y that a consumer is willing to sacrifice to obtain a unit increase in good x. A convex curve implies diminishing marginal rate of substitution. This means that the more units of good x that someone possesses, the less of good y he or she is willing to sacrifice to obtain an additional unit of good x. Point A in Figure 6.1a corresponds to a consumption bundle with a large amount of good y. The bundle at point B contains less of good y but more of x. Accordingly the slope of the indifference curve is flatter at point B than at point A. MRS_{yx} declines as we move down the indifference curve.

By contrast Figure 6.1b shows an indifference curve with a constant slope. In this case the two goods are perfect substitutes.

Various examples

Although the theory of preferences is usually illustrated by comparing bundles of market goods, such as clothes and food, the theory is general. Individuals can have preferences over market and non-market goods, over market goods (income) and leisure, and over current and future consumption (as we saw in Chapter 5). Moreover, each set of preferences can be represented by indifference curves.

Figure 6.2a shows the trade-offs that an individual would be willing to make to obtain more of an environmental good (such as air quality). To obtain a non-marginal increase of an environmental good from E_2 to E_3, the individual is willing to give up ($M_2 - M_3$) market goods. The convex indifference curve implies that as the quality of the environment increases, the individual is willing to sacrifice fewer market goods for additional units of the environmental good.

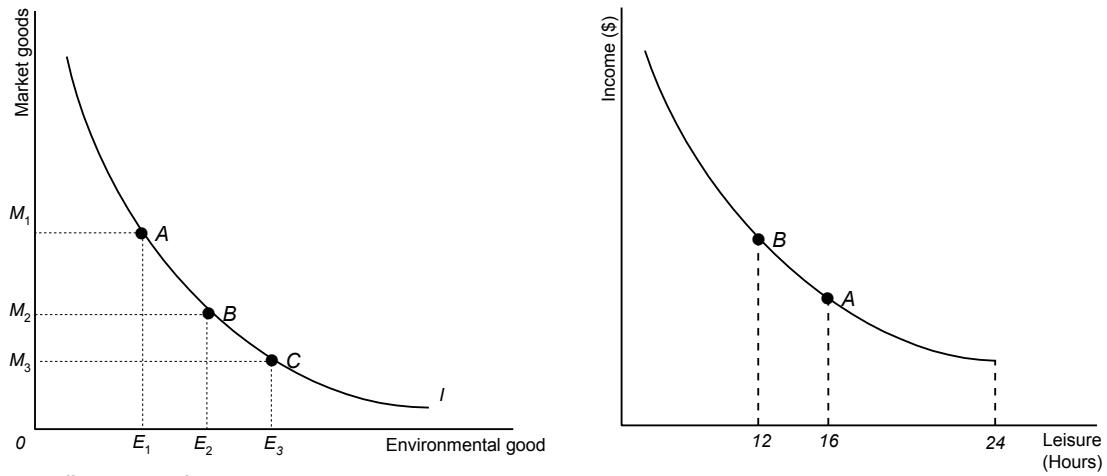


Figure 6.2 Indifference curves for other trade-offs

Figure 6.2b shows the trade-offs that an individual is willing to make between income and leisure. The slope of the indifference curve at any point shows the marginal rate at which individuals require to be compensated for loss of leisure. The slope generally increases as leisure declines. The less leisure that someone has, generally the greater will be the compensation required to give up an extra hour of leisure.

Utility functions

As we will see, a demand curve can be derived from an indifference map. However, the process of estimating demand is more precise and rigorous if mathematics is employed to describe consumer preferences and the indifference curves that depict these preferences. Economists use utility functions to represent these preferences in mathematical form.

Utility function
Shows an individual's utility as a function of their consumption of goods or their income and sometimes also their leisure

A **utility function** shows utility as a function of an individual's consumption of goods. Higher utility numbers indicate greater utility. However, the utility numbers are arbitrary and ordinal rather than cardinal. A bundle of goods with a utility of 200 is preferred to a bundle with a utility value of 100, but it does not necessarily imply twice as much satisfaction.

Consider first a simple linear utility function for an individual, in which goods x and y are perfect substitutes as in Figure 6.1b.

$$U = u\,(q_x, q_y) = q_x + q_y \tag{6.1}$$

If we have the following four consumption bundles for the two goods, (2, 4), (2, 6), (4, 2) and (5, 10), the corresponding utility numbers are 6, 8, 6 and 15. These numbers indicate the preference rank of the bundles. Bundle 4 is the preferred bundle followed by bundle 2 and by bundles 1 and 3. The individual is indifferent between bundles 1 and 3. However, the numbers cannot be interpreted as actual utility magnitudes.

We now rank the same four consumption bundles where the two goods are not perfect substitutes and the principle of diminishing marginal rate of substitution applies. To illustrate this case, suppose that the utility function has a Cobb–Douglas form where the exponents sum to one and exponent α equals 0.7.

$$U(q_x, q_y) = q_x^{\alpha} q_y^{1-\alpha} = q_x^{0.7} q_y^{0.3} \tag{6.2}$$

The estimated utility numbers for the bundles are now 2.56, 2.78, 3.87 and 6.16 respectively. Bundle 4 is still preferred, but followed by bundles 3, 2 and 1 respectively.

Deriving Demand Curves from Preferences and Budget Constraints

To derive demand curves that show what individuals are willing to pay for extra units of a good, we need to know their income (budget constraint) as well as their preferences. Assuming a one-period model in which an individual consumes all his or her income, the budget constraint is given by:

$$M = p_x q_x + p_y q_y \tag{6.3}$$

where M is the income of the individual and p_x and p_y are the prices of goods x and y respectively. The budget line that represents this constraint is shown along with a set of indifference curves in Figure 6.3 overleaf.

The budget equation can be rearranged as:

$$q_y = M/p_y - (p_x/p_y)\, q_x \tag{6.4}$$

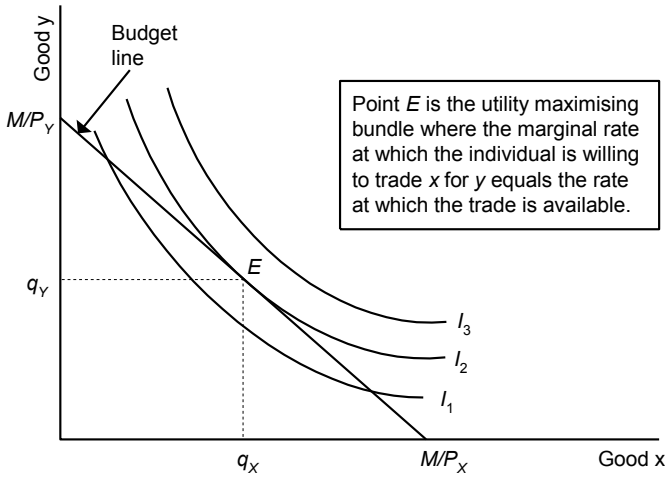


Figure 6.3 The consumption bundle that maximises utility

In this equation, M/p_y is the intercept term for good y and p_x/p_y is the slope of the budget line. Demand for y depends on real income (M/p_y) and the relative price at which a consumer can substitute good y for good x (p_x/p_y).

To know how much someone can obtain from their income, M must be divided by the prices of goods. Also, to know the real cost of something, we need to know relative prices. Suppose that p_x is \$2 and p_y is \$10; the relative price is 0.2. The real cost of good x is 0.2_y. If p_y falls to \$2, the real cost of good x is 1.0_y. The rearranged budget equation (6.4) implies that economic behaviour depends on real income and relative prices, not on nominal magnitudes.

The slope of the budget line is the ratio of the prices of the two goods (the rate at which an individual can trade good y for good x). If we combine this rate with the MRS (the marginal rate at which an individual is willing to trade) we obtain the consumption bundle that a utility-maximising consumer chooses. The assumptions that a consumer has downward-sloping, convex indifference curves and prefers more to less imply that the utility-maximising consumption bundle must lie on the budget line. As shown in Figure 6.3, the utility-maximising bundle is the tangent point (E) of the indifference curve and the budget line. At this point, the MRS equals the relative price.

Demand curves

Figure 6.4 shows the quantity of good x demanded at various prices. To estimate the demand, we shift the budget line to reflect the change in p_x and estimate the new utility-maximising consumption bundles (E, F and G). For simplicity, indifference curves are not shown in this figure. Note that M and p_y are held constant so that the y-intercept is constant. The price consumption curve shows the utility-maximising quantities of good x at each price for x (\$6, \$4 and \$2). The relevant quantities (20, 30 and 40 respectively) can be read off the graph.

Formally, the individual's choice of consumption bundle is a constrained maximisation problem. The consumer chooses quantities of x and y (q_x and q_y) that maximise their utility $U(q_x, q_y)$ subject to their budget constraint ($M = p_x.q_x + p_y.q_y$). The outcome is:

$$q_x^* = D_x\ (p_x, p_y, M) \tag{6.5}$$

$$q_y^* = D_y\ (p_x, p_y, M) \tag{6.6}$$

where q_x^* and q_y^* are the utility-maximising quantities, which can be represented as demand functions D_x and D_y.

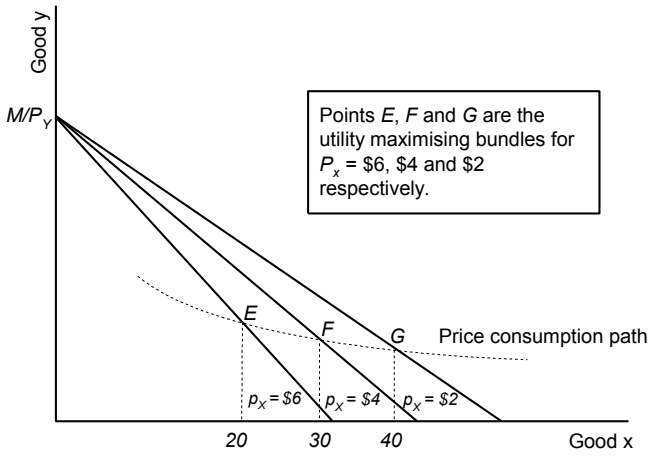


Figure 6.4 Estimating demand from price consumption path

Box 6.1 Estimating a demand function from a specified utility function

Following the notation in the text, suppose that the utility function for an individual takes on the Cobb-Douglas form:

$$U(q_x, q_y) = q_x^{0.7} q_y^{0.3}$$

The budget constraint is given by:

$$M = p_x q_x + p_y q_y$$

The Lagrangian function is formed by combining the utility function and budget constraint:

$$L = q_x^{0.7} + q_y^{0.3} + \lambda(M - p_x q_x - p_y q_y)$$

Differentiating L with respect to q_x, q_y and λ and setting the partial derivatives to zero:

$$dL / dq_x = 0.7 q_x^{-0.3} q_y^{0.3} - \lambda p_x = 0$$
$$dL / dq_y = 0.3 q_x^{0.7} q_y^{-0.7} - \lambda p_y = 0$$
$$dL / d\lambda = M - p_x q_x - p_y q_y = 0$$

Solving the equations for q_x and q_y we have the following utility-maximising quantities:

$$q_x^* = 0.7M / p_x$$
$$q_y^* = 0.3M / p_y$$

With this utility function, expenditure shares are constant. Also, the demand for x does not vary with p_y and the demand for y does not vary with p_x. This feature of the Cobb-Douglas function limits its suitability for applied work.

Thus, the optimal quantities depend on the prices of goods x and y and money income. Different preferences lead to different demand functions. Footnote 1 shows how these general results are derived.[1] Box 6.1 shows how specific demand functions can be derived from a particular (Cobb–Douglas) utility function.

[1] When the utility functions are differentiable, we can use the Lagrange multiplier method to solve for the demand functions. A Lagrangian function is formed by combining the utility function, budget constraint and the Lagrange multiplier (λ):

$$L = U(q_x, q_y) + \lambda(M - p_x q_x - p_y q_y)$$

By differentiating the Lagrangian function with respect to qx, qy and λ and setting the partial derivatives to zero, we obtain a system of three simultaneous equations:

$$U_x(q_x^*, q_y^*) - \lambda p_x^* = 0$$
$$U_y(q_x^*, q_y^*) - \lambda p_y^* = 0$$
$$M - p_x x^* - p_y y^* = 0$$

Demand functions for good x and y can then be obtained by solving for the unknowns q_x^*, q_y^* in terms of p_x, p_y and M.

Substitution and income effects

Substitution effect
The change in consumption of a good or service due to a change in relative prices

With ordinary demand curves, price changes have two effects: substitution and income effects. The **substitution effect** is the effect of a price change on demand due to the change in relative prices, holding the consumer's real income (utility) unchanged. Holding p_y and M constant, if p_x falls, the relative price of good x falls; if p_x rises, good x becomes relatively more expensive. Assuming a diminishing marginal rate of substitution, the substitution effect is always negative. That is, consumption increases as relative price falls and vice versa.

Income effect
The change in consumption due to a change in real income

The **income effect** of a price change is the change in demand due to a change in the real income of consumers. Holding p_y and M constant, a change in p_x changes a consumer's real income. If p_x falls, the budget set expands and the consumer ends up on a higher indifference curve (real income increases). If p_x rises, the budget set contracts and the consumer ends up on a lower indifference curve (real income falls). For normal goods, the income effects of price changes reinforce the substitution effects.

Substitution and income effects are illustrated in Figure 6.5. Amy's initial budget line for two goods, food and clothes, is shown by line AB. Given her indifference curve I_1, Amy maximises her utility at point E_1 and consumes Y_1 units of food and X_1 units of clothes. After the price of clothes falls, the budget line rotates to AC. Amy chooses point E_2 on indifference curve I_2 and consumes Y_2 and X_2 units of food and clothes respectively. To decompose this change in consumption into substitution and income effects, we draw a price line ($A'C'$) parallel to the new set of relative prices and at a tangent to the initial utility curve (I_1). Holding utility (real income) constant, the increase in consumption of clothes from X_1 to X_3 is the substitution effect due to the change in relative prices. The increase from X_3 to X_2 is the income effect.

Ordinary demand curve
A demand curve that shows how quantity demanded varies with price, holding money income constant. Also called observed demand curve

Ordinary and compensated demand curves

Thus, there are two kinds of demand curve. **The ordinary demand curve** (known as the Marshallian demand curve) shows the quantities demanded at all prices holding nominal income constant. This demand curve is derived from the price consumption curve as in Figure 6.4. This curve includes changes in real income. The ordinary demand curve is also called the observed demand curve.

Compensated demand curve
A demand curve that shows how quantity demanded varies with price, holding real income (utility) constant

The compensated demand curve (known as the Hicksian demand curve) shows the effect of price changes on the quantity demanded, holding real income constant (i.e. for a constant utility level). It is derived by varying the price changes as in Figure 6.5 and by drawing out the demand curve associated only with the substitution effects, excluding changes in real income. We cannot hold utility constant when prices change and money income is fixed. With the compensated demand curve, the consumer gains or loses notional income to hold their utility constant. This eliminates the income effect of a price change.

A typical relationship between an ordinary and a compensated demand curve is shown in Figure 6.6. Suppose that the two curves intersect at initial price P_1 and consumption of Q_1. With the ordinary demand curve, if the price falls to P_2, consumption increases to Q_2. If the price rises to P_3, consumption falls to Q_3. Both changes include income effects. The compensated demand curve excludes income effects and is steeper for a normal good. If price falls to P_2, consumption increases only to Q_4. If the price rises to P_3, consumption falls to Q_5.

The expenditure function. The problem confronting the utility-maximising consumer can also be viewed as an expenditure-minimising problem. In this case, the aim is to find the minimum expenditure necessary to achieve a specified level of utility. Instead of moving along the budget line until the highest indifference curve is reached, the consumer now moves

along their indifference curve until the lowest iso-expenditure line is reached.[2] This shows the minimum level of expenditure necessary to achieve a given utility level as a function of prices and the required utility level. We draw on expenditure functions below to estimate the effects of changes on individuals. Box 6.2 overleaf derives compensated demand functions for a

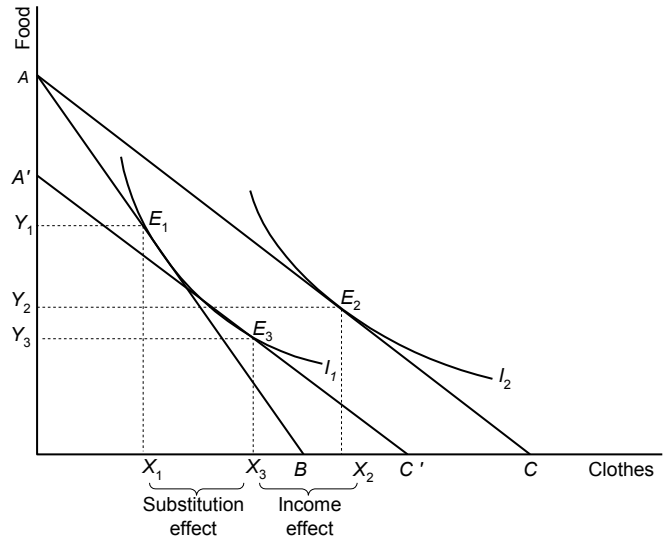


Figure 6.5 Income and substitution effects of a price decrease

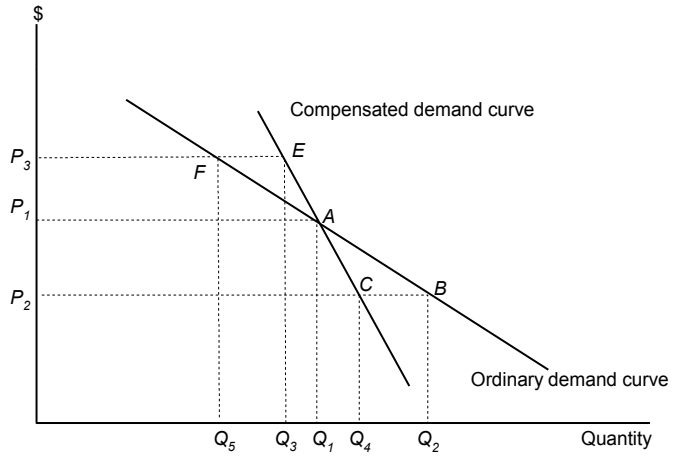


Figure 6.6 Ordinary and compensated demand curves

[2] The isoexpenditure line is similar to the budget line and satisfies the equation $p_x q_x + p_y q_y = M^0$ (M^0 is a fixed expenditure). A Lagrangian function can be written for this minimisation problem as: $L = p_x q_x + p_y q_y + \mu(U^0 - U(q_x, q_y))$. The general compensated demand functions for x and y are: $q_x^{CD} = CD_x(p_x, p_y, U^0)$ and $q_y^{CD} = CD_y(p_x, p_y, U^0)$. Substituting these optimal values in $p_x q_x + p_y q_y$ gives:

$$(p_x . CD_x(p_x, p_y, U^0) + p_y . CD_y(p_x, p_y, U^0) = M(p_x, p_y, U^0),$$

which is called the expenditure function

particular optimisation problem. Box 6.3 shows how income and substitution effects can be estimated,

As we have seen, the demand for a good depends on whether real income changes as with an ordinary demand curve or held constant as with a compensated demand curve. We need therefore to define more precisely the price that individuals are willing to pay for a good. This will also lead us to the question of what individuals are willing to accept as compensation for not having some good or service.

To do this we introduce the concepts of compensating variation (CV) and equivalent variation (EV) measures of value. We then discuss valuation with an ordinary demand curve which we will call the consumer surplus (CS) method. Finally we discuss the relationships between CV, EV and CS measures of value and how to choose between them.

Compensating and equivalent variations

The difference between CV and EV values is the reference point. For CV, the reference point is the individual's current level of utility *before* an economic change; for EV, the reference point is his or her level of utility *after* the change.

Box 6.2 Estimating compensated demand functions

Let the utility function and budget equation be as in Box 6.1. Consider a situation in which the government subsidises consumers so as to leave their utility unchanged. Assume this is done by a lump-sum payment which gives the consumer the minimum income necessary to achieve their initial utility level. Their compensated demand function will give the quantities of commodities that they will buy as functions of commodity prices under these conditions. This can be obtained by minimising consumers' expenditures subject to the constraint that their utility is at a fixed level U^0. The Lagrangian function in this case is:

$$L = p_x q_x + p_y q_y + \mu(U^0 - q_x^{0.7} q_y^{0.3})$$

Differentiating L with respect to q_x, q_y and μ and setting the partial derivatives to zero:

$$dL / dq_x = p_x^{-0.7} q_x^{-0.3} q_y^{0.3} \mu = 0$$
$$dL / dq_y = p_y^{-0.3} q_y^{-0.7} q_x^{0.7} \mu = 0$$
$$dL / d\mu = U^0 - q_x^{0.7} q_y^{0.3} = 0$$

Solving the equations for q_x and q_y, we have compensated demand functions:

$$q_x^{CD} = U^0 / \left(\frac{3}{7}\frac{p_x}{p_y}\right)^{0.3}$$

$$q_y^{CD} = U^0 / \left(\frac{7}{3}\frac{p_y}{p_x}\right)^{0.7}$$

Box 6.3 Substitution and income effects

The expenditure function is given by: $M(p_x, p_y, U^0)$.

Therefore by definition:

$$CD_x(p_x, p_y, U^0) = D_x[p_x, p_y, M(p_x, p_y, U^0)]$$

If we consider a price change of x, that is differentiating with respect to p_x, we get:

$$\frac{\partial CD_x}{\partial p_x} = \frac{\partial D_x}{\partial p_x} + \frac{\partial D_x}{\partial M}\frac{\partial M}{\partial p_x}$$

Rearranging:

$$\frac{\partial D_x}{\partial p_x} = \frac{\partial CD_x}{\partial p_x} - \frac{\partial D_x}{\partial M}\frac{\partial M}{\partial p_x}$$

The above equation shows the effect of a price change on the ordinary demand curve as substitution effect (the first part of the right-hand side shows the effect of a price change on the quantity demanded when utility level is fixed) and income effect (the second term on the right-hand side shows the effect of a price change on demand through the change in purchasing power).

Given that: $\frac{\partial M}{\partial p_x} = q_x$

The income effect is: $-q_x \frac{\partial q_x}{\partial M}$

which is negative for a normal good (because $\frac{\partial q_x}{\partial M} > 0$ for a normal good).

Consider first the **compensating variation**:

- The CV value of a good is the maximum amount that an individual would be willing to pay (WTP) for it and be no worse off with it than without it. This equals the income that can be taken away from someone and leave him or her at their initial utility level.
- The CV measure for loss of a good is the minimum amount that an individual would be willing to accept (WTA) as compensation for the loss and be no worse off than before.

Compensating variation
The amount of money to be given to, or taken from, an individual that would restore him/her to his/her initial level of utility following a change in price or consumption

Turning to the **equivalent variation**:

- The EV value of a good is the minimum amount that an individual would be willing to accept (WTA) and be as well off without the new good as he or she would be with it.
- The EV measure for loss of a good is the maximum amount that someone would be willing to pay to avoid the loss given that it would otherwise occur.

Equivalent variation
The amount of money to be given to, or taken from, an individual that would allow him/her to experience the same level of utility that he/she would experience with a given change in price or consumption

To estimate the CV value of a good, suppose that an individual has an initial level of utility of U_0 with a money income Y_0 and an amount of a public good G_0:

$$U_0 (Y_0, G_0) \tag{6.7}$$

Suppose that government proposes to increase the amount of the public good to G_1, which would increase the individual's utility to U_1:

$$U_1 (Y_0, G_1) \tag{6.8}$$

We want to know by how much the extra amount of the public good increases the individual's utility but we cannot directly measure U_0 or U_1. We therefore seek an indirect measure of the benefit by estimating what an individual would be WTP for the increase in the public good but remain on the initial level of well-being.

$$U_0 (Y_0 - WTP,\ G_1) = U_0 (Y_0, G_0) \tag{6.9}$$

This WTP amount is the (CV) monetary value of the benefit from the increase in the public good from G_0 to G_1.

To estimate the EV value of a good, we ask how much income an individual would be WTA to forgo the increase in the public good and still be as well off as if he or she had received the extra amount of the public good. In this case we consider the combinations of money income and public good that would yield an equal level of utility (U_1).

$$U_1 (Y_0 + WTA,\ G_0), = U_1 (Y_0, G_1) \tag{6.10}$$

Similar measures can be derived for losses in utility. However, in this case, the CV amount is measured by WTA and the EV amount by WTP. Suppose that the change from G_0 to G_1 denotes a loss of some amount of a public good. The CV is the amount of money that will compensate the individual for the loss and leave him or her at their initial level of utility.

$$U_0 (Y_0 + WTA,\ G_1) = U_0 (Y_0, G_0) \tag{6.11}$$

The EV is the amount of money that an individual would be WTP to avoid the change.

$$U_1 (Y_0 - WTP),\ G_0) = U_1 (Y_0, G_1) \tag{6.12}$$

The concepts of CV and EV are illustrated below using the example of (1) a new railway line and (2) a reduction in fares for an existing rail service.

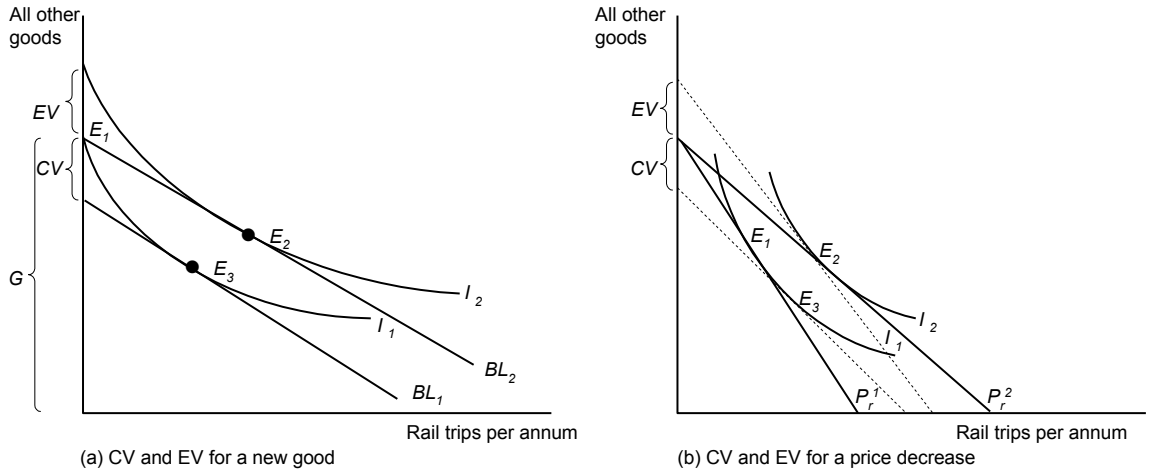


Figure 6.7 Compensating and equivalent variations

Valuations for a new good. Figure 6.7a shows Amy's consumption bundle without and with a railway line. Without a rail line, Amy is at equilibrium point E_1. She consumes G units of all other goods at an average composite price p_g and achieves indifference curve I_1. With a new rail line, given a price (p_r) for a rail trip, her new budget line is shown as BL_2, with a slope of p_r / p_g. Her new equilibrium point is E_2 and her utility has risen to the I_2 curve.

To estimate CV, we draw a budget line BL_1, with a slope of p_r / p_g. Given the indifference curve I_1, the individual would choose equilibrium point E_3. This indicates that, if the rail is built and \$CV are taken away from Amy, she will just be as well off as without the rail line because she is on her initial indifference curve I_1 in both cases. Thus, CV represents the maximum price that Amy will pay for the new rail line given the rail price.

On the other hand, taking the new utility level I_2 as the reference point, the equivalent variation is shown by EV. This is the extra income that would make Amy as well off without the new rail as she would be with it.

Valuations of a price reduction. Now suppose that the rail agency reduces the rail fare from p_r^1 to p_r^2 and the price of other goods p_g does not change. As shown in Figure 6.7b, the CV is the amount of money that can be taken from an individual in the new lower price situation and leave her as well off (at utility level I_1) as with the initial higher price. The benefit of the fall in price is shown by the distance CV. In terms of the consumer expenditure function:

$$CV = M(p_r^1, p_g, I_1) - M(p_r^2, p_g, I_1) \tag{6.13}$$

The EV of the price change is the amount of money that must be given to an individual in the initial price situation to make her as well off as she would be with the new lower price. Here the new utility level (I_2) is the reference point. The equivalent variation for the fall in real fare is shown as the distance EV in Figure 6.7b. In terms of the consumer expenditure function:

$$EV = M(p_r^1, p_g, I_2) - M(p_r^2, p_g, I_2) \tag{6.14}$$

Comparing CV and EV

Of course, the distinction between CV and EV is of no consequence when they produce the same result. If the marginal value of a dollar of consumption is constant over changes in income, CV = EV. In this case, utility is a linear function of income.

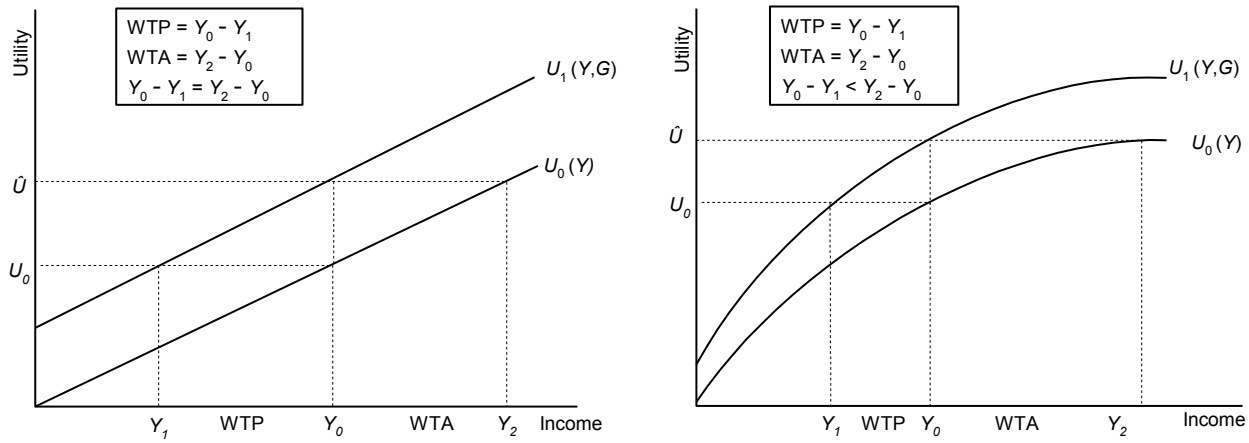
Figure 6.8a shows two linear utility functions. The U_0 curve shows utility simply as a function of private income. The U_1 curve shows utility as a function of private income and a public good. For simplicity we assume here that a public good is either provided or not. We are not dealing with different quantities of a public good. Suppose that an individual has a private income of Y_0 and no public good. She has a utility of U_0. If the public good were provided, she would be willing to pay $Y_0 - Y_1$ for the good and remain at U_0. This is the CV valuation of benefit. On the other hand, if she is deemed to have a right to the public good she would be entitled to utility $\hat{U}$. She would be willing to accept $Y_2 - Y_0$ income for loss of the public good and remain at $\hat{U}$ level of well-being. This may be interpreted as the EV value of the good. However, with linear utility curves, $Y_0 - Y_1 = Y_2 - Y_0$ so that WTP = WTA. In this case the CV valuation = the EV valuation.

This result does not hold if utility is a non-linear function of income. Utility is generally expected to rise with income but at a declining rate as an extra dollar to a rich person has less value than a dollar to a poor one. This is shown in Figure 6.8b. Using the same terms as in Figure 6.8a, it can now be seen that $(Y_2 - Y_0) > (Y_0 - Y_1)$. In words, the WTA compensation for not having the good is greater than the WTP to obtain it. In this case the EV value for a good would be greater than the CV value.

We now turn to valuing the loss of a good again assuming diminishing marginal utility as in Figure 6.8b. In this case the starting point includes the public good, which is $\hat{U}$ on the U_1 curve. The individual is here assumed to have a right to the public good. The WTA amount is again greater than the WTP amount, $(Y_2 - Y_0) > (Y_0 - Y_1)$. However, in this case $(Y_2 - Y_0)$ is the CV value and $(Y_0 - Y_1)$ is the EV value and the CV value is greater than the EV value.

Ordinary (observed) demand and consumer surplus

An ordinary demand curve can be viewed as both a marginal WTP schedule and as a marginal benefit schedule. To see this, consider Figure 6.9a. This figure shows, in a series of discrete blocks, a demand schedule for electricity in terms of kilowatt hours (kWh) per week. The consumer values the first kWh at 90 cents, the second unit at 80 cents, and so on down to the fortieth unit at 9 cents per kWh. These amounts represent marginal benefits.



(a) WTP for a good equals WTA for not having it

(b) WTP for a good is less than WTA for not having it

Figure 6.8 Comparison of willingness to pay and willingness to accept

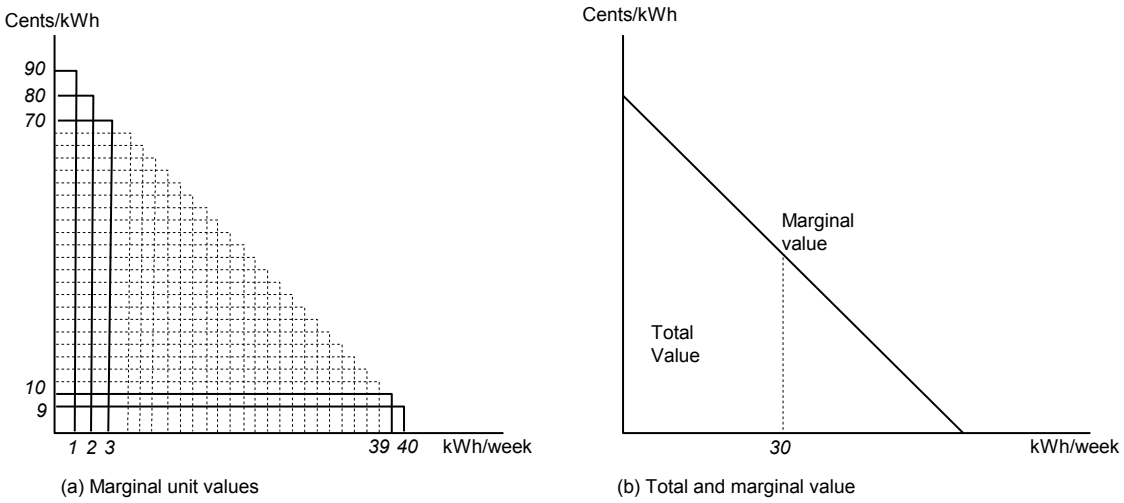


Figure 6.9 Valuing electricity via a demand curve

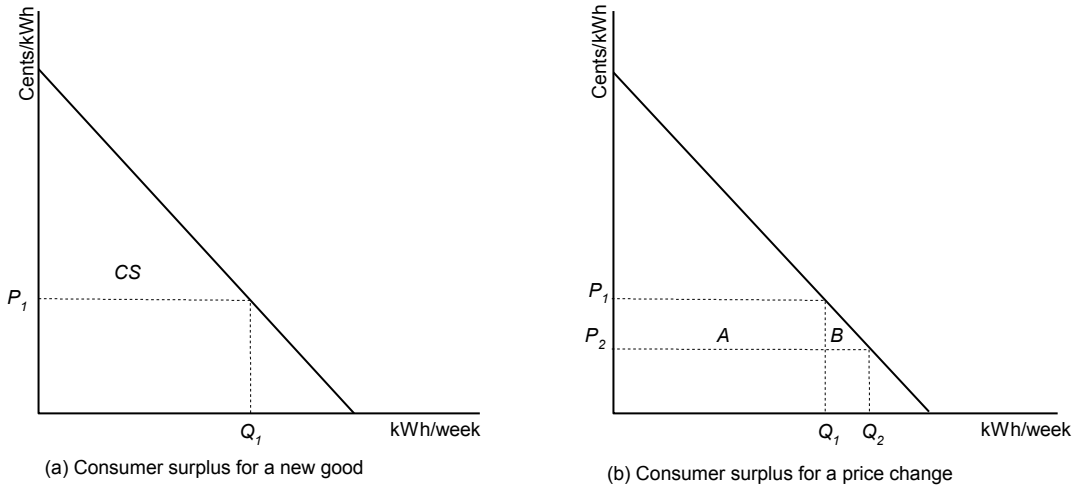


Figure 6.10 Valuation and consumer surplus

Typically, a demand curve is drawn as a continuous downward-sloping curve as in Figure 6.9b. The vertical distance from the demand curve to the horizontal axis is the marginal value of the respective unit. The area under the demand curve represents the total value of electricity consumed.

Consumer surplus
The excess of the benefit that an individual gains from purchase of a good over the amount paid for it

Consumer surplus (CS) is the difference between the maximum amount that an individual is willing to pay for a good and its price. Figure 6.10a shows a consumer's surplus for consumption of electricity. The total CS is the area between her demand curve up to the amount consumed and the market price. Figure 6.10b shows the increase in CS (areas $A + B$) when the price of electricity is reduced from P_1 to P_2. Area A is the gain related to existing consumption (Q_1), which is Q_1 ($P_1 - P_2$). Area B is the surplus associated with the increase in consumption. The consumer gains a large surplus of almost ($P_1 - P_2$) for her first extra units

of consumption but a very small surplus for her last additional units of consumption. When the demand curve can be represented by a straight line, area B equals $0.5\,(Q_2 - Q_1)\,(P_1 - P_2)$. Of course, if the price rises from P_2 to P_1, there would be an equivalent loss of consumer surplus.

Using the change in CS (ΔCS) with an ordinary demand curve as a measure of value, the value of a beneficial change is the maximum amount that an individual is willing to pay to move from their initial level of utility to a higher level. The cost of an adverse change is the maximum amount that he or she is willing to pay to avoid a fall to a lower level of utility. These are payments to achieve, or to avoid, changes in utility levels. They are subtly different concepts to CV or EV payments that would leave an individual on an existing or new utility level respectively.

Consumer surplus, CV and EV

Whereas CV and EV measures hold utility constant, an individual's utility changes as he or she moves along a demand curve. These differences are illustrated in Figure 6.11. Assume an initial price and quantity position of P_0 and Q_0 respectively. The price falls to P_1 and quantity consumed rises to Q_1. The CS is the area between the demand curve and the P_1 line, which equals area $(A + B)$. The CV demand curve holds real income constant at the initial position. This eliminates the income effect and the CV measure of benefit is area A. On the other hand, the EV demand curve holds real income constant at the P_1 Q_1 position. In this case the WTA value for not having the price fall is given by area $(A + B + C)$. If the starting point is the P_1 Q_1 position, the CV and EV curves are reversed.

In summary, CV, EV and CS are equal if the marginal utility of income is constant, as in Figure 6.8a. Equivalently CV = EV if the marginal rates of substitution between public and private goods are constant (if utility is a linear function of different levels of the two goods).

However, when there is a significant change in either (1) real income and the marginal utility of consumption or (2) the relative supply of goods, CV, EV and CS may not be equal.

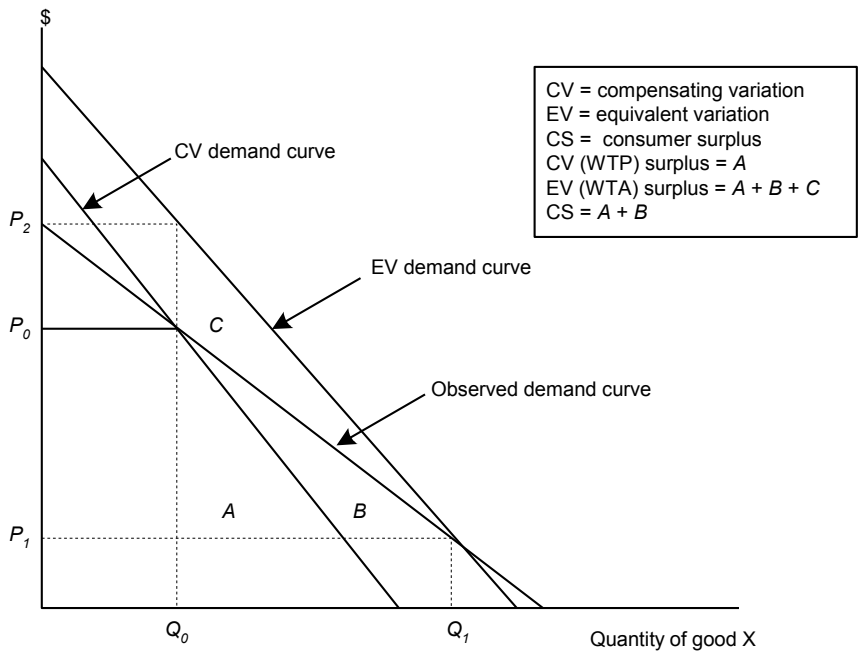


Figure 6.11 Benefits of a fall in price: summarising valuation differences

Table 6.1 Comparison of valuation measures

Income effects	*Beneficial change* *New good or price fall*	*Adverse change* *Loss of good or price rise*
Income effects occur	$CV < \Delta CS < EV$	$CV > \Delta CS > EV$
No income effects	$CV = \Delta CS = EV$	$CV = \Delta CS = EV$

The qualitative relationships between CS, CV and EV measures of value for beneficial and adverse changes are summarised in Table 6.1. When there are no real income effects, the three measures are equivalent. When there are real income effects, CV gives the lowest measure of value for a beneficial change and the highest value for an adverse change. Consumer surplus gives an intermediate value for both beneficial and adverse changes.

Given these potential differences between these three measures of value, which measure should be used? And which measure is most often used?

Choice of valuation principle

The choice of measure of value depends on views about individual rights. If a person has a right to a beneficial change, for example a right to a new hospital service or lower rail prices, the value of the services is the EV amount (the compensation required if the service or subsidy is not provided) rather than a CV or CS amount. On the other hand, if a person loses an existing good, a CV measure of value may be appropriate (the minimum amount that he or she would be WTA as compensation for the loss). Crucially, the choice of measure depends on ethical or political judgements rather than on technicalities.

Where CV and EV values differ considerably, the policy implications may be significant. The CV principle implies that the current situation is an appropriate reference point for policy evaluation and that changes from it must be justified. Suppose that government proposes to clean up a polluted river. Using the CV approach, we would estimate the maximum amounts that the community would be WTP for the clean river and be no better off than before. If the sum of these WTP amounts exceeds the clean-up costs, the river should be cleaned up. On the other hand, if we start from the position that the river should be clean, EV valuation principles apply. We would need to know what people were WTA as compensation for the polluted river. If the sum of WTA amounts exceeds the cost of cleaning up the river, the river should be cleaned up. This could give a quite different outcome.

Or suppose that government is considering whether to provide a higher level of hospital services in a large country town. Under the CV approach we would compare what residents would be WTP for these services and be no worse off with the costs of providing them. If WTP amounts exceeded the costs, the services would be provided. Under the EV approach we would compare what residents would be WTA for not having the services with the costs. The services would be provided if the WTA amounts exceeded the costs.

In practice, CS is the most commonly used valuation method for four reasons. First, it is based on observed demand curves. CV and EV amounts have to be estimated using hypothetical scenarios holding utility constant at the initial or post-change positions. This means either estimating compensated demand curves which take out income effects or running appropriate surveys. Thus, CS measures are generally the most practical measure of value. Estimates of WTA compensation values usually require surveys because these values are not often observed in markets. Second, CS values are often a close approximation to CV or EV values. Willig (1976) estimated that when income effects are small, the differences between WTP and WTA values are less than 5 per cent for most goods. Third, when there are significant income effects, the CS measure is a compromise between the CV and EV measures. Fourth, rights are often not clear. Individuals may have a right to clean air, but also

a right to power supply and travel. Often the same individuals want all these goods. In such cases, it makes sense to estimate how much people are WTP for each potential use of the environment (essentially a CS measure) rather than to start with a preconception about how who owns environmental property and to weight the valuation procedure according to these presumptions.

However, the differences between the measures are not always minor. In a survey of 45 studies of WTP and WTA values, Horowitz and McConnell (2002) found that WTA values can be as much as 10 times WTP values for some health and safety goods or services and are on average seven times higher. The difference between WTA and WTP values rises with the size of the income effect and the income elasticity of demand for the good. It also rises with a scarcity of substitutes for the good. When a good is scarce, individuals require more compensation for its loss.[3]

Differences between WTA and WTP values may also reflect loss aversion as well as on views about rights. Loss aversion occurs when someone places a higher value on a good that is lost than on a gift of the same good, not because of decreasing marginal utility of consumption but because of a dislike of losses, or what Kahnemann and Tversky (1979) famously called the endowment effect. Also, individuals may claim high compensation for a good because they believe they have a right to it and for which they believe they should not have to pay.

Therefore, when individuals lose a good or a property right, a CV measure of value for loss of it may be appropriate. This would be the minimum WTA amount that an individual would accept as compensation for the loss. This is consistent with most popular notions of fairness and indeed the notion of property rights. On the other hand, when individuals have a right to a new good or service, the estimated EV value may be considered more appropriate than the respective CV or CS valuation. This would be the minimum WTA amount that they would accept for not receiving the good or service rather than what they would be WTP to obtain it.

From Valuation Principles to Practice

We now consider some issues in estimating these demand curves for market goods. We also discuss briefly below how to estimate the compensation that individuals are WTA for losing a good (CV values) or the compensation required for not receiving a good to which they have a perceived right (EV values). In Chapter 11 we discuss methods for valuing non-market goods.

Estimating ordinary demand curves

For some policy issues it may be sufficient to estimate only one point on the demand curve: the price that consumers are WTP a particular quantity of goods. However, for many valuations, estimates of the whole demand curve or major parts of it are required.

Many textbooks describe methods for estimating ordinary demand curves (see, for example, Gujarati, 2003). Here we note two major issues: the multivariate nature of the demand function and the identification problem of distinguishing demand and supply.

A demand curve is simply a relationship between the quantity of a good x demanded (Q_x^d) and its price (P_x). However, care must be taken in estimating a demand curve simply by regressing observations of Q_x against P_x because many factors may influence quantity demanded. To estimate the effects of price on demand, it is often necessary to estimate a general demand function that includes other variables such as income and population. For example, if a rail agency wishes to estimate the demand for a railway trips, the agency could estimate a multiple regression equation in which the number of rail trips in a specified period

[3] Hanemann (1991) shows that technically the difference between WTA and WTP depends on the ratio of the income effect to the substitution effect.

is regressed against such variables as rail fares, trip times and service frequency, bus fares, road travel speeds, household income, car ownership and populations at various distances to rail stations.

Evidently, estimating a demand function may require considerable data. Cross-sectional analysis usually provides a richer data set than time series analysis. However, cross-sectional analysis does not always provide data with differing prices that are critical to demand analysis. Time series analysis may provide more information on price changes and their effects.

The second main issue in estimating a demand function (the identification problem) is the problem of estimating the parameters of a structural equation when we observe equilibrium positions. Suppose we observe two prices for a good and the quantities purchased at each price. We may be observing two points on a demand curve or on a supply curve or two equilibrium points reflecting shifts in demand and supply. If we estimate the relationship between quantity and price, we need to know whether we are estimating a demand or a supply function. The identification problem may be resolved if we have additional variables in either the demand or the supply function that allow the two curves to be differentiated. If income is included in the demand function along with price, a change in income will cause the demand curve to shift and each shift in the demand curve creates a new intersection of demand and supply, essentially mapping out a supply curve. Conversely, if a variable measuring weather conditions is included in the supply function, a change in this variable would shift the supply function and the equilibrium points would indicate a demand curve. When demand and supply are determined simultaneously by price, more complex statistical methods, such as two-stage least squares, are required to estimate the demand or supply curve.

Fortunately, in the public sector the supply of services is often exogenous and independent of price and the observed quantity and price data can be assumed to represent a demand curve. However, specification of the demand curve as linear, log-linear or another functional form requires careful analysis and can affect estimates of demand and consumer surplus.

Finally, as a practical matter, when valuing changes to existing services, it is often possible to draw on price and income elasticities that have been estimated in research studies. For example, many research papers have shown that the price and cross-price elasticities for public transport services are usually low (Goodwin, 1992). For new goods, specific market research and econometric work may be required.

Estimating consumer surplus and compensating and equivalent variations

Most econometric estimates of price effects include substitution and real income changes and are thus estimates of consumer surplus. Estimating compensated demand curves, or CV and EV amounts, is more complicated. To estimate a compensated demand curve the analyst must estimate how quantity demanded varies with price holding real income constant. Once the compensated demand curve is estimated, estimating the relevant area under the demand curve would be straightforward. However, we often lack sufficient data to estimate quantity demanded as a function of changes in prices holding income constant.

The following example shows how CS, CV and EV may be estimated in a specific case. Suppose that a household spends $10 000 per annum, representing 25 per cent of its income, on renting housing. To convert this figure of $10 000 into prices and quantities, suppose that the price of housing is $100 per m^2 per annum and that the household rents 100 m^2. Government decides to subsidise rental housing by 10 per cent. Thus the cost of housing falls to $90 per m^2 and, assuming a price elasticity of demand of –1.0, the household purchases 110 m^2 of housing. This is shown in Figure 6.12.

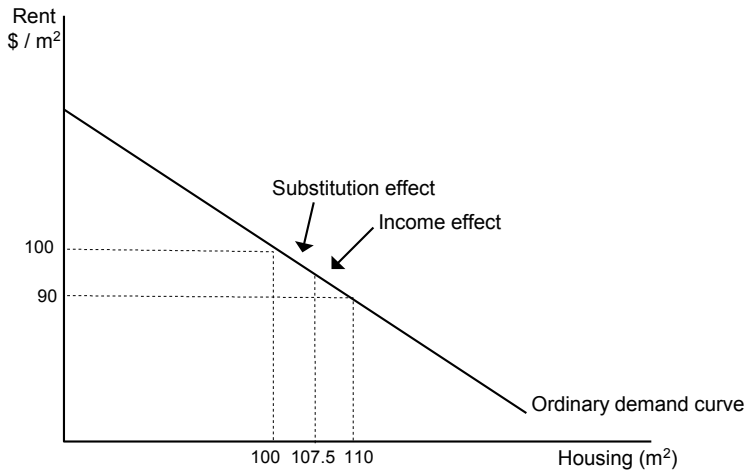


Figure 6.12 Estimating CS and CV for a price fall

Approximating the ordinary demand curve as a linear curve, the benefit to the household measured by the change in consumer surplus can be calculated as:

$$\Delta CS = (100 \times 10) + (10 \times 10 \times 0.5) = \$1050$$

To estimate the CV associated with the fall in price, we need to take out the income effect. Suppose that the income elasticity of demand for housing is 1.0. Because the household receives (initially) a subsidy of $1000 per annum, its real income has increased by 2.5 per cent from $40 000 to $41 000. It follows that the real income effect was responsible for a 2.5 per cent increase in consumption of housing and the substitution effect for the balance of 7.5 per cent. Accordingly, the compensating variation can be calculated as:

$$CV = (100 \times 10) + (10 \times 7.5 \times 0.5) = \$1037.50$$

Now let us calculate the EV. In this case, we have to put the household in its real income position after the price fall and ask what amount would compensate it for not having the lower price. Given that the substitution effect accounts for $7.5m^2$ of the increase in consumption due to the subsidy, to retain the same real income as with the subsidy, the household would consume 102.50 units. Thus the equivalent variation would be:

$$EV = (102.5 \times 10) + (10 \times 7.5 \times 0.5) + \$1062.50$$

As predicted, with a fall in prices, $CV < \Delta CS < EV$. However, even when expenditure is 25 per cent of total income and the price change is 10 per cent, the differences are small. CS is only 1.0 per cent greater than CV and 1.0 per cent smaller than EV. The percentage differences would be larger if expenditure were a higher percentage of income or the income elasticity higher. Zerbe and Dively (1994, p. 113) provide a table showing how these percentage differences vary with the expenditure share of total income and income elasticities.

Box 6.4 shows how differences between CV, ΔCS and EV can be derived formally from a utility function. This demonstrates the rigour of the approach. In general CV or EV values can be estimated by making plausible assumptions about the nature of the (income) utility function. However, because utility functions are not observed, the values derived from this approach may be open to question.

Box 6.4 Deriving consumer surplus, CV and EV from a utility function: an example

Say the consumer has the same utility function as in Box 6.1: $U(q_x,q_y)=q_x^{0.7},q_y^{0.3}$. Initially she faces prices (1, 1) and has an income of \$100. Then the price of x increases to 2. What are the ΔCS, EV and CV?

First we need to derive the demand functions. In Box 6.1, we obtain:

$$Q_x = 0.7M/p_x \text{ and } q_y = 0.3M/p_y$$

Using this formula, consumer demand changes from $(q_{x1},q_{y1}) = 70, 30$ to $(q_{x2},q_{y2}) = (35, 30)$.

Adopting a linear approximation of the demand curve, the loss of consumer surplus equals:

$$(35\times 1) + (35\times 1\times 0.5) = 52.5$$

To calculate the CV we ask how much money would be necessary at prices (2, 1) to make the consumer as well off as she was when consuming the bundle (70, 30)? If the prices were (2, 1) and the consumer had income M, we can substitute into the demand functions to find that the consumer would optimally choose the bundle $(0.7M/2, 0.3M)$. Setting the utility of this bundle equal to the utility of the bundle (70, 30), we obtain:

$$(07M/2)^{0.7}(0.3M)^{0.3} = 70^{0.7}\times 30^{0.3}$$

Solving for M gives us $M = 162$ (approximately). Hence to make the consumer as well off after the price rise as she was before it, she would need about 162 – 100 = \$62 of additional income.

To calculate the EV we ask how much money would be necessary at price (1, 1) to make the consumer as well off as she would be consuming the bundle (35, 30). Letting M stand for this amount of money and following the same logic as before,

$$(07M/2)^{0.7}(0.3M)^{0.3} = 35^{0.7}\times 30^{0.3}$$

This means that $M = 62$ approximately.

Thus if the consumer had an income \$62 at the original prices, she would be just as well off as she would be facing the new prices and having an income of \$100. The EV in income is therefore about 100 – 62 = 38.

As we predicted in Table 6.1, for a price rise,

$$CV > \Delta CS > EV$$

Concluding observations

Identifying, valuing and meeting individual preferences lies at the heart of public economics. In this chapter we have focused on the role of the ordinary demand curve because it is the foundation on which most economic valuations of individual preferences are based. These demand curves can be estimated for most market goods.

However, whether individuals should have to pay for goods and services or be compensated for not having them depends on views about individual rights. Given a decreasing marginal utility of income, WTA values are generally higher than WTP values though the difference is generally small for small changes in income. This means that it may sometimes be appropriate to estimate compensating or equivalent (CV) or (EV) values which are based on compensated demand curves.

Also, there are generally no observed demand curves for non-market goods. As we will see in Chapter 11, other valuation methods are often required to estimate the values of non-market goods. These include other revealed preference methods that analyse individual behaviour in various contexts to infer implicit valuations and stated preference methods that employ survey techniques to elicit the values of individual preferences. However, whichever valuation method is employed, it should be consistent with the valuation principles described in this chapter.

Summary

- The economic approach to public policy is based on valuations of individual preferences. These preferences are generally inferred from the choices that individuals make between various goods, including market and non-market goods.
- Formally, preferences can be represented by utility functions or indifference curves. Demand curves can be derived from utility functions in combination with budget constraints.
- Demand curves supply information on both the total value of a good and the marginal value of an extra unit of a good. Thus, estimates of demand curves or of relevant parts of demand curves are basic to estimates of the values of market goods.
- However, there are two kinds of demand curves. Ordinary (observed) demand curves include substitution and real income effects. Compensated demand curves show demand as a function of price holding real income constant.
- Valuations based on ordinary demand curves use the concept of consumer surplus.
- Valuations based on a compensated demand curve use the concepts of compensating or equivalent variation.
- The compensating variation (CV) value of a good is the maximum amount that may be taken from someone and leave them as well off with the good as without it. The equivalent (EV) value of a good is the minimum amount that someone will accept for not having the good and leave them as well off without the good as with it.
- Turning to losses of goods, the CV of a loss is the minimum amount that someone will accept as compensation and be as well off with the loss as without it. The EV of a loss is the maximum amount that someone will pay to stop the loss and be no worse off than with the loss.
- For most purposes, especially when income effects are small, changes in consumer surplus are a good measure of the value of individual preferences. However, when property rights are important and income effects are large, it may be important to estimate compensating or equivalent variation values.

Questions

1. According to Oscar Wilde, a cynic is a person who knows the price of everything and the value of nothing. Is this a good description of an economist?
2. What assumptions are necessary for supposing that individual preferences can be inferred and valued from the choices that individuals make?
3. Suppose that university fees increase by 20 per cent. How would the income and substitution effects contribute to the change in the quantity of university education demanded?
4. Explain why the marginal rate of substitution varies along a typical convex indifference curve. What is the implication for the relevant demand curve?
5. Economists often assume that for small changes in income, the income utility function can be regarded as linear. Is this a reasonable assumption? What are the implications of this assumption?
6. If Amy has a Cobb–Douglas utility function of the form $u(x, y) = q_x^{0.5} q_y^{0.5}$ where q_x and q_y are the quantities of two goods x and y, and the price of x is twice the price of y, what fraction of her income will she spend on goods x and y?
7. What is the difference between compensating and equivalent variations measures of value? When would it be appropriate to use one or other of these measures rather than a consumer surplus measure?
8. Can willingness to accept compensation values for losses of goods sometimes be observed in market transactions?
9. Ben has an income of $30 000 and spends $10 000 a year on housing. He rents 100 square metres at $100 per square metre per annum. The government now subsidises the rent by 20 per cent. Ben's income elasticity of demand for housing is 1.0 and his price elasticity of demand is –1.0. What is the consumer surplus, CV and EV value of the subsidy to Ben?
10. Amy has the same utility function as in Question 6. She has an income of $1000 and faces prices of $10 for each good x and y. The price of x increases to $20. What is the change in consumer surplus, CV, and EV value of this increase in price?
11. George Best, a famous Manchester United football player, reputedly said: 'I spent 90% of my money on booze, women and fast cars; the rest of my money I wasted'. Does this indicate bounded rationality?

Further Reading

Diamond, P and Vartiainen, H., eds (2007) *Behavioural Economics and its Applications*, see especially Chapter 2 by Bernheim and Rangel, 7–84, Princeton University Press, Princeton, NJ.

Dupuit, J. (1844) 'On the measurement of the utility of public works', translation by R.H. Barback in *International Economic Papers No. 2*, Macmillan Press, 1952, 83–110.

Gruber, J. (2016) *Public Finance and Public Policy*, Chapter 2, Worth Publishers, New York.

Jha, R. (1998) *Modern Public Economics*, Chapter 1, Routledge, London.

Sen., A. (1973) 'Behaviour and the concept of preferences', *Economica*, 40, 241–259

Taylor, J. B. and Frost, L. (2002) *Microeconomics*, J.B. Wiley, Brisbane.

Varian, H.R. (2006) *Intermediate Economics*, 7th ed., Chapters 3–6, W.W. Norton, New York.

Chapter

7

Social Welfare and Economic Evaluation

The interest of the community then is—what? The sum of the several members who compose it.

Jeremy Bentham

The Nature of Public Choices ♦ Opportunity Sets and Social Welfare Functions ♦ Utility and Income ♦ Utilitarian Social Welfare Functions ♦ Working with Income ♦ Social Welfare and Cost-Benefit Analysis

In the last chapter we discussed how to value individual preferences. In this chapter we discuss how to combine these preferences (utilities) into a measure of collective welfare, which we call social welfare. This measure of social welfare is intended to be a criterion for determining public policy. The greater the social welfare, the more desirable is the relevant policy.

Following the early utilitarian, Jeremy Bentham (1789), social welfare would be viewed as the simple addition of individual valuations of utility. However, individual utilities are not readily comparable or additive. Of special importance, society may want to give greater weight and assistance to those members with less welfare (utility). Moreover, if our measure(s) of social welfare are to guide policy making, they must be practical. In this chapter we confront the conceptual difficulties of estimating social welfare and seek to generate practical measures of welfare for use in policy making.

In the first two sections we discuss the nature of public choices, opportunity sets that represent these choices and measures of social welfare (social welfare functions). The third and fourth sections discuss the relationships between individual utility and income and between individual utility and social welfare respectively. The last part of the chapter discusses how economists work with incomes or dollar values, typically using cost–benefit analysis, to evaluate public policy and how these approaches relate to the concept of maximising social welfare.

The Nature of Public Choices

It might be thought that choice is easy in win–win situations when all parties gain and almost as easy when one party gains and no one loses. However, even in win–win cases, someone may feel relatively deprived by the outcome. Several parties may gain from an inheritance but argue vehemently about the perceived unfairness of the distribution! Or suppose that government decides to cut taxes. While the cuts may benefit most individuals and disadvantage no one, the distribution of the cuts may be strenuously debated. Even in win–

win situations, the preferred allocation of the gains must be determined. This almost always involves determining not only the economic impacts of the allocations but also the relative value of the benefits to the individuals concerned.

It might be supposed likewise that cost-effectiveness decisions (i.e. decisions about how to achieve a given target at least cost) are technical decisions that do not involve normative judgements. Again, this is rarely so. Take, for example, such routine decisions as how to service public hospitals or provide IT services, which may be done by in-house labour or outsourced. This often involves gains or losses of economic rents for some workers so there are winners and losers from any such decision and government may not adopt the least-cost option. Or suppose that government wishes to find the least-cost way to achieve an environmental target such as a maximum level of carbon monoxide or nitrogen oxides in cities. The alternatives will impose various benefits and costs on city residents, commuters, tourists, businesses and so on.

The existence of winners and losers is even more evident when government allocates expenditures. Government may fund hospitals rather than schools or fund both by raising taxes and so reducing private consumption. It may build hospitals or schools in one location rather than another. In almost all social assistance, there are gainers and losers. Also, income redistribution generally means both higher taxes and lower output as some taxpayers and income recipients substitute leisure for work.

It follows that the fundamental problem of public choice is to determine how to assess, compare and aggregate these gains and losses. The assessment is based on valuations of individual preferences discussed in the last chapter. Comparison and aggregation of the costs and benefits borne by individuals are discussed in this one.

Opportunity Sets and Social Welfare Functions

Typically, a policy or project evaluation sets out the options, assesses the positive and negative impacts on individuals and determines which option gives the maximum surplus of gains over losses. Here we set out a more formal evaluation procedure, based on the notion of maximising social welfare. While this process is somewhat abstract, it provides a foundation for other more practical evaluation methods.

In this more formal process, the set of options is described as an **opportunity set**. An opportunity set shows the feasible choices available. The concave curves in Figure 7.1 show three opportunity sets. These are opportunity sets for quantities of hospital and housing services, for total income represented by GDP and equality of income (which may be measured, for example, by the Gini coefficient),[1] and for the welfare (utility) of two persons, Amy and Ben. The slope of the opportunity set at any point shows the marginal trade-off available in each case. In practice policy makers are confronted with the kind of choices shown in panels (a) and (b) rather than with a utility map as in panel (c). However, because the end objective of policy making is generally to maximise the welfare of individuals rather than health care or houses or income per se, much of the fundamental analysis of welfare economics is conducted in terms of individual utilities.

To choose the preferred point on any opportunity set, a ranking criterion is needed. This criterion must have two attributes. First, it must be expressed in the same units as the opportunity set. For example, if the opportunity set is composed of various levels and distributions of income, the ranking criterion must reflect these variables. Second, the ranking should reflect social welfare. A higher ranking should indicate greater social welfare and a socially preferred outcome.

[1] The Gini coefficient is described in Chapter 20. A coefficient of 0 is perfect equality and a score of 1 is complete inequality.

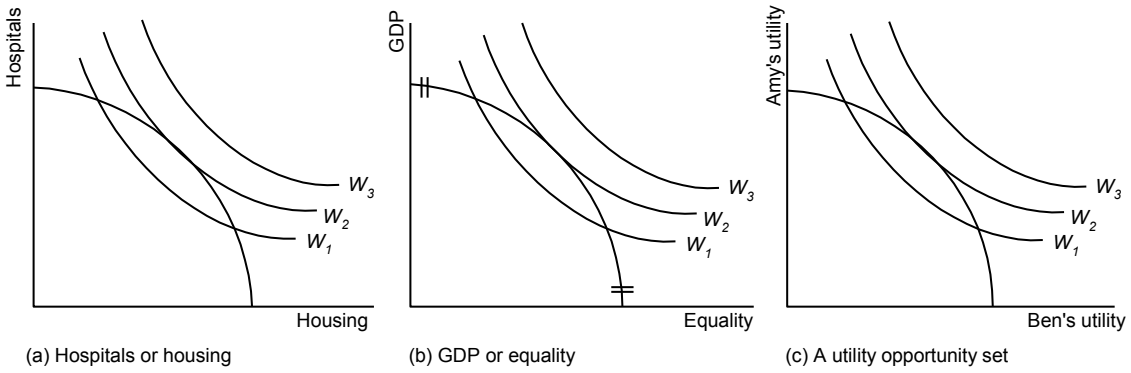


Figure 7.1 Opportunity sets and social welfare functions

In economics, the ranking criterion is usually called a social welfare function. A **social welfare function** (SWF) is any function that defines social welfare by its components. These components are typically individual utilities, but these utilities may be represented by economic or social variables that drive utility. An SWF provides a means to rank social states: the more social welfare the better. Ideally the SWF would rank all possible social states.

In this chapter we focus mainly on utilitarian SWFs with the general form:

$$W = f(u_i) = f(u_1, u_2, \ldots u_n) \tag{7.1}$$

where W is social welfare, u is individual utility (or welfare) and there are $i = 1 \ldots n$ individuals in society. In a utilitarian SWF, social welfare is a function of the utility of the members of society.[2] We focus on utilitarian SWFs because the fundamental aim of policy is the collective welfare of individuals. However, for practical purposes we will also need to consider other ways to rank policies.

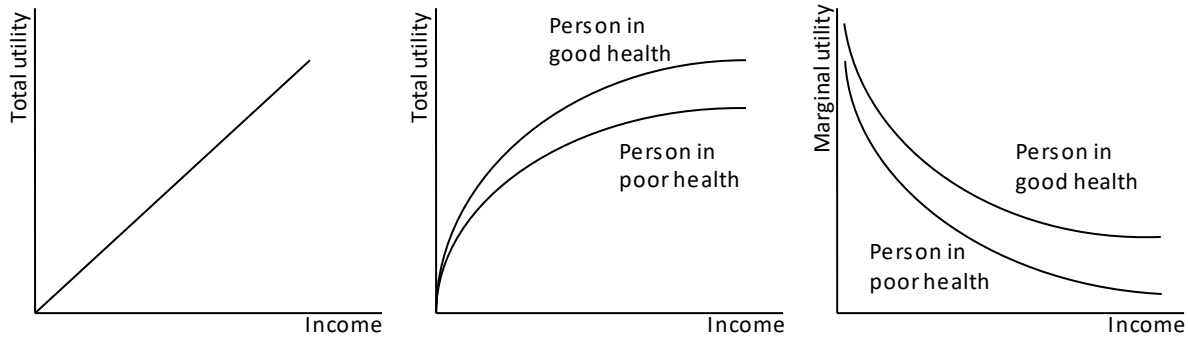
Formally, the aim is to maximise social welfare (W) subject to the constraints of the opportunity set. In each panel of Figure 7.1, there are three convex iso-welfare curves (W_1, W_2, W_3), with W_3 representing the highest level of welfare. An iso-welfare curve shows points of equal social welfare (between which society is indifferent). The slope of the iso-welfare curve at any point shows the marginal rate at which society is willing to trade off the components that make up welfare. Thus, in Figure 7.1c, the slope shows the marginal rate at which society is willing to substitute Amy's utility for Ben's utility (the social marginal rate of substitution, MRS). Welfare is maximised when the iso-welfare curve (W_2) is at a tangent to the opportunity set. At this point, the slopes of the iso-welfare curve and the opportunity set are equal. The social MRS of Amy's utility for Ben's is equal to the marginal rate at which Ben's utility can be transformed into Amy's on the utility possibilities frontier (UPF).

In Figure 7.1b, the opportunity set is various combinations of GDP and equality, and there are again three iso-welfare curves. Welfare is maximised (again on the W_2 curve) when the marginal rate at which society is willing to give up income in return for more equality is equal to the marginal rate at which income has to be sacrificed to achieve extra equality.

Utility and Income

A major problem for policy evaluation is that we want to improve the welfare (utility) of individuals but our main measure of utility is income. Therefore, we need to understand the relationship between utility and income, which we introduced in the last chapter (Figure 6.8).

[2] This form of social welfare function is often called a Bergson–Samuelson SWF.



(a) Linear utility function with constant marginal utility

(b) Total utility of income

(c) Marginal utility of income

Figure 7.2 Utility and income

Figure 7.2a shows a linear utility function where the marginal utility of income is constant. If everyone were to have the same linear utility function, there would be a simple translation of income into utility. A dollar increase in income would have the same marginal value to all individuals regardless of who gained or lost the marginal dollar.

However, there are significant problems with this simple equation between income and utility. First, for most people income has decreasing marginal utility. That is, each extra dollar has a positive but decreasing value of utility. This is shown in Figures 7.2b and 7.2c. It follows that, if everyone has a similar utility function, a transfer of a dollar from a rich person to a poor one would always raise collective utility (assuming no loss of output due to the redistribution). If the redistribution involves an economic loss, we would also need to know the shape of the utility function to determine the optimal rate of dollar transfer between any two individuals.

Second, utility does not depend only on income. It may reflect health, family circumstances or a myriad of factors. Taking only differences in health, for any given level of income and personality, someone in good health is likely to have a higher level of utility than someone with poor health. Also, as shown in Figure 7.2c, an individual in good health may gain more marginal utility from additional income than does a person in poor health. Differences in utility functions complicate policy making. In this case, aggregate utility could be increased by transferring a dollar from the sick person to the healthy one. But this may not reflect social preference which could support more support for the sick person who has a lower absolute level of utility as well as lower marginal utility.

More generally, social valuations of relative need may differ from relative individual utility. Society may take the view that the happier that individual A is relative to B, the less socially valuable is an addition to A's happiness compared with an addition to B's happiness. Suppose that A and B have similar utility functions but that A has more income than B and gets x per cent of B's marginal utility from an extra dollar (where x is less than 100). Any policy that increases B's income by one dollar and reduces A's income by less than $1 / (1 - x/100)$ dollars increases their joint utility. But society may judge that more, or less, weight should be placed on the relative value of a dollar going to B (the less well-off person) than the value associated with their relative marginal utilities. Thus, society may consider that the lower the initial *level* of happiness, the more value should be given to additional units of happiness.

Measurement and interpersonal comparisons

Given that individual utility is a critical component of most SWFs, ideally utility functions would be measurable and comparable. These two requirements are described as a need for cardinality and interpersonal comparability respectively.

Of course, individual utility cannot be measured in the same scientific way that temperature or blood pressure can be measured. The alleged lack of both strict cardinality and a scientific basis for making interpersonal comparisons of utility has led some economists (most famously Robbins, 1938) to conclude that viewing social welfare as an aggregate function of individual utilities is unscientific and not useful. This warning is important, but the conclusion seems exaggerated. As Little (1957) pointed out, people commonly compare the happiness and quality of life of different individuals and conclude that one person is happier or has a higher quality of life than another, and we base our actions on these views. The fact that such comparisons cannot be precise does not render them meaningless or useless.

More recently, economists have undertaken several studies attempting to relate happiness (utility) to income (see Layard, 2006; Layard *et al*., 2008). To estimate the utility function it is typically assumed that the elasticity of marginal utility with respect to income is constant. This elasticity (ρ) is the percentage change in marginal utility associated with a one per cent change in income. This assumption implies that utility (u) is given by:

$$u = \begin{cases} \dfrac{y^{1-\rho}}{1-\rho} & \rho \neq 1 \\ \log(y) & \rho = 1 \end{cases} \qquad (7.2)$$

where y is income.[3] It follows that the ratio of the marginal utilities of two people is given by:

$$\frac{\partial u^B / \partial y}{\partial u^A / \partial y} = \left(\frac{y^A}{y^B} \right)^{\rho} \qquad (7.3)$$

If A has twice the income of B and p = 1, then A obtains half the marginal utility that B obtains from each extra dollar of income. Drawing on four large cross-sectional surveys of subjective happiness and two panel surveys, with data from over 50 countries and time periods from 1972 to 2005, Layard *et al.* (2008) estimated that the value of p varied from 1.19 to 1.36, with a best mean estimate of 1.26.[4] This implies that as income rises by 1 per cent, the marginal utility of income falls by 1.26 per cent. The authors found similar elasticities for sub-groups of the population.

These findings suggest that interpersonal comparisons of utility can be meaningful and that social welfare can be viewed as a function of individual utilities. However, this does not obviate the need for making critical ethical decisions about the nature of social welfare. It is still necessary to determine how changes to relative utilities should be weighted in an SWF, which is a normative decision.

---

[3] Equation 7.2 is known as the Constant Relative Risk Aversion function. Other common and simpler utility function specifications include $U = \log(y)$ and $U = y^p$, where $0 < \rho < 1$.

[4] Layard (2008) used a slightly modified version of the CRRA function to estimate a more realistic model whereby people need a minimum amount of income before they start receiving positive utility.

Utilitarian Social Welfare Functions

Utilitarian SWFs generally have four features. First, they depend only on the utility of individuals. Such a SWF is sometimes described as "welfarist". Rights, such as the right to freedom of speech or worship, matter only in so far as they affect individual utilities. Second, social welfare is assumed to increase with each individual's utility level, other things being equal. Third, iso-welfare curves are generally assumed to be strictly convex. This reflects the egalitarian ethic that inequality in utilities among individuals is socially undesirable. Fourth, it is generally assumed that everyone counts equally—it does not matter who enjoys a high or low level of utility. This is the principle of anonymity or symmetry.

In the following discussion, we identify various forms of utilitarian SWFs and a general SWF function that embraces all forms, depending on the parameter values applied.[5]

The additive utilitarian social welfare function

The most commonly discussed utilitarian SWF is the simple additive utilitarian SWF. The idea underlying this SWF is that society should aim to maximise the total happiness of the community.[6] In this formulation, social welfare is the unweighted sum of the utilities of all individuals:

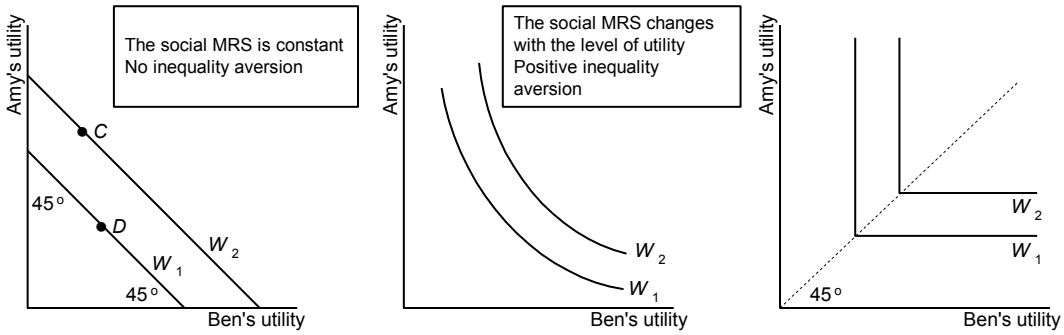
$$W = u_1 + u_2 + \ldots u_n = \sum u_i \tag{7.4}$$

Note that individual utility is usually assumed to be a function not only of income (or goods consumed, x) but also of leisure (l):

$$u_i = f(y_i, l_i) \text{ or } = f(x_i, l_i) \tag{7.5}$$

Also, in most policy or project evaluations, goods include market and non-market goods, such as health and environmental goods.

Figure 7.3a shows the implied social indifference (iso-welfare) curves. With an unweighted additive utilitarian SWF, iso-welfare curves are linear with a 45° angle to each axis. A marginal (or unit) increase in an individual's utility always has the same social value,



(a) Additive utilitarian SWF (b) Weighted additive utilities SWF (c) Rawlsian SWF

Figure 7.3 Social indifference (iso-welfare) curves

[5] In Chapter 9 we discuss another way of ranking social states based on individual preferences over states, drawing typically on voting methods, instead of on individual utilities.

[6] Traditionally, the additive utilitarian SWF was described as 'the utilitarian SWF'. It is also called the classical or Benthamite SWF, named after the 19th century school of utilitarian philosophers founded by Jeremy Bentham. Bentham argued that 'the greatest happiness of the greatest number is the foundation of morals and legislation'.

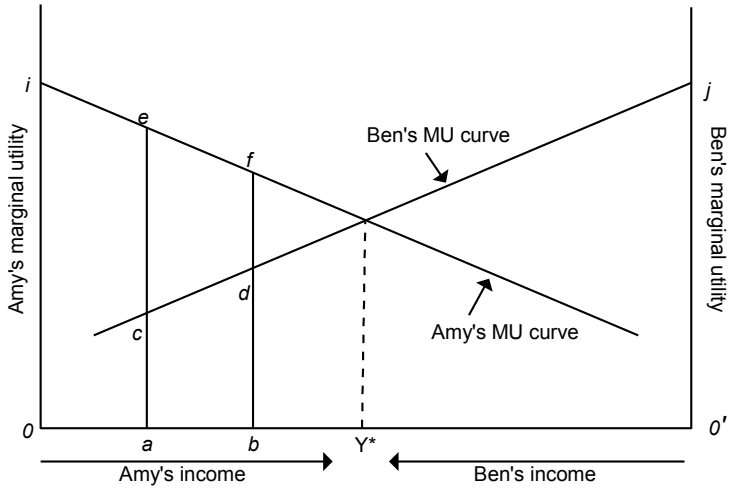


Figure 7.4 How equality of income can maximise social welfare

regardless of the level of their utility. However, this iso-welfare curve does not imply that a dollar to each person has the same social value. If Amy and Ben have a similar decreasing marginal utility of income function, a marginal dollar would provide more utility to whoever has less income. In *income* space, the iso-welfare curves would be convex.

The unweighted additive utilitarian SWF has some attractive features. It is easy to understand. It is democratic in that a marginal increase in the utility of each individual has equal weight. Also, it is often viewed as fair because it justifies redistributing income from high-income to low-income individuals (who usually have greater needs and a higher marginal utility of income). Indeed, under certain assumptions (as discussed below) it justifies complete equality of income distribution.

This is shown graphically in Figure 7.4. Here there is a fixed amount of income (OO') to be divided between Amy and Ben, who are assumed to have similar utility functions. Initially, say Amy has $O'a$ income and Ben's income is $O'a$. Then according to the utilitarianism calculus, total welfare is $Oiea + O'jca$. If ab income is transferred from Ben (the richer) to Amy (the poorer), total welfare will be $Oifb + O'jdb$, which increases welfare in the society by $cefd$. When Amy has OY^* amount of income and Ben has $O'Y^*$ income, their marginal utilities are equal and their total utility is maximised. Because Amy and Ben have the same utility function, total utility is maximised when total income is divided equally between them.

Nevertheless, the unweighted additive utilitarian SWF is often criticised for being concerned only with total welfare and not with its distribution. In Figure 7.3a, C is on a higher social welfare curve than D and, with an additive utilitarian SWF, would be preferred to D. But many people would disagree and prefer D to C. They may argue that income should be transferred to individuals with lower absolute levels of utility, without regard to the marginal utility that they obtain from income or to the possible loss of aggregate utility. With a simple additive utilitarian function, a transfer of income from a poor unhealthy individual who has little enjoyment in life to a rich healthy person could increase total utility. But this may not be regarded as desirable. An example of such a scenario is given in Box 7.1 overleaf.

Introducing equity explicitly into social welfare functions

Equity can be introduced explicitly into a SWF in various ways, for example by a weighted additive function. A multiplicative form of SWF also allows for the distribution of utility. In

Box 7.1 An additive utilitarian social welfare function may not be desirable

Say in a two-person (Ben and Amy) economy, total fixed income of \$100 is split between them. For Ben, the marginal utility of income is: $MU_B = 400 - 2y_B$ while for Amy the marginal utility is $MU_A = 400 - 6y_A$, where y_B and y_A are amounts of income to Ben and Amy respectively. If the welfare function is simple additive, to maximise W set marginal utilities equal subject to the constraint that $y_B + y_A = 100$.

So, $400 - 2y_B = 400 - 6y_A$.

Substituting $y_A = 100 - y_B$

$2y_B = 6(100 - y_B)$.

Therefore, $y_B = 75$, $y_A = 25$.

This shows that if, for the same amount of income, Ben gets more marginal utility than Amy, the welfare-maximising utilitarian solution is to provide Ben with more income. The additive utilitarian SWF, which aims to maximise the sum of individual utilities, may not be an attractive ethical objective.

such a SWF, welfare is a function of the product of individual utilities. This can be expressed in logarithmic form with social welfare a function of the sum of the natural log of individual utilities.

More generally, we may adopt a weighted multiplicative utilitarian SWF:

$$W = u_1^{a_1} \times u_2^{a_2} \times \ldots \times u_n^{a_n} = \prod u_i^{a_i} \tag{7.6}$$

where the a_i are weights between 0 and 1 assigned to each individual's utility and Π represents multiplication. Equation 7.6 allows us to give greater weight to increases in the utility of people with greater needs.

Equation 7.6 is represented by a convex iso-welfare curve (see Figure 7.3b). Convexity implies that society prefers an even distribution of individual utilities to an uneven one. As inequality increases, society is willing to accept a decrease in the utility of the poor only if there is a larger increase in the utility of the rich. The acceptable trade-off depends on the degree of inequality.

Another SWF that allows even more strongly for welfare distribution is the Rawlsian maximin SWF. Rawls (1971) proposed that social welfare should depend only on the welfare of the least well-off individual (or group of individuals) in society. This can be interpreted as:

$$W = \min (u_1, u_2, \ldots, u_n) \tag{7.7}$$

This implies an L-shaped iso-welfare curve (Figure 7.3c). An increase in the utility of anyone other than the least well-off person makes no difference to the welfare of society.

Rawls argued that people could choose a fair society only if they were unaware of their position, income-earning capacity and so on in society (sometimes described as a veil of ignorance). He held that individuals who did not know their position, being risk averse, would choose a society in which the welfare of the least well-off person was as high as possible. This approach attempts to generate an unbiased ethical ranking of social states. However, it has been questioned on the ground that the selfish views of individuals in an *original state*[7] do not necessarily have ethical content. Nor is there any evidence that individuals so placed would be so risk averse as to be concerned only with the welfare of the least well- off person.[8] The exclusive emphasis on the welfare of the least well off individual is limiting. The Rawlsian SWF is indifferent to the welfare of anyone else.

[7] The position was original because no social or political system existed, and therefore no individual in the imagined position had a known place in the society. Rawls described this as the principle of justice.

[8] Harsanyi (1955) argued that if individuals in their original position were to maximise their expected utility, they would choose an additive utilitarian SWF.

A general social welfare function

All the SWFs described above can be viewed as special cases of a general SWF.

$$W = \frac{1}{1-\varepsilon}\sum_i \left(u^i\right)^{1-\varepsilon} \tag{7.8}$$

where ε is a parameter that reflects social concern for equality of welfare.[9] With this formulation, the elasticity of substitution along each iso-welfare curve is constant and given by $1/\varepsilon$. This means that the ratio of the proportional change in marginal utilities to the proportional change in absolute utilities is constant.

When $\varepsilon = 0$, there is no explicit concern for equality. Each marginal unit of utility has equal value regardless of the individual's level of utility. Equation 7.8 then reduces to the simple unweighted additive utilitarian SWF (Equation 7.4) and the iso-welfare curve in Figure 7.3a. When ε is positive, increases in individual utility are transformed into less than proportional increases of

$$\frac{1}{1-\varepsilon}\left(u^i\right)^{1-\varepsilon}$$

This implies that more weight is attached to a marginal increase in utility for someone with low utility than for someone with high utility. As ε increases, the weight for equality increases. In the limit $\varepsilon \to \infty$ and Equation 7.8 reduces to the Rawlsian SWF, Equation 7.7.

Table 7.1 shows how the social valuation of marginal utility may vary with differences in utility levels and social values of ε. If Ben's level of utility is M times that of Amy's, the weight attached to an additional unit of utility for Amy is M^{ε}. Suppose that $M = 2$ (Amy has half the absolute level of utility of Ben). If $\varepsilon = 0$, a unit increase in Amy's utility has the same social value as a unit increase in utility for Ben. If $\varepsilon = 2$, the social value of marginal unit of utility for Amy is four times the social value of a marginal unit of utility for Ben.

Note that these **social** valuations of utility are *additional* to any differences in the marginal private utilities of consumption. Suppose that a dollar provides Amy with one extra unit of utility and Ben with only 0.5 extra units of utility. And suppose that $M = 2$ and $\varepsilon = 2$. The net social value of a transfer of a dollar from Ben to Amy is then 3.5 units, because Amy's one unit attracts a social value of 4 units, whereas Ben forgoes 0.5 units of utility.

From utilitarian social welfare functions to public policy

SWFs are intended to assist in ranking social states and thus in determining public policies. However, utilitarian SWFs are a rather abstract concept with few direct practical applications.

Table 7.1 Valuations of marginal utility of individuals with differing social weights

Values of social weights (ε)	*Individual utility levels relative to average utility*				
	2/1	*1/1*	*½*	*1/3*	*1/4*
0	1	1	1	1	1
0.5	0.71	1	1.41	1.73	2
1	0.5	1	2	3	4
2	0.25	1	4	9	16
4	0.0625	1	16	81	256

[9] When $\varepsilon = 1$, Equation 7.8 is indeterminate and is replaced by $W = \sum_i \log(u^i)$. It may also be noted that this SWF is ordinal and that the value of W may be negative, but it rises with increases in any individual utility.

One utilitarian SWF with a possible practical implication is the Rawlsian SWF. This implies that income should be redistributed to the least well-off person or group of persons in society. Of course, individual welfare must be defined, for example in terms of income, health, dependants and so on.

The simple additive SWF was also regarded traditionally as implying a strong case for income redistribution. Indeed, under the following assumptions social welfare is maximised by complete equality of income:

1. Individuals have similar utility functions (they gain equal utility from any given amount of income).
2. Marginal utility of income falls as income increases.
3. The total amount of income available is fixed and independent of its distribution.

If these assumptions hold, any transfer of income from a richer person to a poorer one increases the welfare of society, because a marginal dollar gives the poorer person more utility. This was shown in Figure 7.4. Moreover, the larger the transfer the greater is the increase in social welfare.

A related policy question is: how should government raise a given level of tax revenue with least welfare cost? If the marginal value of a dollar falls with income, a dollar tax on a high-income person results in less loss of total utility than does a dollar tax on a low-income person. To achieve a tax revenue target with least loss of welfare, tax should be levied at a 100 per cent rate on the highest incomes downward until the target is reached. For the marginal dollars of tax paid, all those paying tax would have an equal sacrifice. But this marginal sacrifice is lower than the sacrifice that would occur with a tax on lower incomes. People with low incomes would pay no tax.

However, these conclusions depend on the assumptions. If the assumptions do not hold, neither do the conclusions. Assumption (1) is questionable because utility depends on many factors besides income. Assumption (2) is more plausible. However, assumption (3) is very doubtful because income redistribution almost always reduces the amount of income available for distribution. Therefore, the conclusions that total utility is maximised by complete equality of income are questionable.

These examples show that maximising a utilitarian SWF subject to constraints may produce significant policy guidelines, but not operational outcomes. We have also seen that it is possible to estimate utility functions which allow us to convert income into utility equivalents. However, for practical purposes, it is generally necessary to work directly with economic variables such as income or consumption along with leisure. For example, macroeconomic states may be ranked as a function of GDP and some measure of income equality. To assess particular policies or projects, money measures of benefits and costs are generally required. We review these approaches below and their implications for evaluating social welfare.

Utilitarian social welfare functions: other considerations

Other important SWF issues are the components of utility, the consequentialist nature of SWFs and the related treatment of equity.

The components of welfare. In the eloquent words of the American Declaration of Independence (1776), all men "are endowed … with certain inalienable rights … among these are life, liberty and the pursuit of happiness". Whether or not one agrees with the notion of natural rights, most people would consider that their utility depends not only on income and leisure but also on their rights to vote and elect governments, to freedom of speech and congregation and freedom from arbitrary arrest. These rights are hard to measure.

Outcomes and process. SWFs are often treated as consequentialist: welfare depends on outcome, not on process. In the consequentialist view of social welfare, government should distribute output according to need and so as to maximise the total welfare of society, subject to the constraint that redistribution may reduce the output available for distribution. In effect, all output belongs to the state. No individual has a right to their output. .

However, taking process into account, there is another equity principle—the principle of just reward. According to this principle, people should be compensated for both labour and savings. Compensation for labour can be represented formally by including leisure in the utility function as in Equation 7.5. Other things being equal, individuals with more leisure have more utility. Those who work longer hours may be compensated by higher money income.

But what is a fair return to labour? Nozick (1974) argues that individuals have a right to all legally acquired holdings, acquired either through earnings or through inheritance of wealth providing this was justly acquired. This is an extreme view. A more moderate view is that individuals are entitled to keep what they can earn in a competitive market, given equal positions of opportunity and ability. Individuals should not be able to appropriate surplus income, above their opportunity cost, when this is due to market power in imperfectly competitive markets. Nor should they be able to appropriate economic rent due to superior innate capacities. This rent should be shared with less gifted individuals.

Some writers, such as Holcombe (1998), argue that fair process is as important as fair shares of output. If the process is fair (e.g. there is equality of opportunity and markets are competitive) the economic outcome may be viewed as fair. If a football game is fair, the outcome is generally viewed as fair. However, given differences in individual abilities and in resources available to individuals, and the random nature of misfortune, society can scarcely be indifferent to variations in economic outcomes even in competitive markets.

Conclusions. Most utilitarian SWFs imply that resources should be shifted from individuals with high utility to those with less utility. But utility (or need) are not easily defined concepts. On the other hand, the just reward approach argues that individuals should be entitled to their earnings provided there is equality of opportunity and markets are competitive. Such issues arise at many points in the text. For example, in our analysis of public finance below, we will discuss ability to pay versus user benefit principles of equity and the meaning of vertical and horizontal equity. In our discussion of social assistance, we will meet a variety of concepts of need especially in relation to differences between households.

Working with Income

For most practical purposes, economists work with incomes or willingness-to-pay dollar amounts rather than with utilities. At the macroeconomic level, we may trade-off total income (GDP) against equality of distribution. Thus a SWF may be represented as:

$$W = W(S, \theta) \tag{7.9}$$

where S is represents total income, which captures the efficiency aspect, and θ represents the inequality of income, which captures the equity aspect. This SWF is increasing in S and decreasing with respect to θ. We have seen previously that a utilitarian SWF is generally increasing with individual utility and that individual utility increases with income. To satisfy this feature the above SWF has to satisfy the following principle:

$$\frac{\partial W}{\partial S}\frac{\partial S}{\partial y_i} > \left|\frac{\partial W}{\partial \theta}\frac{\partial \theta}{\partial y_i}\right| \tag{7.10}$$

An increase in someone's income affects the SWF in two ways: it increases the total income of the society and changes the inequality. The above condition states that an increase in someone's income has a greater effect on total welfare than does an increase/decrease in inequality. Suppose that a policy change increases the income of one person (or group) in the society. This increases S, but this may increase or decrease inequality depending on whose income has increased. If inequality increases its effect on SWF is negative. However, with the above SWF, whoever may be the beneficiary of a policy change, social welfare will increase. The effect of an increase in income on S always exceeds the effect of this increase in income on θ. From an axiomatic perspective, the above SWF can be derived as:

$$W = \text{GDP per capita} \times (1 - \text{GC}) \tag{7.11}$$

where GC is the Gini coefficient measure of income inequality.

However, the core components of social welfare are *individual* utilities and where possible economists estimate the benefits and costs of economic changes to individuals. As economists have stressed for a long time and as Stiglitz *et al* (2009) discussed in a major OECD report, GDP is not a good measure of welfare even when adjusted for distributional effects because it does not include any value for non-marketed goods, health, externalities, consumer surpluses or leisure. A full accounting of welfare effects for individuals includes values for all these elements of welfare drawing on the valuation principles discussed in Chapter 6.

Weighting incomes. Turning to individuals, the core problem is the relationship, or the lack of it, between income and utility. Even if individual utility can estimated as a function of income, social valuations of marginal utility may differ from individual valuations.

Suppose that the SWF embodies the idea that the social value of an additional unit of income is some function of an individual's income. In essence, we can replicate the welfare function shown in Equation 7.8 but substitute income (y) for utility and some form of social weighting factor (w) for equity. Thus, suppose that social welfare can be re-expressed as a function of incomes:

$$W(u_1, u_2 \ldots, u_n) = \sum_i (w_i y_i) = \sum_i y_i^{(1-w)} / (1-w) \tag{7.12}$$

where w represents the elasticity of the social valuation of the marginal utility of income for each individual (or the degree of aversion to income inequality).

Table 7.2 provides summary values of the *social* value for an additional dollar for three individuals depending on the choice of social weight. This is of course simply *a reduced version of Table 7.1* with income differences instead of utility differences. If $w = 0$, the social value of extra income is the same for everyone. Society would be indifferent to distributional changes in income. If $w = 1$, the social value of marginal income is inversely proportional to income. If Ben has twice the income of Amy, the social value of an extra dollar to Amy is twice the social value of an extra dollar to Ben. The higher the value of w, the higher is the social value of marginal income of low-income individuals. In the extreme case, if $w = \infty$,

Table 7.2 Values of marginal income as function of individual income and social weights

Social weight (w)	*Individual income relative to mean income*		
	2/1	*1/1*	*0.5/1*
0.0	1.00	1.0	1.00
0.5	0.71	1.0	1.41
1.0	0.50	1.0	2.00
2.0	0.25	1.0	4.00

all weight would be attached to the income of the least advantaged person. The weighting approach is attractive in that it recognises welfare differences between individuals and attempts to provide an operational procedure for dealing with these differences. Policy makers must make such judgements in many situations.

However, determination of any such weights requires value judgements. While such valuations might be elicited from social surveys and econometric analysis, as in Layard et al. (2008), weighting dollars according to the estimated average marginal utility of income (w = 1.26) would itself represent a value judgement. Therefore, most economists are cautious about employing such weights and treat all dollar values equally in an evaluation procedure like cost-benefit analysis. Of course, this approach does imply a social weight (i.e. $w = 0$). However, this is essentially an attempt to separate efficiency and equity considerations. It does *not* imply that equity considerations are irrelevant to public policy. Rather, as discussed below, they need to be considered alongside expected efficiency outcomes.

Social Welfare and Cost-Benefit Analysis

In practice, decision makers deal mainly with specific policies and projects rather than with overall design of the economy. Typically, we want to know whether a policy or project will increase social welfare. To answer this question, we generally focus on the change in social welfare (ΔW). Adopting utility units, we want to know whether:

$$\Delta W = \sum_i w_i \Delta u_i > 0 \tag{7.13}$$

where w_i are social weights where applicable.

However, given the difficulty of estimating u_i and w_i, as a practical procedure, economists generally estimate whether the benefits exceed the costs by drawing on unweighted income-based valuation measures described in Chapter 6. Therefore, we estimate whether:

$$\Delta W = \sum_i (b_i - c_i) = \Sigma y_i > 0 \tag{7.14}$$

where b_i and c_i are estimated benefits and costs for individuals (or other economic agencies) and y_i are income-based measures of changes in individual welfare, which may be positive or negative. If the estimated total benefit exceeds total cost, there is said to be a net social benefit (NSB > 0). Not surprisingly, this approach is called cost–benefit analysis (CBA).

The monetary valuations of goods usually reflect changes in consumer and producer surpluses based on ordinary demand schedules, but they may reflect estimates of compensating or equivalent variations when appropriate. As we saw in Chapter 6, the differences between these valuation measures are generally small. When they are not, the choice of measure can have significant implications. However, the aggregation of individual values into an overall measure of social welfare is generally more contentious than the valuations of goods to individuals or other economic agents. This takes us back to the issue of how to treat equity if not by weighting valuations.

Aggregating income-based measures of welfare changes

Evidently when a project or policy produces winners and losers and when a dollar has a different value to different individuals, NSB = $\Sigma y_i > 0$ may not represent a net social welfare improvement if the distributional effects are regressive.

However, $\Sigma y_I > 0$ does represent a *potential* net social welfare improvement (it is a potential Pareto improvement). If estimated benefits exceed costs, the gainers from a change can fully compensate the losers. If compensation is made, someone is better off from the change and no one is worse off. The resource reallocation would then be an actual Pareto improvement. This approach is formalised in three related compensation criteria:

1. **Kaldor criterion:** a change from *A* to *B* is desirable if those who gain from *B* could compensate the losers and still be better off at *B* than at *A*.
2. **Hicks criterion:** a change from *A* to *B* is desirable if the losers from *B* cannot bribe the gainers into not wanting the change to *B*.
3. **Scitovsky criterion** (sometimes described as the double criterion): a change from *A* to *B* is desirable if those who gain from *B* could bribe the losers to accept the change and the losers could not bribe the gainers into not making the change.[10]

There is relationship between these compensation criteria and the concepts of compensating and equivalent variation (CV and EV) described in Chapter 6. The Kaldor test requires that the sum of CVs is positive. The Hicks test requires that the sum of EVs is positive.

Scitovsky (1941) pointed out that the Kaldor and Hicks criteria could create inconsistent outcomes and that strictly both criteria would need to be met. The Scitovsky reversal paradox is illustrated in Figure 7.5. Suppose that the initial utility point is *A*. Government then decides to build a new road that provides benefits only to *Y*, but for which both *X* and *Y* pay. The new utility levels are given by point *B*. However, *Y* can compensate *X* and both can be better off at point *C*. But if compensation is *not* paid, the starting point on the utility map with the road is point *B* and the comparator (without the new road) is point *A*. Given that *X* prefers non-road goods, she is better off at *A* with no road. She is also willing to bribe *Y* into having other goods and not having the road at point *D*, which both prefer to point *B*. Therefore, when compensation is not made and the distribution of income changes, the Kaldor and Hicks criteria can justify both a change and the reverse decision. Fortunately, this inconsistency is rare and for most purposes the Kaldor criterion is adopted.

Kaldor and Hicks designed their criteria in response to the Robbins (1938) critique that individual utilities cannot be compared and to avoid the need for a social welfare function. If losers are compensated, changes that satisfy their criteria are Pareto efficient. There would be no need for interpersonal comparisons of utility or indeed for any SWF. Moreover, such an approach may be justified pragmatically. Because governments undertake many projects, adopting projects with a positive NSB (or where there are alternatives, options with highest NSB) will raise the welfare of most people more than would projects with lower NSBs,

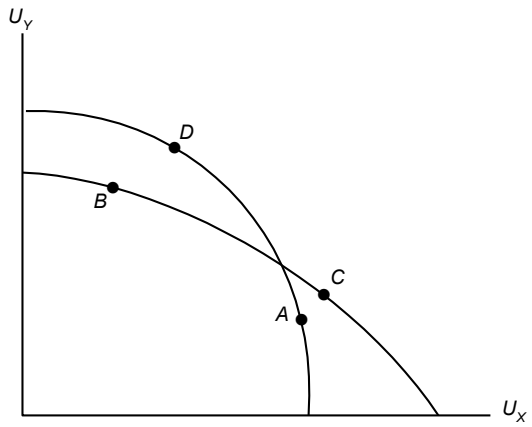


Figure 7.5 The Scitovsky reversal paradox

[10] See Kaldor (1939), Hicks (1939) and Scitovsky (1941).

including projects with negative net benefits. However, the compensation principle is based on the notion of hypothetical compensation. It does not require that compensation is paid.

If compensation is not paid and someone loses from a policy change, the argument that compensation could be paid is of questionable relevance. The NSB criterion (Equation 7.14) gives equal weight to all dollar valuations of benefits and costs regardless of their distribution. Of course, some projects are redistributive from rich to poor as well as efficient. However, if high income individuals gain from a project and low-income people lose, a project may satisfy the net social benefit criterion but worsen the distribution of income.

Evaluation options

There are three main ways to evaluate policies or projects. All are based on estimates of individual valuations of benefits and costs. The differences relate to the treatment of aggregate net benefits.

First, we may evaluate policies with a social welfare function that applies social (equity) weights to WTP values for different groups of individuals. This approach aims to fully integrate efficiency and equity outcomes. Equity-weighted measures of NSB are used occasionally, for example in evaluations of economy-wide tax and income transfer systems (see optimal taxation literature in Chapter 28). However, they are used here mainly for academic demonstration purposes because there is no objective basis for establishing equity weights for groups of individuals. This approach is rarely used to evaluate specific policies or projects because establishing equity weights for individuals for individual cases is even more problematic. Moreover, an estimated equity-weighted NSB is an unclear mixture of efficiency and equity impacts. It does not have a clear welfare meaning. A policy could show a positive NSB even if the distributional impacts were regressive.

Second, policies may be evaluated simply with an efficiency criterion such as unweighted NSB. Government would adopt policies that pass the NSB test, because this makes most efficient use of resources. Government would deal with equity separately as needed through income transfers or other compensation programs. This could include indirect compensation whereby government would compensate losers via fiscal policies to ensure that overall changes do not disadvantage less well-off individuals. This approach may achieve high efficiency outcomes and reasonable equity at low cost because fiscal instruments are generally better instruments for redistribution than are projects. However, acceptance of it requires faith in the political process. When a policy disadvantages some individuals, they may not take the sanguine view that they will receive appropriate compensation via other policy changes.

The third approach to evaluation accepts that equity may be important at the individual policy or project level. However, it adopts a dual track consideration of efficiency and equity rather than attempt to integrate efficiency and equity into a single overall estimate of social welfare (as in the first approach). The economist would present decision makers with both (1) the estimated aggregate (unweighted) NSB of a policy and (2) a description of the distributional impacts, possibly including potential compensation packages. This is a common approach.

The estimated NSB, based on cost–benefit analysis, indicates the potential efficiency gains from a policy. But the economist generally does not attempt to convert this into an aggregate measure of social welfare. Rather the economist provides separate information on the efficiency and equity consequences of any policy. Government then makes the policy determination.

Conclusions

Welfare economics provides a rigorous framework for determining the allocation of resources and related public policy decisions. Valuations of individual benefit and cost are based on individual preferences. The principles of compensating and equivalent variation, or of consumer or producer surpluses, provide exact monetary measures of benefit and cost. Thus, welfare economics provides a means to quantify costs and benefits.

However, technical analysis cannot resolve all issues. A dollar may be more valuable to one person than to another. In any case, social valuations of individual needs may differ from private valuations of the marginal utility of income. Thus, it is often necessary to make trade-offs when someone gains from a policy and another person loses. Unless full compensation is possible, policy making requires a trade-off between the efficiency gain of a policy change and the distribution of the gains and losses.

It is sometimes feasible to treat efficiency and equity separately. At the macroeconomic level, policy makers would aim first to maximise the total value of output by the most efficient use of resources and then redistribute the income earned in an equitable way. At the microeconomic level, government would adopt efficient policies that provide the highest net social benefit (those policies with greatest excess of total benefit over cost) and deal with distributional issues separately, where appropriate by compensating losers. Of course, a social welfare function, reflecting values judgements, is still required (at least implicitly) to determine the optimal redistribution.

A rationale for treating efficiency and equity as separate issues is that efficiency of resource use is a technical issue and distribution of output a normative matter. Also, when policies have minor equity implications, the separation of efficiency and equity may be practical and reasonable. However, in the absence of individualised lump sum transfers, it is impossible completely to separate equity from efficiency—one will affect the other. It is then impossible to judge between alternative economic states without an ethical social welfare function or a political judgement.

Summary

- The fundamental problem of public choice is to determine how to assess, compare and aggregate gains and losses from the use of resources.
- The assessment is based on valuations of individual preferences. Comparison and aggregation of the valuations raise greater problems as the marginal utility of income varies across individuals.
- Also social valuations of marginal income may vary according to the level of wellbeing of individuals in society.
- Ideally policies or projects would maximise social welfare, using some agreed measure of social welfare, that is, a social welfare function (SWF).
- The SWF describes the relationship between social and individual welfare. However, individual utilities can be combined into an overall measure of social welfare in many ways and this aggregation requires value judgements.
- Most public policy involves possible changes from the present rather than design of a whole social state. In this case we want to measure changes in the welfare of individuals (i.e. their benefits and costs) and to find a way to sum these changes.
- If the estimated total benefits of a policy exceed the costs, the policy is described as efficient in the sense that the gainers could compensate the losers. If compensation is made directly or indirectly in the longer run, someone would gain, and no one would lose, from the policy.
- If compensation is not made and a project or policy has some expected adverse effects, policy makers must decide whether the efficiency benefits justify the distributional impacts.
- There is no ready technical basis for weighting the benefits and costs into an overall measure of social welfare.

Questions

1. What is a social welfare function?
2. Is it possible to compare individual welfare levels? If such comparison is not possible, is it possible to make judgements about social welfare?
3. Does a Pareto-efficient outcome avoid the need for a social welfare function?
4. Can a utilitarian social welfare function be equitable?
5. Assume an economy with only two people (Amy and Ben).
   i. Let the social welfare function be $W_1 = u_a + u_b$ where u_a and u_b are the utilities of Amy and Ben. Graph the iso-welfare curves (the social indifference curves). What importance, if any, is attached to their relative wellbeing?
   ii. Repeat for $W_2 = 2u_a + u_b$. What are the implications?
   iii. Draw a utility possibilities curve and show how the SWFs in i and ii affect the desired outcome.
   iv. What are the main weaknesses of the additive utilitarian social welfare function?
6. Should government be concerned with the distribution of utility or the distribution of income?
7. Is it possible to estimate the marginal utility of income?
8. Explain why social valuations of marginal needs may differ from individual valuations of the marginal utilities of income. Is it possible to measure these social valuations?
9. When does cost–benefit analysis maximise a utilitarian social welfare function?
10. Is maximising an additive utilitarian social welfare function always inequitable?
11. Define the compensation principle. What is the relationship between this principle and a potential Pareto improvement? What value judgements underlie the compensation principle?
12. When may the Kaldor and Hicks compensation tests give different policy answers?
13. Amy's utility is u_A and her income is y_A while Ben's utility is u_B and income is y_B. Suppose: $u_A = 10y_A^{0.5}$ and $u_B = 10y_B + 0.8u_B$. Suppose initially that both of them have \$100. If you consider a (simple) additive SWF, what will happen to social welfare if \$36 is taken away from Ben and given to Amy?
14. How can cost–benefit analysis deal with distributional impacts of policies?

---

15. Society contains three individuals (groups) with the following income, health and perceived utility.

	1	*2*	*3*
Income (\$ per annum)	20 000	40 000	80 000
Health	Good	Poor	Good
Utility	100	120	200

The government has six policy options (A to F) which give forecast income changes as shown below.

Option	*1*	*2*	*3*	*Total*	*Rank*
A	+1000	0	-3000	-2000	
B	+500	+500	-2500	-1500	
C	+1500	-500	-500	+500	
D	+500	0	+500	+1000	
E	-100	+1100	0	+1000	
F	+1000	-250	+500	+1250	

i. How would you rank the options?
ii. What does this imply about a social welfare function?
iii. What does it imply about compensation schemes?

Further Reading

Buchanan, J.M. (1959) 'Positive economics, welfare economics, and political economy', *Journal of Law and Economics*, 2, 124–138.

Kaldor, N. (1939) 'Welfare propositions of economics and interpersonal comparisons of utility', *Economic Journal*, 49, 549–552.

Layard, R. (2006) 'Happiness and public policy: a challenge to the profession', *Economic Journal*, 116, C24–33.

Layard, R., Nickell, S. and Mayraz, G. (2008) 'The marginal utility of income', *Journal of Public Economics, Special Issue: Happiness and Public Economics* (eds) T. Besley and E. Saez, 92, Nos 8–9.

Little, I.M.D. (1957) *A Critique of Welfare Economics*, Chapters I and XV, Oxford University Press, Oxford.

Mueller, D.C. (2003*) Public Choice III*, Chapter 23, Cambridge University Press, Cambridge.

Ng, Y-K. (2003), *Welfare Economics*, Macmillan, Basingstoke, UK.

Pearce, D.W. (1983) *Cost–Benefit Analysis*, Chapters 1–3, Macmillan, London.

Stiglitz, J.E., Sen, A. and Fitoussi, J-E. (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, OECD, Paris.

Tresch, R.W. (2002) *Public Finance: A Normative Theory*, Chapter 4, Academic Press, New York.

Chapter

8

Cost-Benefit Analysis

The age of Chivalry is gone: that of sophisters, economists and calculators has succeeded.

Edmund Burke

Cost-Benefit Applications ♦ Overview of Cost-Benefit Analysis ♦ Measures of Net Social Benefit ♦ Valuing Benefits and Costs ♦ The Social Discount Rate ♦ Uncertainty and Risk ♦ Distributional Issues ♦ CBA and Other Evaluation Methods

Cost–benefit analysis (CBA) aims to estimate the net social benefit (NSB) of a project or policy. The NSB equals total benefit less total cost. Benefits and costs include market and non-market goods. NSB may be positive or negative. CBA is applied to a wide range of public projects and policies, based on the principles of welfare economics. It is more comprehensive than other forms of economic evaluation, such as cost-effectiveness analysis and financial analysis. CBA is widely used across international, national and state jurisdictions.

This chapter describes the method of cost–benefit analysis. The first part describes applications and outlines the method. The chapter then describes the main features of CBA: measures of net social benefit, the valuation of costs and benefits and the treatment of time, risk and distributional issues. The last part of the chapter briefly reviews other less comprehensive evaluation methods.

Cost-Benefit Applications

CBA was developed initially to assess public capital expenditure projects, notably water and transport projects. Would the benefits justify the expenditure compared with investing in other projects?

Figure 8.1 overleaf shows various kinds of benefits that may occur with: a new good, improved service, increased capacity or lower costs. In Figure 8.1a, the benefit of a new good net of recurrent costs in a given period is represented by the shaded area between the demand curve and the marginal cost curve. The net benefit may change over time. The estimated net benefits over the life of the project are compared with the capital expenditure, which is not shown in this figure. The shaded area in Figure 8.1b shows the net benefit of an improvement in quality for some service, as demand shifts from D_1 to D_2. Figure 8.1c shows the net benefit of an increase in the capacity of a service, such as water supply, from S_1 to S_2. Figure 8.1d shows the benefit of a reduction in operating costs from MC_1 to MC_2.

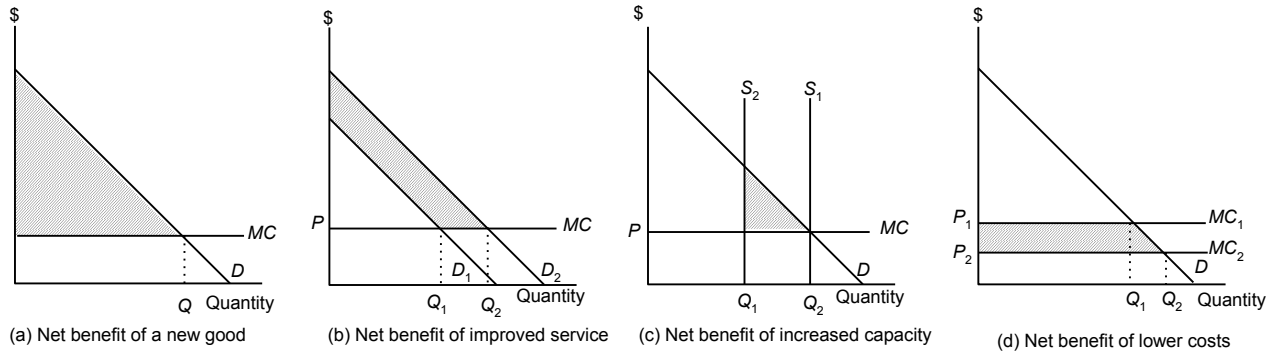


Figure 8.1 Examples of benefits from capital expenditure (shaded areas)

Figure 8.1 shows the direct benefits of consumers and marginal costs of producers of the service. CBA also accounts for indirect effects where appropriate. These may be of two kinds. First, there may be impacts in related markets for complementary or substitute goods. For example, a major event may generate international visitors (the direct effects) and complementary benefits may arise for businesses servicing tourists (indirect effects). Where prices diverge from marginal costs in related markets, there may changes in producer surplus in these markets. Second, there may be indirect non-market effects, including environmental impacts, where third parties are affected (positively or adversely). Table 8.1 shows examples of direct and indirect effects in road and education projects.

Importantly, in addition to capital projects, CBA is used to evaluate recurrent programs and all kinds of policies. For example, it may be used to assess optimal class size in schools. And most regulations impose costs and confer benefits on various parties that can be assessed using CBA. For example, environmental regulations typically impose costs on firms with the aim of improved health and environmental amenity. Likewise, regulations dealing with occupational health and safety and consumer protection, for example for medicines and food, generally impose costs on firms and aim to improve individual welfare.

Table 8.1 Examples of costs and benefits in CBAs for roads and education projects

	Roads		***Education***	
	Costs	***Benefits***	***Costs***	***Benefits***
Direct effects	Construction costs Maintenance costs	Savings in travel time, vehicle costs and accidents by firms and households using new road	Public costs of providing education Student income forgone Student out-of-pocket expenses	Benefits to students (higher incomes and improved quality of life) Increase in tax returns
Indirect effects	Losses to rail operators Noise, air pollution and lower amenity around new road	Savings in travel time, vehicle costs and accidents on other roads Lower noise and improved amenity around other roads	Displacement of existing workers	Benefits to employers (higher productivity) Reductions in crime and other social costs

Australian legislation requires that all new regulations are subject to a Regulatory Impact Statement that demonstrates the net public benefit of the regulation, which in effect is a requirement to provide a CBA.

Applications of CBA can be found in many places in this book, including the evaluation of education and health care services, transport, environmental regulations and public safety policies, and in public interest tests of restrictions on competition.

Overview of Cost-Benefit Analysis

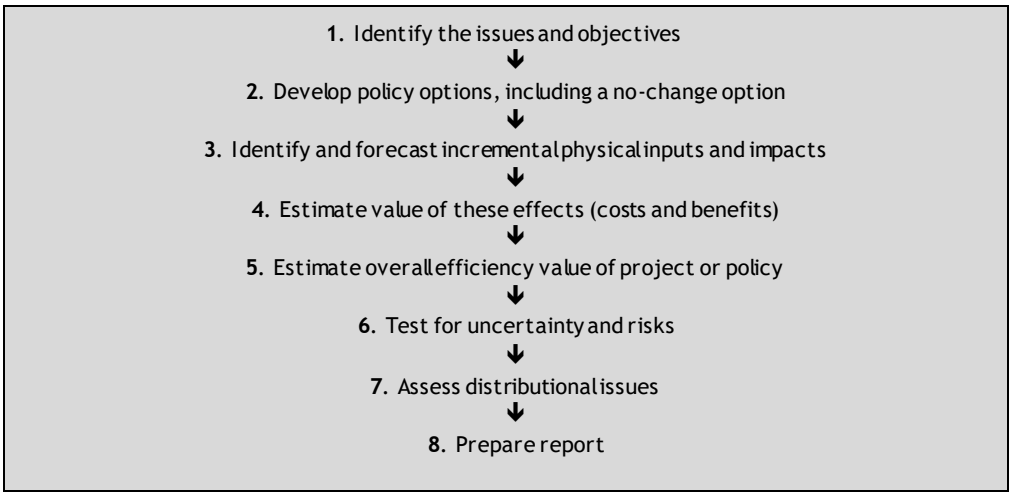
Box 8.1 provides an overview of the CBA process. The first three steps in the process are common to any rational evaluation procedure. However, some points about them should be noted as they are critical to the outcome of a CBA.

Development of appropriate options (step 2) is fundamental. This should include a 'no change' option, which is often called the Base Case. However, the easiest way to 'prove' that a project or policy is desirable is to evaluate it only against a poor Base Case and to ignore other options. The options should generally include small changes as well as large ones.

Also critical is the forecast of incremental physical impacts (step 3). These may be forecasts of traffic, power consumption, environmental impacts, health effects and so on. These effects should be estimated as incremental to the Base Case. Many forecasts are provided by other disciplines such as traffic modellers, statisticians or epidemiologists, but in some cases economic factors such as income and prices can critically affect forecast outcomes.

While it is generally desirable that economists contribute to steps 1 to 3, the key economic contributions are steps 4 to 7: valuing effects on individuals, combining them into an overall measure of project value (net social benefit) and dealing with risk and distributional issues. These contributions are discussed below. To provide a context we first describe overall measures of project worth.

Box 8.1 Key steps in cost-benefit analysis



Measures of Net Social Benefit

There are three main cost–benefit measures of net social benefit (NSB): the net present value, internal rate of return and benefit–cost ratio. Each measure includes all quantified costs and benefits and discounting for time differences. Each measure indicates whether the NSB is positive. If NSB is positive, a project is described as efficient—the gainers could compensate

the losers so that at least one person would be better off, without making anyone else worse off. When a project passes the efficiency test by one measure, it also passes on the other two measures. However, the three tests may rank efficient projects differently.

Net present value
The present value of benefits net of costs

Net present value (NPV) is the most common efficiency measure of a project. The estimated NPV is the present value of estimated benefits net of costs:

$$\text{NPV} = \sum_{t=1}^{n} \frac{b_t - (K+c)_t}{(1+r)^t} \tag{8.1}$$

where b stands for benefits, K and c are capital and operating costs respectively, there are $t = 1...n$ years, and r is the selected discount rate (usually the rate of return on a marginal alternative project—see discussion below). If NPV > 0, the estimated total benefit exceeds total cost. Because benefits are the value of consumption gained and costs are the value of consumption forgone, a positive NPV indicates an increase in the total value of consumption. If there are several options, the one with the highest NPV would be the most efficient.

Suppose that, in constant prices, a project has a capital cost of $150 million, generates net benefits of $40 million annually for five years and the discount rate is 7 per cent per annum.

$$\text{NPV}=-\frac{150}{1.07}+\frac{40}{(1.07)^2}+\frac{40}{(1.07)^3}+\frac{40}{(1.07)^4}+\frac{40}{(1.07)^5}+\frac{40}{(1.07)^6}=\$13.1 \text{ million} \tag{8.2}$$

This project would be efficient. Indeed, if the discount rate were 10 per cent, the NPV would be $1.5 million and still be marginally efficient.[1]

Internal rate of return
The rate of return that would give a NPV equal to zero

The internal rate of return (IRR) on capital employed is the rate of return that would give NPV equal to zero. Equivalently, the IRR equates discounted net benefits to discounted capital costs. It is obtained by solving for i in (8.3):

$$\sum_{t=1}^{n} \frac{K_t}{(1+i)^t} = \sum_{t=1}^{n} \frac{(b-c)_t}{(1+i)^t} \tag{8.3}$$

where the symbols are as above except for i, which is the IRR. If the estimated project IRR exceeds the chosen discount rate (r), NPV > 0. The project is efficient. Using the same figures as in Equation 8.2, the IRR would be 10.4 per cent. The project would be efficient because the IRR exceeds the selected discount rate of 7 per cent.

$$\frac{150}{1.104}=\frac{40}{(1.104)^2}+\frac{40}{(1.104)^3}+\frac{40}{(1.104)^4}+\frac{40}{(1.104)^5}+\frac{40}{(1.104)^6} \tag{8.4}$$

However, the NPV and IRR measures may rank projects differently. Table 8.2 shows three projects, *A*, *B* and C, along with their NPV and IRR returns. Project *B* has a higher NPV (using a 7 per cent discount rate) than both *A* and *C*, but a lower IRR. The IRR measure favours projects with high returns in early years and small projects. However, in both cases the high IRR is misleading. Take project *A*—the early surpluses cannot be reinvested at the internally determined discount rate. Thus, if the $120 million surplus in year 2 were reinvested

[1] The year to which costs and benefits are discounted affects the size of NPV but not the sign. The text discounts first year costs and benefits, as the default formula for NPV in Excel does. If the analyst prefers to discount from year 2, the Excel formula can be adjusted to allow for this.

Table 8.2 Project outcomes with the IRR and NPV criteria ($m)

Project	*Year 1 Capital*	*Year 2 Net benefit*	*Year 3 Net benefit*	*IRR (%)*	*NPV 7% discount rate*
A	-100	120	0	20.0	12.1
B	-100	0	135	16.2	17.9
C	-50	0	70	18.3	10.4

at the available 7 per cent rate, the return in year 3 would be $128.4 million ($120 m × 1.07). This is less than the $135 million surplus achieved by *B* in year.

In the case of project *C*, the savings in capital expenditure cannot be invested at the same rate as that achieved by the small project. If the opportunity cost of capital is estimated correctly, the additional $50 million in capital expenditure would generate precisely $50 million in benefits (i.e. a net benefit of $0).

In general, the NPV measure provides the preferred ranking because project benefits should be discounted by the chosen rate of discount that reflects the real opportunity cost of capital rather than by an arbitrarily determined mathematical rate.

Nevertheless, the IRR can be a useful measure of project value because some people understand it more easily than the NPV. Also, it provides a check against large projects. An NPV of $12.1 million is significant for a $100 million project but not for a $1000 million project. The estimated IRR in the latter case would be only marginally above the test discount rate and a warning that this might not be a viable project.

The benefit–cost ratio (BCR) is generally defined as:

$$\text{BCR} = \sum_{t=1}^{n} \frac{(b-c)_t}{(1+r)^t} \Big/ \sum_{t=1}^{n} \frac{K_t}{(1+r)^t} \qquad (8.5)$$

Benefit-cost ratio
The ratio of the present value of net recurrent benefits to the present value of capital expenditure

where the symbols are as above. Note that recurrent costs are placed here in the numerator.[2] If the BCR > 1, NPV > 0, and vice versa. Using the figures in Equation 8.2, and a 7 per cent discount rate, the BCR would be 1.09. This indicates that the project is just efficient.

Box 8.2 NPV versus BCR criterion

Suppose that three projects (*A*, *B* and *C*) have the following capital costs and benefits. If there were no capital constraint, *A* would be preferred. Compared with say *B*, *A* has an incremental cost of $20 million and generates incremental benefits of $35 million. On the other hand, if the agency has a capital constraint of, say, $50 million, it would maximise the surplus obtained from the $50 million by selecting *B* and *C* instead of *A*.

Project	*Capital cost ($m)*	*Discounted benefits ($m)*	*NPV ($m)*	*BCR*
A	50	105	55	2.10
B	30	70	40	2.33
C	20	50	30	2.50

[2] If the purpose is to estimate the return to scarce capital expenditure, which it usually is (as in Box 8.2), only capital expenditure should be included in the denominator. However, some texts and many practitioners include recurrent costs in the denominator.

Again, the NPV and BCR measures may rank projects differently, with the BCR favouring projects with low capital expenditure. The NPV measure is generally preferred because it has no size bias. The NPV measure ensures that any additional capital required for a large project is discounted at the appropriate marginal opportunity cost of capital. If employment of additional capital increases the estimated NPV, this is the most efficient use of the capital.

However, the BCR is relevant to decision making if the capital available to an agency is constrained so that the marginal return on the agency's use of capital exceeds the marginal return obtainable elsewhere. In this case, the agency should select projects in order of their present value per unit of constrained capital (i.e. by the BCR) until the capital constraint is exhausted. This maximises the surplus (the NPV) from the use of scarce capital. Box 8.2 provides an example.

Valuing Benefits and Costs

Like any measuring system, CBA must be based on a standard unit of measure (a numeraire). The most convenient numeraire is the local currency unit in present-day domestic prices. The choice of unit does not affect the result of a CBA. Costs and benefits could also be measured in an international currency unit. What matters is the valuation principles and consistency of measurement.

We could also use forecast prices (known as current or nominal prices) instead of present-day (constant) prices. However, constant prices are simpler to use and facilitate inter-temporal comparisons. However, three implications of constant prices should be noted.

1. Use of constant prices presumes that all prices change at a similar rate, so that price relativities are constant. If the price of some good, for example electricity, is reliably forecast to change relative to the general price level, this change in real value should be allowed for.
2. Future effects should be discounted by a real rate of discount, which excludes inflation. As can be seen from Equation 8.6, the net present value of a stream of future net benefits is the same whether it is estimated with a constant price and discounted by a real rate of discount or estimated in inflated current prices and discounted by an inflated nominal discount rate:

$$\text{NPV} = \sum_{t=1}^{n} \frac{(b-K-c)_t (1+\pi)^t}{(1+r)^t (1+\pi)^t} \tag{8.6}$$

   where r is the real rate of discount and π is the rate of inflation.
3. However, a financial analysis (as distinct from a CBA) may need to allow for expected inflation. Policy makers may require forecasts of nominal cash flows, including interest payments.

The value of foreign currency can also be an issue if the exchange rate does not reflect the real value of foreign currency. This may occur because government controls the exchange rate itself or because it controls currency movements or trade. Most often, such controls result in an overvalued local currency and an undervaluation of foreign currency. In a CBA it may then be necessary to allow a premium for the real value of foreign currency. The standard approach is to adopt a price for foreign exchange based on what purchasers of foreign currency are willing to pay for it rather than the official price. This raises the effective price of imports so as to measure their social value in terms of the prices of domestic goods.

However, in the absence of government controls on the exchange rate or on international trade and capital, the market-determined exchange rate reflects the real value of foreign

currency and can be used to convert foreign currency costs or earnings to the local currency value. This is the case in most OECD countries, including Australia.

Valuing goods

The general principle of valuation in CBA is that *goods (whether market or non-market goods) should be valued as the individuals concerned would themselves value them*. More precisely, adopting the compensating variation principle discussed in Chapter 6, goods would be valued at the maximum price that individuals or firms would be willing to pay (WTP) for them and be no worse off than at present.

In practice, as we saw in Chapter 6, valuations of market goods may be derived from observed or compensated demand curves (the latter hold utility constant and can provide compensating or equivalent variations). However, when income effects are small, the valuation differences arising from the different measures are small. Given the practical advantages of using observed demand curves, valuations in CBA studies are generally based on them. The points along an observed demand curve are interpreted as marginal WTP values or marginal benefits. The area under the demand curve is the total benefit.[3]

The WTP value of a good equals the sum of the price paid plus any consumer surplus. The lower the price, the larger consumer surplus likely to be. If goods are provided free, WTP is entirely consumer surplus. Estimates of consumer surplus are therefore required in many CBA studies.

When goods are marketed, standard econometric methods can be used to estimate demand curves. A critical parameter is the price elasticity of demand which shows how demand responds to changes in price. The price elasticity of demand (η_d) is the ratio of the percentage change in quantity demanded (Q_d) to the percentage change in price (P):

$$\eta_d = \frac{\Delta Q_d / Q_d}{\Delta P / P} = \frac{\Delta Q_d}{Q_d} \times \frac{P}{\Delta P} \tag{8.7}$$

where Δ represents change. A price elasticity of –1.0 means that the percentage change in quantity demanded is inversely proportional to the percentage change in price. The price elasticity varies along linear curves and is constant for log linear curves (curves that are linear in logarithms). Price elasticities are typically estimated from time series or cross-sectional studies that relate quantities purchased to prices paid and other determinants of demand, such as income, household characteristics and the prices of other goods. In practice, most cost–benefit studies draw on established elasticities rather than estimate separate elasticities.

When goods are not marketed, other valuation methods are required. WTP amounts for non-market goods can be estimated either by other revealed preference methods or by stated preference methods. Revealed preference methods infer individual values from the behaviour of individuals in various contexts. Stated preference methods obtain values by asking people what they are willing to pay for goods. These valuation methods are described in Chapter 11, where we discuss various ways to estimate the value of non-market goods. Box 8.3 overleaf introduces the concepts by showing how they are used to estimate the value of travel time savings.

Valuing use of resources: factors of production

Resource costs are a product of the quantity of resources used and their real unit cost. Usually experts other than economists provide estimates of the quantities of resources required for production. The economist then estimates the real unit cost of these resources. Errors in cost estimates arise more often because of poor estimates of resource quantities than of unit costs.

[3] As discussed in Chapters 4 and 6, this assumes generally rational consumer behaviour.

Box 8.3 The value of travel time savings

Analysts conventionally distinguish between two kinds of travel time savings: working time and leisure time savings. Working time savings are generally valued at the wage rate of the respective tripmaker, sometimes with an addition for directly related on-costs. In a competitive market, workers are paid the value of their marginal product. If a worker earns $50 an hour, each hour spent in travel represents the loss of $50 of output and savings in travel time are worth $50 per hour. This assumes that the time spent in travel has zero productivity. Some studies have shown that this is not true. Also, the tripmaker is assumed to be personally indifferent between time spent in travel and in the office.

Leisure time savings have been valued in several ways (Hensher, 2011). Traditionally, the most common way was by revealed preference analysis of mode or route trip choices. Analysts examined how much trip makers were willing to pay to save travel time by choosing a faster but more expensive mode or route, for example travel by train rather than by bus or by a time-saving toll road rather than by a slower toll-free route. Other studies have estimated the value of travel time from house prices that vary with accessibility to employment. More recently, many analysts have used stated preference survey techniques. This may involve simply asking people what they would be willing to pay to save a certain amount of travel time or eliciting travel time values from responses to more complicated choices involving travel time, money and other travel attributes. Many studies indicate that people are willing to pay on average between 30 per cent and 40 per cent of their hourly wage to save an hour of travel time for leisure purposes. The results vary partly with the disutility of certain kinds of travel.

It may be questioned why the values of working and leisure time are not equal since microeconomic theory tells us that, at the margin, individuals forgo an hour of leisure for an hourly wage. One reason is the tax wedge. A more critical reason is that leisure time is not actually saved. If non-work travel time is saved, leisure is spent in a preferred manner—at home or in the pub rather than on a congested road. Studies of the value of leisure travel time savings are therefore valuing lifestyle preferences rather than actual leisure time savings.

Here we focus on the cost of resources (also known as factors of production). These factors include land or more generally natural resources, labour and purchases. Purchases may include purchases of materials, utility services, property and plant and equipment. Of course, purchased goods may themselves embody land, labour and capital, and other purchases.

Opportunity cost
The value of output forgone by using a resource in one way rather than another

Generally the real cost of a resource is its **opportunity cost**—this is the value of output forgone in its most productive alternative use. The key valuation principles are:

- If a resource (an input to production) is in fixed supply, its opportunity cost is the highest price that another producer would be willing to pay for it. This price would generally be the market price inclusive of any indirect taxes.
- If an input can be increased to meet project demand, its opportunity cost is the marginal cost of supply. This excludes any indirect tax and any mark-up on the costs of factors of production in an uncompetitive market. This is generally less than market price.

So when are resources or inputs to projects in fixed supply? They tend to be in fixed supply when there is full employment, natural resources or land are in fixed supply, one or a few firms control the output of intermediate goods, or government regulates the supply of the good or service. In such cases the supply of the relevant input cannot be readily expanded to meet an increase in demand. Thus the market price (the amount that another user is willing to pay for the resource) is the effective opportunity cost. In most OECD countries, including Australia, market prices are usually good indicators of the real cost of employing factors of production and of purchased inputs and few adjustments to market prices are required.

However, market prices do not represent real costs when market failures, notably imperfect competition, or government regulations cause prices to diverge from marginal cost. Thus, monopoly prices may overstate the real marginal cost of supply. Or, if there is unemployment, wages paid may not reflect the opportunity cost of employment. Where there are tariffs on imported goods, the prices of imports may not reflect their real cost. In such cases, the real

opportunity cost of using a resource may vary from observed prices. In the cost–benefit literature, these real costs are often described as 'shadow prices'. This term indicates that the price does not actually exist in the market.

Let us now apply our valuation principles to the main factors of production, namely to natural resources including land, purchases of various kinds and labour.

The real cost of land and other natural resources. Given that these resources are in fixed supply, the opportunity cost (OC) of using them is their value in the highest alternative use. This applies whether the resources are privately or publicly owned. If the resource can be sold in a competitive market, its OC is the competitive market price. If land is publicly owned and cannot be sold for some reason, the cost of using it for one purpose rather than another is again its highest value in the forgone use. For example, the cost of urban parkland is typically the value of housing land than is forgone.

The real cost of purchases. Consider first purchases supplied by domestic producers. If supply is fixed, the material must be diverted from another user. The OC is the highest price that another user would be WTP for it, inclusive of any taxes. On the other hand, if the purchase can be supplied by increasing local output, the OC is the marginal cost of production exclusive of indirect taxes and subsidies and any monopoly mark-up on the price. In these cases any indirect tax or subsidy is a transfer payment and not a resource cost.

The OC principle also applies to imported goods. If there are no restrictions on imports, the cost of an imported good is simply the price paid less any tariff. The tariff is a transfer payment between importer and government and not a net cost to society. However, if there is a quota on an imported good, the supply of imports is fixed, and the cost of the imported good is its full market price.

The cost of labour. If a worker (Ben) is indifferent between occupations including leisure, the opportunity cost of employing him is the value of forgone output. If Ben would be otherwise employed in a competitive economy, the OC is his highest wage in alternative employment. If he would be otherwise unemployed, the real cost of employing him is zero. Nothing is lost by the employment. Suppose that an employer pays a wage of $120 a day, which also represents the value of Ben's marginal output in this employment. In this case the employer receives no marginal benefit. But because Ben is indifferent between work and leisure, he obtains a benefit of $120 (and society as a whole is $120 better off).

Valuation of labour is complicated when labour has occupational preferences, preferring leisure to work or working in one occupation rather than another. The real cost of employing Ben is then his 'reservation' wage—this is the minimum amount that will attract the worker into a specific employment. Suppose that Ben would require $100 a day to forgo surfing. The government offers $120 for a day's work, which he accepts. In this case the real cost of Ben's employment is $100 because this is the minimum amount that he would accept for this kind of work and be no worse off than before. Any lower figure ignores the value of his leisure.

Valuing labour at less than its nominal wage has other significant implications. Suppose that a public project costs $100 million, which includes $30 million for employing out-of-work labour who would be willing to work for $15 million. The real cost of the project is $85 million. If the project generates user benefits of $110 million, the net benefit of the project would be $25 million. This would be $110 million in user benefits plus $15 million in worker surpluses less $100 million cost to taxpayers. If the increase in employment were counted separately as a benefit of the project, the employment benefits would be counted twice.

Two further notes. First, all expenditures on factors of production, including capital expenditures, should be recorded in full when the expenditure is incurred. No interest

payments are included in CBA. The discounting process captures the OC of all resources tied up in the project, whatever the source of funds. In a CBA, depreciation of capital and payments of interest on loans are not treated as expenses because this would be double counting.[4]

Second, there is an identity between WTP and OC on the one hand and consumer and producer surpluses on the other. Recall that NSB equals total benefit less total cost. Excluding externalities, for any good the benefit is the WTP amount and the cost of resources is OC. Therefore:

$$NSB = WTP - OC \tag{8.8}$$

However, consumer surplus (CS) and producer surplus (PS) equal:

$$CS = WTP - P \text{ and} \tag{8.9}$$

$$PS = P - OC \tag{8.10}$$

where P equals price. Therefore,

$$NSB = WTP - OC = CS + PS \tag{8.11}$$

This identity is useful because it is often convenient to show the net social benefit as the sum of consumer and producer surpluses.

Valuing other costs

Other costs arise may arise with negative externalities. This may raise forecasting problems regarding the impacts but does not create new valuation issues. When firms are affected, the cost is the increased cost of production or loss of output, or both.

When households experience indirect costs, such as loss of amenity or health, there may also be valuation issues associated with losses of rights discussed in Chapter 6.[5] Suffice to say here that in these cases there may be an equity argument for allowing the minimum amount that people would be *willing to accept* as compensation for the adverse impacts and be no worse off than before. However, the general practice in CBA studies is to adopt the values that individuals would be willing to pay to avoid the adverse effects.

Secondary benefits

Secondary benefits are the flow-on consequences of primary benefits. A project that generates additional income for owners of capital or for labour may create a second round of expenditure, and indeed further rounds, that raises household incomes through a multiplier effect.

Secondary benefits are not usually included in cost–benefit studies for various reasons. First, in a fully employed economy, additional expenditure cannot generate extra employment income; instead, it increases imports or displaces exports. Second, if secondary benefits occur because some resources are underemployed, secondary benefits would also arise from expenditure on an alternative project. If the secondary multiplier is M, then if for any project $B < C$, $M \times B$ is also likely to be less than $M \times C$. Thirdly, aggregate demand is a function of national monetary and fiscal policy rather than of individual projects.[6]

Of course, secondary benefits may have distributional consequences. The location of projects may advantage one region or community rather than another. These distributional effects are often of interest to government.

---

[4] In a financial analysis, loans may be included as income and repayments of capital and interest as expenses.

[5] Of course, households ultimately bear any indirect costs borne by firms.

[6] If unemployment varies by region, labour is immobile and there is no alternative project for a region, multipliers may vary and secondary benefits may be counted, but these are unusual, indeed exceptional, conditions.

Transfers

A transfer occurs when a gain to one party is offset by a loss to another. Such transfers do not affect the net social benefit of a project and so may be excluded from a CBA. However, they may also be included (as gains and losses) for completeness and transparency.

Consider, for example, changes in the government budget due to changes in tax revenues, welfare payments or subsidies. These impacts can be ignored in CBA—a dollar increase in tax revenue to government is a dollar loss to the taxpayer. The transfers affect the incidence of the benefits but do not affect resource use or net social benefit. However, for completeness it is sometimes useful to record these changes.

There are many other examples of transfers. Indeed, any pricing arrangement has transfer impacts. Certainly, the price charged for a good or service, for example a road toll or a rail fare, may affect the amount of consumption of a good, and this in turn affects the benefits and the costs in a CBA. However, the actual payment of a road toll or a rail fare is a transfer that determines the financial incidence of the benefits from a road or rail improvement.

The Social Discount Rate

Discount rate
The interest rate at which future income or payments are discounted to determine their present value

The social discount rate is an inter-temporal weighting. It generally places less weight on future consumption. Following CBA valuation principles, the rate of discount of future consumption would reflect the consumption time preferences of individuals. On the other hand, the cost of capital should reflect its opportunity cost (the rate of return foregone on the marginal alternative project). Because income from capital is taxed, the gross rate of return on capital (the alternative producer rate of discount) is generally higher than the rate of return to savers (the consumer rate of discount). This difference lies at the heart of the debate about the appropriate discount rate. In addition to the consumer and producer discount rates, we also discuss below a social time preference rate of discount, a project risk-adjusted rate of discount and a synthetic (combined) discount rate.

The consumer rate of discount, known as the private time preference rate (PTPR), is the rate at which individuals are willing to exchange marginal consumption in the present period for additional consumption in the next period. Although often described as a time preference, it may also reflect a declining marginal utility of consumption in so far as individuals expect their income to increase over time. The fact that the real rate of interest is generally positive implies that most people prefer an extra dollar of income this year to an extra dollar next year and require compensation for forgoing present consumption.

The PTPR can be calculated from market interest rates adjusted for inflation and tax. Suppose that the market interest rate (mr) for savings is 7 per cent, the marginal tax rate (t) is 30 per cent and the expected inflation rate (π) is 2 per cent. The PTPR is:

$$PTPR = \frac{1 + mr\,(1-t)}{1+\pi} = \frac{1+(0.07\times 0.7)}{1.02} = 1.028 = 2.8\% \qquad (8.13)$$

A PTPR of 2.8 per cent means that, in constant prices, individuals place equal value on \$100 of marginal consumption this year and \$102.80 of marginal consumption next year.

The social time preference rate (STPR) is a variant on the PTPR, which attempts to derive temporal weights that would maximise inter-temporal social welfare rather than private welfare. It is intended, *inter alia*, to protect the interests of future generations who do not have a say on present-day interest rates or policies on climate change. Thus, the STPR is designed to equalise the marginal social utility of consumption over time. This STPR is given by:

$$STPR = \mu + (g \times e) \qquad (8.14)$$

where μ is pure time impatience, g is the expected growth in per capita consumption and e is the elasticity of the marginal utility of consumption, that is, the percentage change in marginal utility associated with a 1 per cent change in consumption. There is considerable debate about whether it is appropriate to allow any value at all for pure time impatience (μ). The value of g is typically about 1.5. And, as we saw in Chapter 7, the estimated value of e is somewhere between 1.0 and 1.5 (Layard, 2008). The result is that the estimated STPR is usually a little lower than the PTPR.

The alternative producer rate of discount, traditionally described as the social opportunity cost of capital (SOC), is the marginal rate of return that can be achieved with alternative investment.[7] The SOC is generally higher than PTPR because it is measured before tax and includes the return to government, whereas PTPR is measured after tax.[8] Using the same nominal interest rate and inflation numbers as before, the SOC would be 4.9 per cent.

$$SOC = \frac{1+mr}{1+\pi} = \frac{1.07}{1.02} = 1.049 = 4.9\% \tag{8.15}$$

The project risk-rated rate of discount is the rate of return that projects require to attract scarce risk capital to the project. This rate varies with the risk for each project.

A synthetic discount rate (SR) incorporates the SOC and PTPR as follows:

$$SR = (w \times SOC) + (1 - w)(PTPR) \tag{8.16}$$

where w is the proportion of capital employed in a project that would be invested elsewhere and $(1 - w)$ is the proportion that reflects consumption forgone.

Selecting the discount rate

Following normal valuation principles, the forecast stream of consumption benefits would be discounted by the consumer rate of discount (PTPR) or possibly by the STPR where the impacts are intergenerational. On the other hand, efficiency requires that the return on each project should be at least as high as the return on the alternative project in the private or public sector (the SOC in the terminology used here).

This dilemma can be resolved by estimating a shadow price for capital expenditure by simulating the benefit stream foregone and discounting these foregone benefits with the PTPR. This may be explained by an example. Suppose an agency plans a project costing an estimated $100 million in year 1 and producing an estimated $4 million of benefits each year for 30 years, with a residual value of $100 million in year 31. Suppose also that the SOC is 5 per cent and the PTPR is 3 per cent. Would this be a viable investment?

If the SOC is 5 per cent, an alternative investment of $100 million would provide the equivalent of $5 million benefits per annum plus a residual value of $100 million. Discounting this income stream by 3 per cent gives a present value of $135.2 million. In other words, when the rate of return forgone is 5 per cent, an investment of $100 million for 30 years is equivalent to giving up consumption with a present value of $135.2 million. This is the real cost, or shadow price, of the capital investment. On the other hand, with a discount rate of 3 per cent, the present value of $4 million per annum over 30 years plus $100 million in year 31 is only $116.1 million. As this is less than the $135.2 million of forgone consumption, the NPV is negative and the project is not efficient.

---

[7] In some recent literature, the term "social opportunity cost of capital" has been applied to the synthetic rate of discount (the cost being a weighted function of investment and consumption displaced by the project).

[8] The return on alternative investment may also include an allowance for non-diversifiable market risk.

Although technically attractive, this 'shadow price of capital' approach is rarely employed. One reason is complexity in application. Another is that applying the SOC rate of discount to all costs and benefits provides similar results. This can be illustrated as follows. Suppose that a project has a capital cost K and a perpetual stream of annual benefits b. Discounting the benefits by the PTPR (r), their present value is:

$$PV(b)_t = \sum_{t=1}^{\infty} b_t / (1+r)^t = b / r \tag{8.17}$$

Turning to the capital cost and discounting the (perpetual) stream of benefits forgone also by the PTPR, the present value of these costs is:

$$PV(K) = (K \times \rho)/r \tag{8.18}$$

where ρ is the SOC. Therefore, a project has a positive NPV if

$$b/r - (K \times \rho/r) > 0 \tag{8.19}$$

Equation 8.19 implies

$$b > K \times \rho \tag{8.20}$$

In words, if the PTPR is applied both to the benefits (consumption obtained) and the costs (consumption forgone), a positive NPV requires that annual benefits exceed the product of the capital invested and the SOC.

While this is a special case, if all capital expenditure represents forgone capital investment rather than foregone present consumption, using the SOC discount rate generally produces the same result as using a shadow price of capital along with the PTPR. Suppose again that a $100 million investment produces $4 million of benefits each year for 30 years and a residual value of $100 million. Employing a SOC discount rate of 5 per cent, the NPV would be – $14.6 million and the project would not be efficient.

Most Australian jurisdictions recommend use of the alternative producer rate of discount in the order of 7%.[9] This allows for non-diversifiable market risk which is the return generally available in private markets. This ensures that government adopts efficient projects that benefit **both** current and future generations (providing some surpluses are reinvested). Lower discount rates are likely to lead to adoption of more inefficient projects. The reasons for not adopting a project risk-adjusted discount rate are discussed below.

Two exceptions to this general strategy should be mentioned. First, when private investors provide capital for a project and bear part of the risk, the discount rate for their contribution should reflect the rate at which they are willing to part with their capital as this represents their perceived cost and required rate of compensation. Second, if the opportunity cost reflects loss of consumption rather than investment, a consumer or synthetic rate of discount would be appropriate. However, once capital is raised, there is always an investment opportunity cost.

Uncertainty and Risk

The technical literature on risk distinguishes between uncertainty and risk. An outcome is said to be uncertain when the probability distribution cannot be forecast. An outcome is risky when the outcome is variable, but the probability distribution can be estimated. There are rational ways of dealing with risk. It is much more difficult dealing with pure uncertain events. Accordingly, wherever possible analysts attempt to estimate probability distributions for uncertain events.

[9] On the other hand, many European jurisdictions recommend use of a time preference or synthetic discount rate. See for example HM Treasury (2011).

Similar issues arise in CBA studies. Where possible, economists attempt to quantify all major project costs and benefits. Such valuations, even when approximate, clarify trade-offs and reduce the risks of wild exaggeration. However, if impacts are highly uncertain they may not be quantified. Poor quantification can discredit the rest of the work. For example, it may be difficult to forecast biodiversity loss or to value a loss of Aboriginal heritage. Such impacts should be listed in a CBA. Decision makers can then decide whether they outweigh the quantified results.

Turning to the treatment of risk, Figure 8.2 illustrates the kind of choice that may be required. This shows the probability distributions of two projects A and B. Project B has the higher forecast mean and greater variance. A CBA typically estimates the mean (weighted average) outcome. However, government may prefer the project with less variance. What advice can economists give about this trade-off between means and variances?

Risk neutrality. We start by assuming risk neutrality (an indifference to the dispersion of outcomes) for two reasons. As we will see, there are strong technical arguments for government adopting a risk-neutral strategy.

To define risk neutrality, we must first define expected value (EV). EV is the weighted average value of the possible outcomes, weighted by the probability of their occurring. Assuming discrete outcomes with a probability summing to one, the EV of a project is the product of the forecast outcomes and their probability of occurrence:

$$EV = p_1 Y_1 + p_2 Y_2 + \ldots p_n Y_n \tag{8.21}$$

where p represents the probability, Y is income outcome and there are $n = 1 \ldots n$ scenarios.

An individual is risk neutral if he or she is indifferent between a certain outcome of $\$x$ and a prospect with uncertain outcomes but an expected mean value of $\$x$. To take a simple example, risk neutrality implies that a person is indifferent between receiving a certain $1000 and a prospect offering a 50/50 chance of $2000 or $0, which would have an EV of $1000. Risk aversion means that an individual would prefer receiving a certain outcome of $\$x$ to a prospect with uncertain outcomes and an expected mean value of $\$x$.

Although most individuals are apparently risk averse most of the time, Arrow and Lind (1970) showed that risk neutrality is generally the best strategy for a community. The reasons

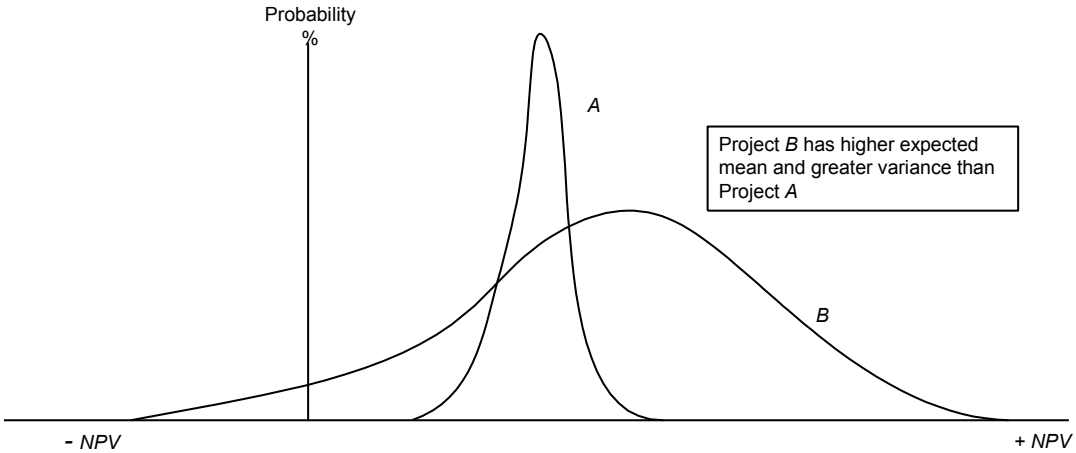


Figure 8.2 Forecast probability distributions for two projects

are risk pooling and risk spreading. **Risk pooling** occurs when risks are spread over a large number of projects. By the law of large numbers, the weighted average mean return on all projects will be achieved with little overall variance. Projects that exceed expectations offset those that underperform. Unless project risks are correlated, there is little risk to the aggregate return over all projects. **Risk spreading** occurs when the impacts of a project are spread widely over a large population and the impact on individual members of the community is small. This occurs when project costs are spread over many taxpayers. The smaller the impact, the closer are most people to risk neutrality.

Given risk neutrality, the efficient strategy is to select projects with the highest expected NPV. Suppose that project *A* has a certain NPV of $1.5 million and that project *B* offers a 50/50 chance of $5 million and –$1 million respectively. The expected NPV of *B* is $2.0 million and the expected value approach leads us to select *B*. Project selection based on expected values produces a higher aggregate return to society than does selection based on risk avoidance principles.[10]

However, risks cannot always be pooled or spread, for example for large projects in a small community. Where the risks are large relative to individual incomes, outcome variance becomes a significant issue.

Risk aversion. When variance matters, a measure of risk aversion (the preference of safety) is required. In effect, this means that we would like to know the certainty equivalent (CE) of the uncertain prospect. CE is the certain amount of income that would confer the same utility as a distribution of expected outcomes. For example, for a risk-averse individual, the CE for a prospect offering a 50/50 chance of $2000 or $0 could be, say, $700.

At community level, the CE is the amount of certain income that society collectively would be prepared to exchange for the set of uncertain outcomes generated by a project. Taking projects *A* and *B* above, distributional issues aside, society would presumably be willing to pay up to $1.5 million for *A*. However, it might not be willing to pay this amount for *B*. If the CE for *B* were less than $1.5 million, *A* would be preferred.

This implies that we can estimate CE amounts. This can be done in two ways. First, we can assume that risk aversion reflects the diminishing marginal utility of income as described in a standard concave utility function. In this case, CE values are the monetary equivalents of expected utility values which can be derived from assumptions about the nature of the utility function (see Rosen and Gayer, 2014, pp. 173–174). Secondly, CE values can be derived from household surveys. However, neither approach is commonly practised and, as a practical matter, alternative ways of handling risk are generally required.

One possible way is to raise the discount rate. This allows for the risk factor to affect the results directly and is a common practice in the private sector. However, it is not clear how much discount rates should be raised to allow for risk. Moreover, risks are not necessarily correlated with time so increasing the discount rate does not necessarily deal accurately with risk. Therefore, risk is rarely dealt with by raising discount rates in CBA studies.

In the absence of a valuation of the risk (via the CE principle or the discount rate), risk is generally treated descriptively. A comprehensive description would involve a complete mapping of the probability distribution, as shown in Figure 8.2. Where the number of uncertain variables is large, the input combinations can become very large. Various computer programs use sampling procedures with the number of iterations specified by the user. For continuous variables, the sample is drawn from a continuous distribution. The set of sample values is derived randomly and a probability distribution of results is generated from repeated

[10] Accurate calculation of expected values may require a large number of assumptions about probability distributions for variables and correlations of these distributions. This is impractical for most studies. Estimates of expected NPV are generally based on estimated mean values for each variable.

runs. Estimates of EV and variance can be derived from the resultant probability distribution. Decision makers then use this information to determine the preferred EV–variance trade-off.

A more common descriptive approach is sensitivity testing. Sensitivity tests show how the estimated NPV of a project (based on estimated mean values of variables) changes with variations in the values of inputs, such as construction cost or forecast demand. Sensitivity tests are practical and may reassure decision makers that outcomes are not sensitive to plausible changes in key variables. However, they provide an incomplete mapping. They also usually ignore the probabilities of high or low values of variables occurring (and so may over-emphasise extreme values that are unlikely to occur) and they ignore interdependencies between variables. For example, in a commercial project lower output may be offset by higher prices.

In conclusion, when risks are widely distributed and not correlated there are sound reasons why policy makers should aim to maximise expected NPV using a risk-free discount rate. When risks are not widely spread, they may wish to know the risk profile. This may be done by providing a comprehensive probability distribution of outcomes or by presenting sensitivity tests to policy makers rather than a single NPV figure. Policy makers must then decide whether to accept the risks. This may be informed by community attitudes to risk. But there is no mechanical formula to determine whether the risks should be accepted.

Distributional Issues

There are strong reasons for adopting an efficiency criterion of project value (i.e. estimated NPV). If this criterion were applied generally, most people would be better off (directly or indirectly) than if less efficient projects were regularly adopted. The indirect benefits arise because efficiency maximises collective project surpluses which can be redistributed via income transfers to compensate losers. Also, distributional objectives can generally be achieved more effectively and at less cost by income transfers than by individual projects.

However, income transfers may not produce an acceptable distribution of income. Also, losers from projects are often not compensated. Some projects with positive NPVs may benefit the rich and hurt the poor. Moreover, WTP values are based on the existing distribution of income. Because WTP rises with income, WTP measures may bias projects towards higher income households, especially when goods are provided free or at below-cost prices.[11] Therefore governments are often interested in the equity as well as the efficiency impacts of projects. This involves two steps: an analysis of distributional impacts and a policy determination.

Analysing distributional impacts. The analysis of distributional effects may appear straightforward, but it often has complications. First, the community must be divided into selected social groups because it is impractical to show the impacts on every individual. Groups may be chosen according to their nationality, income, age, sex, health status, area of residence and so on. Inevitably the selection of social groups involves value judgements. Second, when a project benefits firms, the benefits may accrue to shareholders, managers or employees, or government via taxes or consumers via lower prices. The role of prices is especially important. For example, lower production costs initially benefit firms. But, if the market is competitive, consumers will benefit from price reductions. Therefore, final impacts may differ from initial ones as impacts are shifted between groups. Third, transfer payments (mainly indirect taxes and subsidies) affect the distribution of costs and benefits although they do not affect total value of consumption. Fourth, secondary (flow-on) benefits can have significant distributional effects especially in regions where there are unemployed resources

[11] This problem does not arise when consumers pay the full costs of any service.

and potentially significant local multipliers. Evidently, a distributional analysis may require information, estimates and analysis that are not part of the basic CBA study.

Distributional impacts and decision making. As we saw in Chapter 7, estimated WTP dollar values of impacts can be converted into social utility values by using equity weights. In principle this provides an overall measure of social welfare inclusive of equity as well as efficiency.

However, we also observed that significant problems arise with the use of equity weights. The choice of weights is an ethical judgement not a technical exercise. There is no one standard set of generally acceptable weights available. Also, equity weights can produce confusing results. An unweighted NPV measures the gain in aggregate consumption. A weighted NPV has no clear meaning. Weights can lead to the uninformed adoption of inefficient projects and rejection of efficient ones. Nor does use of weights prevent adoption of projects with positive weighted NPVs that harm the poor.

In summary, policy makers often require information about distributional effects. Subject to careful analysis, this information can be provided along with the CBA. Policy makers may then weight the impacts as they see fit, but they should be well informed of the opportunity costs of the choices. There is no technical basis for the use of equity weights and their use may confuse rather than clarify the results for decision makers.[12]

CBA and Other Evaluation Methods

We consider below five other methods of evaluation. These include four that use a dollar metric and may be considered economic evaluation methods and one general non-economic method. Although all five evaluation methods are often used, none is as comprehensive or as rigorous as CBA.

Financial analysis. This is an account of the cash flows (revenues and expenditures) of a project for a specified public or private agency. The format is like that of CBA in that cash flows are recorded over the life of a project and summarised via discounting in an aggregate figure, such as net present value. Unlike CBA, a financial analysis may incorporate financing arrangements and allow for projected inflation. Thus, a financial analysis may include interest payments, use market interest rates and include forecast general price changes.

Financial analysis
An account of the cash flows of a project for a specified agency

A financial analysis shows how a project is to be financed and whether a financial profit or loss will be incurred. This is important for government budgeting and for sustaining an investment. However, it includes cash flows only for the specified agency and it does not include non-cash benefits or costs. Except in a perfectly competitive market with no externalities, a financial analysis presents only a partial picture of the costs and benefits of a project.

Cost-effectiveness analysis (CEA). This assesses the cost of achieving an outcome. Typically it is used to assess the least cost way to achieve a politically determined outcome. For example, CEA may be used to determine the least-cost way of servicing a forecast number of hospital bed days or of meeting an environmental target (say, an ambient air quality target). CEA may also be used to assess the least-cost way to raise a tax revenue target. In this case the aim is to determine the tax (or taxes) that minimise the deadweight loss arising from distorting the behaviour of firms and households.

Cost-effectiveness analysis
Analyses the minimum expenditure required to achieve a given target

[12] Despite these caveats against equity weights, CBA typically adopts standard equity values for life and travel time savings in order to avoid biasing outcomes in favour of higher income households.

CEA is employed widely where outcomes are determined politically. In these cases, CEA helps to achieve productive efficiency. However, compared with CBA, CEA is very limited because it does not value outcomes. It cannot be used to compare options with varying levels of output or to evaluate options with multiple outputs. Many policy issues cannot be reduced to cost-effectiveness issues.

National output analysis. Many policies are evaluated by reference to gross domestic product (GDP). For example, changes in interest rates or in fiscal policies, such as the carbon tax, may be assessed by their estimated contribution to GDP.

GDP can be a useful measure of the desirability of short-run macroeconomic management measures. Other things being equal, an increase in GDP is generally welfare enhancing. However, GDP is not a generally satisfactory measure of social welfare. GDP does not allow for externalities, non-marketed goods, the value of leisure, terms of trade effects or national versus foreign interests. It is therefore an inappropriate evaluation criterion for many major public policies or projects.

Economic impact analysis (EIA). This usually means analysis of the income impacts of a project in a selected region. Typically, the local income impact (ΔY) of a project is estimated as a function of the local expenditure (E) and the multiplier (M):

$$\Delta Y = E \times M \tag{8.22}$$

$$M = 1/(1 - MPCL) \tag{8.23}$$

where $MPCL$ is the marginal propensity to consume locally produced goods. The multiplier falls as $MPCL$ declines. $MPCL$ declines with three main leakages: taxes, savings and expenditures on imports to the region. Suppose that taxes and savings account for 30 and 10 per cent of marginal income respectively, and that imports account for half of the 60 per cent balance that goes to expenditure. Then $MPCL$ = 0.30 and M = 1.43. Suppose that an externally financed project costs \$10 million, with half spent on local resources. The estimated gross regional income impact would be (\$5m × 1.43) = \$7.15 million.

However, this assumes that local employment (at primary and secondary stage) has no opportunity cost. Suppose that of the \$5 million spent on local resources, \$4 million employs resources that would be otherwise employed in the absence of the project. The net local income effect would be \$1m × 1.43 = \$1.43m. This is a much more modest result than is typically found in an economic impact analysis. The result would be lower still if the secondary (multiplier) employment has any opportunity cost. We should also note that a CBA would include a \$1 million project benefit for the employment of otherwise unemployed resources.

Done properly, EIA studies can show the net regional income impact. But they have several weaknesses. They are limited geographically and ignore non-monetary effects. They regard expenditures as benefits rather than as (opportunity) costs. Thus, the economic effect is always positive, which is why industry groups like such studies. But an effect that is always positive is of little help to policy makers.

Multi-criteria analysis (MCA). This has various forms and names, for instance planning balance sheets, goals achievement matrices, point scoring methods and so on. But essentially there is a single underlying method.

Table 8.3 Simple example of multi-criteria analysis

Criterion	*Criterion weight*	*Project A*		*Project B*	
		Score out of 100	*Weighted Score*	*Score out of 100*	*Weighted score*
Traveltime	40	80	32	50	20
Environment	30	40	12	80	24
Total cost	30	30	9	60	18
Total	100		53		62

First, the analyst establishes the objectives (or criteria). In the simple example in Table 8.3 the criteria are travel time savings, environmental benefits and cost. Second, weights are attached to each criterion that reflects its importance, in this case 40, 30 and 30 respectively to an arbitrary total (here 100). Third, each option (here projects *A* and *B*) is scored against each criterion (again here they are each scored out of 100). The score measures the extent to which each option meets the respective objective. Fourth, weighted score (the product of the criterion weight and the score) is estimated. Finally, the weighted scores are added to a total.

The MCA method helps to organise issues systematically and transparently. Policy makers can then make trade-offs between objectives. However, there is subjectivity and arbitrariness at all the main points in the process: in selecting criteria, in allocating weights to each criterion and in estimating points for each option for each criterion. The fundamental weakness of MCA and virtually all non-economic methods of valuation is that, unlike CBA, they are not based on any valuation principles. In the absence of such principles the evaluation process is wide open to manipulation.

Conclusions

CBA is the only form of economic analysis that presents a complete welfare assessment. CBA is based on principles of value and provides practical methods to implement these principles. CBA provides transparent, testable and efficient results. The other forms of economic analysis are all incomplete in one or another way.

However, positive net social benefits are not always equitable. The simple aggregation of WTP amounts can result in well-off individuals gaining at the expense of less well-off individuals. Also, the future is inherently uncertain and forecasts, often based on hard-to-forecast inputs from other disciplines, are inevitably subject to ranges or variance.

Economists have an important role in estimating efficiency/equity trade-offs and plausible ranges of outcomes so that decision makers are fully informed. However, determining the preferred efficiency/equity outcome and the preferred amount of risk aversion/safety are ultimately political issues.

Summary

- Cost–benefit analysis is the most common and comprehensive form of economic evaluation.
- CBA is based on the valuation principles of welfare economics. It values the costs and benefits in monetary terms as they would be valued by the parties affected.
- The estimated net present value shows whether expected benefits exceed costs. If the estimated NPV is positive, a project or policy is described as efficient.
- The internal rate of return and the benefit–cost ratio are other summary measures of project efficiency.
- Market prices often provide accurate indicators of marginal benefits and costs. However, where prices do not fully reflect social values, estimates of consumer surpluses and shadow prices are needed.
- Future costs and benefits are discounted to present values. In Australia, this generally involves using the return on alternative investment as the social rate of discount.
- When risk is widely spread, maximising expected NPV is usually an efficient and fair strategy. However, when a policy entails significant risk for some group(s), policy makers may wish to know the likely variance in outcomes as well as the expected mean outcome.
- When the distributional impacts of a policy adversely affect some social group, especially a disadvantaged group, an analysis of distributional impacts is likely to be required. However, equity weights are rarely employed.
- Other methods of evaluation include financial analysis, cost-effectiveness analysis, GDP studies, economic impact analysis, and multi-criteria analysis. None is as comprehensive or as rigorous as cost–benefit analysis.

Questions

1. Consider two projects, *A* and *B*, with capital costs of \$100 million and \$20 million respectively. Given an opportunity cost of capital of 7 per cent, *A* provides an NPV of \$30 million. *B* provides an NPV of \$10 million. Explain why *A* is better than *B*, although it costs five times as much and produces only three times the net benefit.
2. Is the net present value always the best measure of project value? When might other measures of project value be preferred?
3. In a CBA, what is an appropriate price for using a tonne of steel: the domestic market price, the market price less indirect taxes or an import or export price?
4. Suppose that the government pays Amy \$120 per day to work on a new project. Amy has a leisure preference of \$60 a day and currently receives an unemployment benefit of \$30 per day. What is the social cost of employing Amy on a public project?
5. Does the cost of finance for a project affect the discount rate? Why is the government borrowing rate generally not the appropriate discount rate?
6. Should the opportunity cost of capital discount rate or the social time preference rate of discount be applied to assess projects to reduce greenhouse gas emissions and long-term global warming? Explain your answer.
7. What are the weaknesses of sensitivity tests as a means of informing policy makers about the variance of project outcomes?
8. Can certainty equivalent amounts be estimated? Is this a useful tool for evaluating uncertain outcomes?
9. Why may cost–benefit analysis show that it is efficient to locate polluting industries in poor countries? If this is the case, does this discredit cost–benefit analysis?
10. If a national highway is substantially upgraded and freight costs between two cities are reduced by 30 per cent, who would benefit: the trucking firms using the highway, the business shipping freight or consumers? Give reasons.
11. The government is considering constructing a new urban freeway for \$500 million. This would save 30 000 commuters 10 000 hours a day (10 minutes each way) for 200 days a year. Suppose that the road can be built in one year, commuters value their travel time savings at \$20 per hour and the road has a life of 30 years after year 1 and a residual value estimated at cost at \$500 million in year 32. There are no other costs or benefits.
    i. Would the road provide a net social benefit with a discount rate of 7 per cent?

    Now suppose that the government charges a road toll of \$2.50 each way and that traffic falls to 20 000 commuters twice a day.

ii. What would be the financial outcomes using a 7 per cent discount rate?

iii. What would be the net social benefit if the remaining commuters had an average value of travel time of $25 per hour?

iv. What are the implications of the road toll?

12. When, if ever, should other evaluation methods be preferred to cost–benefit analysis?

Further Reading

Asian Development Bank (2013) *Cost-Benefit Analysis for Development: A Practical Guide,* ADB, Manila.

Boardman, A., Greenberg, D., Vining, A. and Weimer, D. (2011) *Cost–Benefit Analysis, Concepts and Practice*, 4th ed., Prentice Hall, New York.

Department of Finance and Administration (2006) *Handbook of Cost–Benefit Analysis*, Commonwealth of Australia, Canberra.

European Commission (2014) *Guide to Cost-Benefit Analysis of Investment Projects*, DG Urban and Rural Projects.

HM Treasury, 2011, *The Green Book, Appraisal and Evaluation in Central Government*.

Layard, R. and Glaister, S. (1994) 'Introduction', in R. Layard and S. Glaister (eds), *Cost–Benefit Analysis*, Cambridge University Press, Cambridge.

Lesser, J. and Zerbe, R.O. Jr. (1998) 'A practitioner's guide to benefit–cost analysis', pp. 221–268 in F. Thompson and M.T. Green (eds), *Handbook of Public Finance*, F. Marcel Dekker, New York.

New Zealand Treasury (2015) *Guide to Social Cost-Benefit Analysis*.

NSW Treasury (2017) *Guide to Cost-Benefit Analysis*, TPP 17-03, NSW Treasury, Sydney.

Rosen, H. and T. Gayer (2014), *Public Finance*, 10th ed, Chapter 8, McGraw-Hill, New York.

Victorian Department of Treasury and Finance (2013) *Economic Evaluation for Business Cases, Technical Guidelines*.

Vining, A. and Weimer, D.L. (2010) 'An assessment of important issue concerning the application of benefit-cost analysis to social policy', *Journal of Cost–Benefit Analysis*, 1, 1–37.

Part

4

Public Choice

Chapter

9

Public Choice and Individual Values

Politics is not an exact science, as the professors are apt to suppose. It is an art.

Otto von Bismarck, German Chancellor, Prussian Upper House

In a democratic society, collective decisions usually involve voting procedures of one kind or another. It is natural therefore to ask whether these voting processes result in preferred collective outcomes, however these may be defined, and indeed whether the processes can be said to maximise social welfare. Again, a major problem is aggregation of preferences. How can collective decision-making processes take the sets of individual preferences and produce overall collective outcomes that satisfy these preferences? In this chapter, we examine methods for making collective decisions and their strengths and weaknesses.

Unanimous Collective Choice

An attractive feature of market trades is that exchanges are voluntary and all participants expect to gain. No one loses: no one is forced to purchase something that he or she does not want or obliged to pay more for a good than its expected value. In the economics literature, this is described as a Pareto- beneficial outcome.

In the public sphere, collective provision of a good creates a Pareto-beneficial outcome only if someone votes for it and no one votes against (this is described as "weakly unanimous".[1] If anyone votes against a proposal, a decision to adopt it involves some coercion. A coercion cost exists when someone must accept a collective decision to which he or she would not voluntarily assent. Thus, a coercion cost exists when a public good is over- or undersupplied for some individual. Oversupply occurs when an individual contribution for a marginal unit of the good exceeds what he or she would willingly pay for it. Undersupply occurs when an individual would be willing to pay for a marginal increase in the good, but this is not provided. Because a (weak) unanimous vote is necessary to avoid coercion, we need to consider how this vote might be achieved and whether it is practical.

[1] Strong unanimity requires everyone to vote for a proposal. The statements in this section require only weak unanimity.

Consider provision of a pure public good. All citizens enjoy the same quantity of the good, but if their preferences vary they are unlikely to be willing to pay the same amount for it. To achieve unanimous agreement on the quantity to be supplied, each individual must be able to make a different payment for the good (or at least for a marginal unit of the good). An individual's payment is a function of the total expenditure on the public good and his or her share of this (which is described as the tax price). The tax price may vary for each individual. This approach differs from private exchange where each person pays the same price for a good but consumes different quantities.

The literature (Mueller, 2003, Chapter 4) suggests that a public agency can determine the unanimously demanded amount of a public good in two ways. Both involve a central planner asking citizens what they want. Using one approach the planner announces various tax prices and asks citizens to state their desired quantity of the public good. Unanimity is achieved when everyone wants the same quantity.[2] Alternatively the planner can announce various quantities of the public goods and ask citizens what tax prices they would be willing to pay for these quantities. Unanimity is achieved when these voluntary tax prices sum to 100 per cent of the cost.

The following example illustrates the first approach. The problem is to determine the quantity of play pieces in a local park. Figure 9.1 shows Anne and Bruce's demand curves (D_A and D_B) for play equipment. Because the equipment is a non-excludable public good, the same amount of equipment must be available to both families. Therefore, total demand is obtained by adding up what Anne and Bruce are willing to pay for each quantity of the good. The marginal cost of producing an extra play piece is assumed to be constant. The efficient quantity is Q_E play pieces, because at this point the collective benefit (as measured by collective willingness to pay) of an extra play piece equals the marginal cost of supplying it. At any other level of supply, a change in the amount supplied could make at least one of them better off without making the other worse off. Note that neither Anne nor Bruce alone would be willing alone to pay for the Q_E equipment, so that individual action will be inefficient and collective action is required.

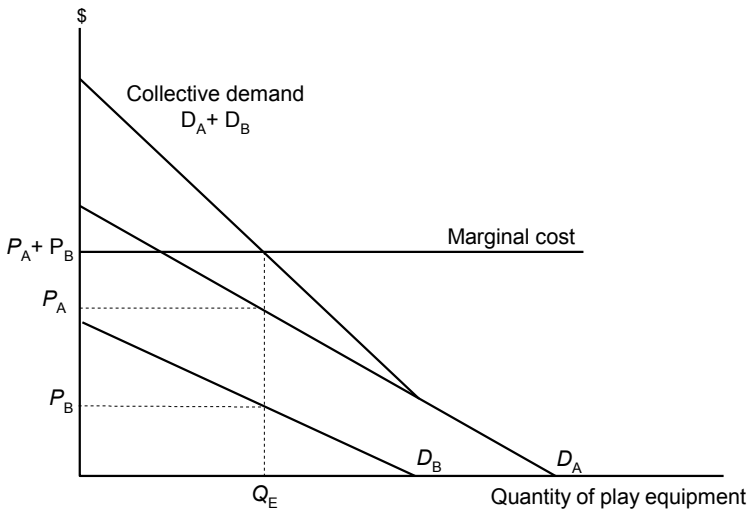


Figure 9.1 Efficient provision of a public good

[2] These personalised tax prices are often described as Lindahl prices. Lindahl (1919) proposed that tax shares should be adjusted until citizens agreed unanimously on the quantity of the public good to be supplied.

Table 9.1 Example of unanimous agreement

Option	*Proposed tax shares for Anne and Bruce*	*Play pieces demanded by Anne*	*Bruce*
1	50-50	10	5
2	75-25	6	8
3	66-33	7	7

A unanimous outcome is achieved when Anne and Bruce agree on both their tax contributions and their preferred number of play pieces. Suppose that the planner proposes that they share equally the cost of the equipment and finds that, as shown in Table 9.1, with these tax rates Anne wants more play pieces than Bruce does. The planner then adjusts the tax shares for Anne and Bruce to 75 per cent and 25 per cent respectively and finds that Bruce wants more pieces. Finally, when the planner adjusts the tax shares to 66 per cent and 33 per cent respectively, Anne and Bruce each vote for seven play pieces. Given these tax shares, there is unanimous support for this number of pieces.

Analysis of unanimity. Unanimous agreement is attractive. Each voter nominates the amount that he or she is willing to pay for the marginal unit provided. All collective decisions are based on the benefit principle. Given their initial income, no one loses from the outcome. Unanimity achieves a Pareto-efficient outcome for public goods like competitive markets achieve for private goods. At this efficient point, the marginal rate of substitution of public for private goods for each individual is equal to their tax price.

However, there are major difficulties with obtaining this Pareto-efficient outcome. One is the possibility that individuals will not reveal how much they are willing to pay for public goods when they have an incentive to conceal their preferences. Both Anne and Bruce have an incentive to understate what they are willing to pay for play pieces because this would reduce their tax share. This is known as free riding. **Free riding** occurs when individuals attempt to obtain a benefit without paying for it.

The traditional public finance assumption is that the incentive to free-ride encourages people to understate their demand for public goods and results in under-provision of public goods. The presumption is that each public good is financed by a separate tax.[3] On the other hand, when public goods are financed from consolidated government revenue, which is much more common, responses to willingness-to-pay questions have minimal effect on taxes paid. In this case, voters have an incentive to over-state their demands for public goods, in the hope that this will influence government to provide more of the good, and over-provision of public goods is the likely outcome.

However public goods are financed, individuals often have an incentive to misrepresent the amount they would be willing to pay for public goods. Such dishonest responses are generally described as strategic behaviour! In the discussion of public goods in Chapter 11, we examine further whether individuals do respond strategically and whether they can be given incentives to respond honestly. It appears that strategic behaviour is a chronic problem in collective choice situations. Although economists have devised a survey method that should elicit honest answers, it is too complex to be practical. Generally, the requirement of unanimity encourages strategic behaviour because each voter knows that no decision can be made without his or her agreement.

[3] The traditional assumption is derived from Wicksell (1896).

The second major problem with a unanimity requirement is the cost and time needed to achieve unanimity, if indeed it can be achieved. The cost of finding the tax prices required for unanimity for each public good would be very high. Indeed, given the difficulty of achieving this, it may be asked why economists consider such an impractical notion in the first place. The answer is that (weak) unanimity provides a welfare benchmark, like Pareto efficiency in private exchange, in which there is no coercion. All non-unanimous approaches to collective decisions involve some coercion. It is therefore necessary to be sure that there is no perfect collective solution.

Other Collective Choice Methods

Given that all non-unanimous decisions coerce someone, which non-unanimous method of social choice provides the best social outcome? To answer this question, Buchanan and Tullock (1962) proposed the **optimal majority rule**. This rule recognises that collective decisions have two main costs: in their terminology, these are external and decision-making costs. An external cost (also described as a coercion cost) occurs when an individual loses utility due to a collective decision. Decision-making costs are the costs incurred in reaching decisions. The optimal majority is the proportion of voters required to agree to a decision that minimises the sum of external and decision-making costs.

The concept of the optimal majority is illustrated in Figure 9.2. There are N voters. External costs fall as more voters must agree to a collective decision. On the other hand, decision-making costs rise with the number of people involved in making the decision. The optimal majority is K/N voters, where K is the number of voters that must agree to a proposal. This is the majority that minimises external (coercion) plus decision-making costs.

The optimal majority varies with the issues to be determined. The optimal majority is lower when there is consensus among individuals and low potential for coercion, and when individuals have a high cost of time. Conversely, the optimal majority is larger when the interests of minority groups are severely threatened. It is often argued that the more important the social issue, the greater should be the majority required.[4] Thus changes to constitutions

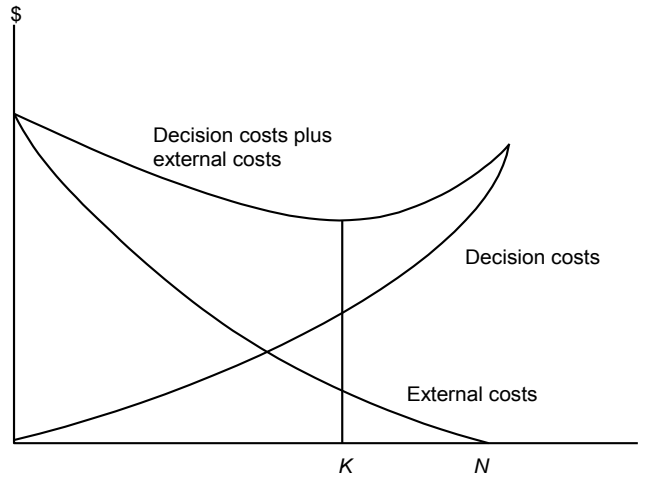


Figure 9.2 An optimal majority

[4] Rousseau (1762) argued that 'the more grave and important the questions discussed, the nearer should the opinion that is to prevail approach to unanimity'.

often require special large majorities.[5] However, requirements for large majorities imply a conservative view—the votes of supporters of the status quo have more weight than votes of supporters of change. This may be justified by the potential costs of change. It does not seem to be justified by appeal to the importance of the issue (however that may be defined).

Simple majority voting

The most common decision rule is simple majority voting (SMV). Where there are N voters, SMV requires that at least one voter above $N/2$ must vote for a proposal before it is approved.[6] If voters have a similar intensity of preferences, a simple majority vote would represent a potential Pareto improvement because the gainers could compensate losers and no one would be worse off. However, the assumption of equal intensity of preference is unrealistic.

Nevertheless, SMV is practical and efficient in that it has lower decision-making costs than decision methods requiring larger majorities. A simple majority vote is the smallest possible majority that can make a decision without self-contradiction. A less than majority vote (say, a 45 per cent vote) could cause a proposal to be both passed and rejected. This would be impractical and lead to high decision-making costs. Note that this would imply a kink in the decision cost curve shown in Figure 9.2, with decision costs rising when fewer than $N/2$ voters are required for a collective decision.

SMV may also be regarded as fair in that it gives each vote an equal weight. Higher majority requirements allow a minority to block the interests of the greater number. On the other hand, out of all majority vote methods, a simple majority vote provides the greatest potential and incentive for coercion. A 51–49 vote maximises the number of possible losers and the potential gain per individual from redistribution. Thus SMV is fair when one voter–one vote is appropriate, but not when it results in high coercion of some voters.

SMV has other problems. A major one, known as the voting paradox, is the potential for inconsistency or intransitivity. Suppose there are three options, X, Y and Z. Transitivity requires that if X is preferred to Y, and Y is preferred to Z, then X should be preferred to Z. The choice is intransitive if Z is preferred to X.[7] As an example, suppose that Anne, Bruce and Cathy must choose how to allocate a million dollars—on tertiary education, health care or Olympic athletes. Table 9.2 shows their preferences in rank order (1 = first). A 2–1 majority (Anne and Bruce) prefers spending on health care to tertiary education. But a 2–1 majority (Bruce and Cathy) prefers support for Olympic athletes to health care. Yet another majority (Anne and Cathy) favours tertiary education over Olympic athletes. The result is intransitive.

When social choice is intransitive the outcome from SMV depends on the order in which votes are taken. If the order is random, the outcome is determined by chance. Alternatively, if a chairperson knows the preferences of voters, he or she may rig the voting procedure to obtain a personally desired outcome. This process, called **agenda setting**, is clearly undemocratic. On the other hand, if voters realise what is happening, they may vote strategically. For example, although Anne prefers health care to tertiary education, she may know that if health care defeats education, the majority will then support Olympic athletes, which is her third preference. However, she can ensure that Olympic athletes lose if she allows education to defeat health care!

---

[5] For example, as noted in Chapter 2, changes to the Australian Constitution require the support of both a majority of all voters and a majority of voters in a majority of the states.

[6] If more votes are required to approve a proposal, the majority is described as a 'qualified majority'.

[7] If an individual prefers X to Y and Y to Z, she is said to be irrational if she then prefers Z to X. Thus the requirement for transitivity is sometimes described as the axiom of rationality.

Table 9.2 Ranking over issues

Voters	*Tertiary education*	*Health care*	*Olympic athletes*
Anne	2	1	3
Bruce	3	2	1
Cathy	1	3	2

Another concern with SMV is that it may result in cycling over issues and instability. In the resource allocation example above, a majority can be found to vote against any winning proposal, leaving open the possibility of its reversal. SMV may also produce unstable outcomes on distributional issues. This reflects the phenomenon known as **cycling across gains**. Suppose that our three voters must decide how to distribute $1000 of income benefits between themselves. The following could occur.

1. Anne and Bruce could split the gains 50–50 and take $500 each, leaving Cathy with nothing.
2. Cathy then proposes a 40–60 split to Anne who gets $600, leaving Bruce empty handed.
3. So Bruce proposes a 40–60 split to Cathy who gets $600, which leaves Anne with nothing!

Cycling across gains is common in politics. Evidently the outcome may not be equitable. In addition, there may be no stable outcome to this process.

The nature of preferences and the median voter

Inconsistent voting is most likely to occur when voters allocate expenditures over different goods. But it can also arise when voters allocate expenditure to a single public good. Suppose that Anne, Bruce and Cathy must vote for expenditure on the local public school where there is a private substitute. Anne is rich and will send her children to a private school and gives first rank to low public expenditure. Bruce has a medium income. He favours high expenditure on the public school because he would then send his children to it, but his least preferred option is medium expenditure because he would pay medium taxes and send his children to a private school. Cathy is poor and wants medium expenditure on public schools but does not want higher taxes. Their preferences are ranked in Table 9.3. A majority prefers low expenditure to medium, medium expenditure to high expenditure and high expenditure to low expenditure. The outcome is intransitive.

Evidently, majority voting can also produce inconsistent results for single-issue decisions. This may arise from multi-peaked preferences.[8] A voter has a single-peaked preference if her

Table 9.3 Voters' ranking of expenditure on local public school

	Expenditure on public school		
Voter	*low*	*medium*	*high*
Anne (rich)	1	2	3
Bruce (medium income)	2	3	1
Cathy (poor)	3	1	2

[8] Multi-peaked preferences do not always create inconsistent voting.

Median voter
Where preferences on an issue can be represented on a single continuous scale, the median voter is the person with as many voters on either side of his or her position

utility falls as she moves away from her optimum outcome in any direction. Her preferences are double-peaked if her utility falls and then rises as she moves away from her peak. As shown in Figure 9.3, Anne and Cathy have single-peaked preferences for public school expenditure. However, Bruce has a double-peaked preference.

On the other hand, if voter preferences are single-peaked and the issue is uni-dimensional, an equilibrium result can be obtained and cycling avoided. Suppose there are five voters, now including David and Emma, and that each voter obtains utility from public expenditure as shown in Figure 9.4. All preferences are single-peaked. Cathy is in the central position and is the median voter. David and Emma prefer high expenditure. Anne and Bruce prefer less expenditure. In this case, majority voting yields a stable result (Q_e), which is the amount preferred by Cathy. All other possibilities are defeated by majority vote.

Median Voter Theorem
Where preferences on an issue can be represented on a single continuous scale and voting is by simple majority vote, the preference of the median voter is decisive

In general, if voters vote on a single issue and all preferences are single-peaked, majority voting produces a stable equilibrium outcome that reflects the peak preference of the median voter. This is known as the **median voter theorem**. Because many voting situations approximate to these situations, the median voter theorem has many significant practical implications (see Chapter 10).

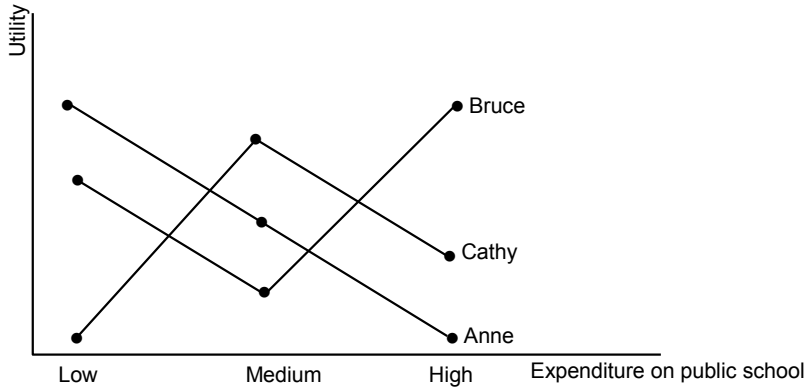


Figure 9.3 Single and double-peaked preferences

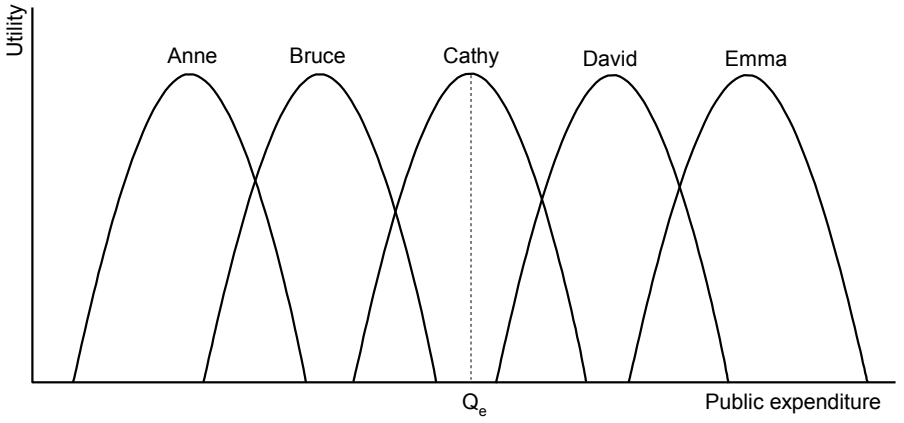


Figure 9.4 Single-peaked preferences, public expenditure and the median voter

Table 9.4 Voter utility and vote trading

	Utility from increased expenditure on	
Voter	*Environment*	*Defence*
Anne	+7	−2
Bruce	−1	+5
Cathy	−3	−4
Total	+3	−1

What are the welfare implications if the median voter's preference is decisive? There is in general no reason to suppose that the median voter's preference for a public good will necessarily coincide with the efficient amount of the good—which would require that the sum of marginal benefits from the good equal the marginal cost of the good.

Intensity of preference and vote trading

In simple majority voting, each vote has equal weight irrespective of the strength of a voter's preferences. This makes voting different from market transactions in which individuals can express preferences by spending more on some goods than others.

However, when voters feel strongly they may be able to influence the outcome by trading votes (American texts describe this as log-rolling). An example may show how this works. Table 9.4 shows the utility that Anne, Bruce and Cathy obtain from increased expenditure on environmental protection and defence. Anne strongly supports more expenditure on the environment but not on defence. Bruce's preferences are the opposite. Cathy opposes more spending on the environment and defence. With separate votes on each issue, a 2–1 majority would vote against more spending on either item. However, both Anne and Bruce prefer an overall outcome in which spending increases for both environmental protection and defence. They can achieve this by vote trading—Anne votes for more defence if Bruce votes for environmental expenditure.

Vote trading is common. There are many opportunities for it as the political process provides a continuous stream of choices. For vote traders, trading provides mutually beneficial exchange. It is how most democracies function.

Is vote trading desirable? It has some desirable characteristics. It allows for intensity of preference and it tends to reduce cycling over issues. On the other hand, it may encourage strategic behaviour. Moreover, it may not increase aggregate welfare. In our example, total utility increases with more expenditure on the environment, but falls with higher expenditure on defence. It is easy to produce other examples showing vote trading resulting in good or bad projects as judged by overall net benefit. Thus, there can be no unequivocal judgement on vote trading.

Rank-order voting or point scoring

Rank-order and point voting are two other common methods for making collective decisions. Both allow for intensity of preferences, albeit crudely. With rank-order voting, each voter ranks the options, for example 1 for the best option, 2 for the next best and so on. The option with the lowest total score is the preferred option. With point scoring, voters may be given a number of points, say 100, to allocate to the options. In this case the option with the highest number of points would win.

Suppose that Anne, Bruce and Cathy rank four options (W, X, Y and Z) as in Table 9.5 overleaf. Option X has the lowest score and wins. However, if a choice were made between

Table 9.5 Rank-order voting

	Rank given to the four options			
Voter	***W***	***X***	***Y***	***Z***
Anne	1 (1)	2 (2)	3	4
Bruce	2 (1)	3 (2)	4	1
Cathy	4 (2)	1 (1)	2	3
Total	7 (4)	6 (5)	9	8

Note: Figures in brackets are ranks when *Y* and *Z* are not considered.

only the two best options (*W* and *X*), with the irrelevant alternatives (*Y* and *Z*) eliminated, *W* has a lower total score than *X* and would win (see figures in brackets).

There are several concerns about rank-order voting. First, the choice between the two most preferred options may depend, as in this example, on how options are ranked relative to options that are irrelevant to the main choice. The outcome depends on the alternatives included. Second, voters do not have an equal say in the outcome. Cathy's strong dislike of *W* causes *X* to be preferred when all options are included. This is of course the consequence of allowing for intensity of preference. Third, rank-order voting is vulnerable to strategic behaviour. For example, if Anne strongly wants *W* to win, she could rank *X* equal to four instead of two and so ensure that *W* wins even with all four alternatives. Similar disadvantages apply to point scoring voting methods.

Arrow's General Impossibility Theorem

Given the problems with each of these voting methods, is there any acceptable way to rank social states or make collective decisions based on individual preferences over these social states? In other words, is it possible to determine a social ordering of states based on the preferences of individuals over these states rather than on the utilities of individuals in these states?

In an attempt to answer these questions, in a famous book entitled *Social Choice and Individual Values*, Kenneth Arrow (1951) proposed that an acceptable social choice process should satisfy six ethical criteria.[9]

1. *Decisiveness*. The decision rule should be able to rank all possible alternatives or outcomes.
2. *Unrestricted domain*. The decision rule must produce a social ordering over all outcomes whatever the nature of individual preferences towards these outcomes, including, for example, multi-peaked preferences.
3. *Transitivity*. The rule should provide a consistent ranking of all possible alternatives. If *A* is preferred to *B* and *B* is preferred to *C*, then *A* should be preferred to *C*.
4. *Independence of irrelevant alternatives*. Society's ordering over any two alternatives A and *B* should depend only on its ranking of these two alternatives. It should not depend on how these alternatives are ranked relative to any other (irrelevant) alternatives.
5. Responsiveness (the Pareto postulate). The decision rule must respond to individuals' preferences. If one individual prefers *A* to *B* and no one prefers *B* to *A*, then society prefers *A* to *B*.
6. *Non-dictatorship*. The social ordering of alternatives must not reflect the preferences of only one individual when others have opposing preferences.

[9] These criteria are sometimes called axioms of collective choice.

Arrow's Impossibility Theorem shows that no collective decision rule satisfies all these criteria. The proof runs broadly as follows.[10] The unrestricted domain criterion allows all possible individual preferences. When preferences differ, as they often do, no unanimous choice emerges. It is necessary to choose between alternatives that are opposed by some voters. To satisfy the axiom of independence from irrelevant alternatives, society must choose between only two alternatives at a time. However, procedures such as majority voting that yield social choices between two alternatives can produce inconsistent ranking when three successive pair-wise choices are made. A choice must be made between the various options or the result is indecisive. But there is no way to produce consistent choice that is not imposed or dictatorial.

Arrow's Impossibility Theorem
No collective decision-making rule complies with six generally accepted ethical criteria for voting situations

Arrow's approach attempts to aggregate the ordinal preferences of individuals. These preferences are simply individual orderings of alternatives. The approach precludes interpersonal utility comparisons. It is difficult to obtain consistent and acceptable decisions using only ordinal preferences.

By contrast, cardinal preferences would imply that individual utilities could be measured. If the preferences of individuals could be quantified and if a social welfare function could be agreed, a consistent social ordering over outcomes could be obtained. A social welfare function allows preferences (utilities) to be weighted in some way. But this is inconsistent with the criterion of non-dictatorship. Preferences could also be weighted indirectly by restricting the preference domain or by adopting a point scoring system. However, these approaches are ruled out by the criteria of unrestricted domain and responsiveness or independence of irrelevant alternatives respectively

The criteria and their implications. Arrow's theorem depends on the choice of criteria. These criteria are normative. Accordingly, opinions as to their desirability may differ. Although Arrow's criteria are generally regarded as ethically mild, there is an extensive literature on whether any of the criteria may reasonably be modified and, if so, what the implications would be. There is little support for relaxing the responsiveness and non-dictatorship criteria (5 and 6 above). However, the importance and applicability of the other criteria have been questioned.

Consider first the criteria of decisiveness and unrestricted domain. Limiting the range of alternatives to be determined or the range of permissible preferences can improve the chances of reaching agreed and consistent collective decisions. For example, limiting preference orderings over uni-dimensional choices to single peaks produces a stable result based on the preferences of the median voter. For many such choices, preferences are indeed single-peaked. Also, the more homogeneous are preferences over the choice set, the more easily are consistent decisions achieved. For example, as discussed in Chapter 33, the process of "voting-with-the-feet" whereby individuals choose the local government area where they live and the associated public goods/taxpackage encourages homogeneous local communities and consequently improves the allocation of resources to local public goods. It follows that the range of preferences is often compatible with consistent collective decisions. The range of issues may also be restricted by placing legal constraints on issues that are subject to collective decision making, for example by protecting freedom of religious association. However, stringent restrictions on issues for collective decisions or on preferences would not be acceptable in most democracies.

Some commentators (such as Buchanan and Tullock, 1962) argue that transitivity is not critical. Intransitivity is not common when choices involve uni-dimensional issues or few alternatives. More fundamentally, it may be argued that although intransitivity is irrational for an individual, it is not critical for society. For example, it may be argued that the lack of a

[10] The explanation here draws on Mueller (2003).

stable majority for expenditure on tertiary education, health care or Olympic athletes does not matter. Society can be regarded as indifferent between the alternatives. What is needed is a practical choice process that produces an acceptable alternative when social deadlock exists. So long as the process of choosing between alternatives is fair, the outcome may be regarded as acceptable. Suppose that three leading tennis players (*A*, *B* and *C*), are playing in a tournament and that because of their varying styles *A* would defeat *B*, *B* would defeat *C* and *C* would beat *A*. The winner depends on the draw. Providing the draw is fair, the result is considered fair.

Possibly the most contentious criterion is independence of irrelevant alternatives (IIA). In a simple pair-wise choice, ordinal preferences are sufficient. The introduction of other alternatives allows for intensity of voter preferences and implicitly for interpersonal comparison of utilities. Thus relaxing the IIA criterion has potential advantages. Many decision procedures employ some form of point scoring to rank many alternatives. On the other hand, it is not clear how, or by what criteria, additional alternatives should be included. Also, when additional alternatives are included there is the possibility of strategic behaviour.

In conclusion, no collective decision rule fully satisfies the mild ethical criteria and the process proposed by Arrow. Nor have subsequent studies been able to show that any single collective decision rule would satisfy other ethical criteria that might be generally acceptable. However as Sen (1985) pointed out, social justice requires that society collectively makes interpersonal comparisons and Arrow's approach was avowedly designed to avoid interpersonal comparisons. As a practical matter we must be willing to make interpersonal judgments. This implicitly or explicitly involves agreeing a social welfare function of some kind.

Conclusions

All methods of making collective decisions produce unsatisfactory results in some circumstances. Even weak unanimity is generally impractical. The most common collective choice procedure (simple majority voting) may produce inconsistent or unstable results, tends to reflect the preferences of the median voter and may disadvantage minorities. Larger majority requirements protect minorities but are costly in decision time and biased against change. Vote trading enables strong preferences to be allowed for and reduces intransitive voting, but it does not prevent strategic behaviour or coercion of minorities.

Inevitably government decisions suffer from the inherent difficulties of collective choice. Many decisions are inconsistent or involve log-rolling or strategic behaviour. Many decisions fail to reflect individual values. Even a well-intentioned public agency may not always produce efficient outcomes. Also, outcomes may not be stable. For any policy package that a majority supports, another majority may be found for a package that overturns some of the provisions of the original package.

However, the technical difficulties of public choice should not be over-rated. The technical difficulties associated with making collective decisions do not mean that most collective decisions are unsatisfactory any more than the existence of some market failures means that most markets are inefficient. The lack of a perfect collective choice rule is not a critical practical issue. The issue is whether reasonable decision rules can be found. The criteria of decisiveness, domain, transitivity and independence of irrelevant alternatives can each be relaxed in some circumstances without significant social cost.

However, the most important social choice problem is the problem of interpersonal comparisons. Ultimately interpersonal comparisons are unavoidable and collective decision methods designed to avoid them cannot produce a socially just society.

It is also important to distinguish between procedural (constitutional) rules and methods for deciding particular outcomes. It may be easier to achieve unanimous (Pareto-efficient)

decision procedures than unanimity on outcomes. Individuals may agree unanimously that certain decisions should be made by a simple majority vote, even though they may not vote unanimously for the outcomes that follow. Suppose that a group of people wishes to choose a restaurant. The group may agree unanimously to a less-than-unanimous decision rule such as a majority vote. If the majority can agree on a restaurant, all other members of the group will agree to it. A group can agree unanimously to accept a less-than-unanimous decision and be better off than without the rule (when no decision might be reached). We enter games willingly and accept that there will be winners and losers providing that the game is fair. It is rational and consistent with self-interest to accept a constitutional rule that may produce some outcomes counter to our preferences.

Finally, our analysis of the difficulties associated with making collective choices has some significant implications for the design of government. Most of the difficulties arise from differences in individual values. Groups with similar values can reach agreed collective decisions more easily. Thus, private clubs, locally empowered groups and local government units are often efficient instruments for the delivery of public goods. This reduces the coercion that individuals feel when they are part of a larger public authority which has to satisfy a wide variety of individual preferences.

Summary

- Economic analysis usually assumes that the aim of public policy is to maximise social welfare where welfare is a function of individuals' utilities. However, most public decisions are based on voting. Thus, the question arises as to whether voting methods produce outcomes that reflect individual values. The core issue is again one of aggregation: how to move from individual to social preferences.
- Voting methods include unanimous agreement, qualified and simple majority voting, rank-order voting and point scoring methods.
- Unanimous agreement is possible if public goods are financed by personalised taxes. However, determining individuals' preferences is difficult because individuals may gain by concealing their preferences. Also, unanimous agreement often requires time-consuming decision processes.
- More common collective decision-making methods also have problems. Simple majority voting may produce inconsistent or unstable results. The outcomes may reflect the preferences of the median voter, may be inefficient and may disadvantage minorities. Larger majority requirements may protect minorities but are costly in decision time and biased against change.
- No voting method satisfies quite mild ethical criteria that ensure satisfactory processes and decisions in all situations. All voting methods have some unsatisfactory features, such as coercion, intransitivity or instability, or dependence on irrelevant alternatives, in some circumstances.
- Nevertheless, reasonable collective decision rules can often be found. In some cases, individuals may agree unanimously on a decision process and accept the outcome. In other cases, public goods may be delivered in smaller, more homogeneous groups which reduce the conflicts of values that arise in larger groups.

Questions

1. What are the limitations of unanimous voting as a method of public choice? Would unanimous voting be in conflict with any of Arrow's axioms for collective choice?
2. Is free riding likely to lead to under- or over-statement of demands for public goods? Give reasons for your answer.
3. According to Buchanan and Tullock, an optimal majority is one that minimises the sum of coercion costs and decision-making costs. If this is an optimal majority, why may it still fail to meet one or more of Arrow's axioms for an acceptable collective choice rule?
4. What problems arise with the use of point scoring as a method of collective choice?
5. Given that collective choice methods often fail to reflect individual values, what are the implications for the role and design of government?
6. Under what conditions will simple majority voting produce a determinate ranking? When does the outcome of simple majority voting depend on the order in which alternatives are presented?
7. Suppose that government is considering a proposal to increase expenditure on public education. If voter preferences for government funding on education are inversely proportional to their income, what is likely to determine the level of expenditure?
8. Anne, Bruce and Cathy each value publicly provided child care hours differently. Anne's demand for child care hours per week is $Q^d = 80 - 10P$; Bruce's demand is $Q^d = 150 - 25P$; and Cathy's demand is $Q^d = 200 - 20P$. If the marginal cost of providing an hour of child care is \$20, what is the efficient number of child care hours per week? And what would be the share of the tax burden for each of them under Lindahl pricing?
9. Two families (*A* and *B*) live on a street without lights. The local council will provide street lights if the families are willing to pay for the running costs, which are \$30 per month per street light. The demand of the two families for street lights is estimated to be $P_A = 40 - 2Q_A$ and $P_B = 30 - 2Q_B$, where P is the price the family is willing to pay for a street light per month.
   i. Express the total demand function for street lights.
   ii. How many street lights should be provided?
   iii. If the council charges each family \$15 per street light per month, could unanimity be achieved?
   iv. What would be the share of expenses between family *A* and *B* if unanimity is to be achieved?
10. Should different voting rules be applied to different public decisions such as changes to constitutions?

Further Reading

Arrow, K.J. (1950) 'A difficulty in the concept of social welfare', *Journal of Political Economy*, 58, 324–346.

Bowen, H.R. (1943) 'The interpretation of voting in the allocation of economic resources', *Quarterly Journal of Economics*, 58, 27–48, reprinted in *Readings in Welfare Economics*, K.J. Arrow and T. Scitovsky (eds), 1969, Allen and Unwin, London.

Hindricks, J and Myles, G.D. (2006) *Intermediate Public Economics*, Chapter 10, MIT Press, Cambridge, MA.

Hillman, A.L. (2003) *Public Finance and Public Policy*, Chapter 3, Cambridge University Press, Cambridge.

Mueller, D. (2003) *Public Choice III*, Chapters 4–7, 24, Cambridge University Press, Cambridge, UK.

Sen, A. (1985) 'Social Choice and Justice', *Journal of Economic Literature*, XXIII, 1764–1776.

Chapter

10

Public Choice In Practice

Politicians are the same the whole world over. They promise to build a bridge even when there is no river.

Nikita Krushchev

Politicians and Public Policies ♦ Public Servants and the Supply of Public Services ♦ The Role of Special Interest Groups ♦ General Causes of Government Failures ♦ Predictions of Public Choice Analysis ♦ Policy Implications

In this chapter we discuss how public policies are actually made and whether these processes are likely to produce outcomes that enhance collective welfare. The analysis focuses on the roles of three groups (politicians, public servants and special interest groups) and the environment in which they work. The actions of members of these groups are influenced by social norms, individual ethics and private incentives. On the other hand, the public sector often offers opportunities for acquisition of private benefits or economic rents. And, certainly, some politicians and public servants as well as members of special interest groups may pursue these benefits depending on the opportunities available and the constraints imposed by the political environment. This environment includes the voting processes, the requirement that politicians must get elected, the monopoly nature of much bureaucratic supply and the asymmetric nature of information in the public service.

In the first half of this chapter we examine how politicians, public servants and interest groups may influence public policy. The second half of the chapter discusses general causes of government failures, describes empirical results of public choice analysis and discusses policy implications.

Politicians and Public Policies

In democracies, elected representatives make most important collective decisions. This reduces the cost of making these decisions. But it raises the principal–agent problem. In general, the issue is how can the principal *A* motivate their agent *B* to work in the interest of *A* rather than follow their self-interest, when and where the objectives of the principal and the agent diverge? In a democracy, how can citizen voters (the principals) ensure that their elected representatives (their agents) represent the preferences of voters rather than their own agendas? To answer this question, we need to identify the objectives of politicians, the constraints on their actions and the likely outcomes of the representative process.

The public choice literature usually assumes that an elected representative seeks to maximise their own utility subject to the constraint that he or she must be elected. A

politician's utility is assumed to depend on their personal position, income and power and, in some cases, on achievement of social objectives. However, because these elements of utility are difficult to measure, the literature usually assumes that a politician's main objective is to get elected. As a recent Australian Prime Minister, John Howard, remarked: 'I am in the business of winning elections'[1] (not, it may be noted, in the business of maximising social welfare)! The public choice analysis then examines how election processes affect collective decisions and how freedom from election constraints may allow a politician to pursue their personal agenda rather than represent the preferences of the voters.

There are two difficulties with this approach. One is the pejorative starting assumption that politicians are largely out to further their own interests and, by implication, put little or no value on social ideals. The second difficulty is that outcomes depend on the constraints on politicians posed by the election process. Because election processes and voting rules vary significantly across countries, outcomes may also vary. We consider below some implications of two-party and multi-party models of legislatures. A key issue is whether all votes for politicians are equally important or whether some votes are more important than others.

Two main parties and critical votes

Traditionally, many legislatures have featured two main parties. Examples include the House of Representatives in Australia (Liberal and Labor), the House of Representatives in the United States (Republicans and Democrats) and the House of Commons in the UK (Conservatives and Labour). Notwithstanding the emergence of minority parties in balance of power positions in Australia and the UK in recent years, dominant two-party systems may emerge for several reasons. One is the winner-takes-all nature of elections. To be elected to represent a constituency in the House of Representatives in Australia, a candidate must obtain 50 per cent plus one vote. If this does not happen in the first round of voting, the candidate with least votes is eliminated. This process continues until one candidate achieves the required vote. For the UK House of Commons, the winning representative is the candidate who achieves most votes in the first round of voting. There are no prizes in either case for coming second. Second, national parties require substantial organisation, marketing and resources. Third, the party with most seats usually forms the government. A candidate who seeks power must join a party that can win such a majority.[2]

A key finding of public choice theory is that in a two-party state and under certain conditions, the median voter has the decisive vote. This applies both to the election of representatives and to public policy making. The median voter is the voter whose preferences are in the middle of the distribution of preferences. In voting for representatives, the conditions are that the winning candidate must attract 50 per cent plus one vote, voting is compulsory and voter preferences are single-peaked about a single major issue (such as the amount of government expenditure). Under these conditions, a candidate maximises his or her vote by winning the support of the median voter. In Figure 10.1 a candidate to the right or left of centre will not obtain the required majority. If this holds, this implies that both parties must offer policies that reflect the preferences of the median voter.

There is much casual evidence for the median voter hypothesis. As Persson and Tabellini (2002) point out, many welfare programs are oriented to assist the middle classes. Also, governments tend to favour labour market policies that protect the wages and working conditions of the employed at the expense of unemployed workers who are not represented by the median voter.

---

[1] As reported in *The Sydney Morning Herald*, 1 March 2000.

[2] These conditions do not always result in two main parties. Italy changed from proportional representation to the UK's first-past-the-post system, but still has a multi-party outcome.

Turning to econometric studies, as Mueller (2003) shows, many studies have attempted to test the median voter hypothesis by estimating a cross-section demand equation of the following form:

$$\ln G = a + \alpha \ln t_m + \beta \ln Y_m + \gamma \ln Z + \mu \qquad (10.1)$$

where G is government expenditure (e.g. on schools) in various jurisdictions, t_m and Y_m are the tax price and income of the median voter and Z is a vector of preference parameters, such as number of children in a household. Although many studies find that the coefficients α and β have the expected sign and are statistically significant, similar results are obtained in equations employing mean tax price and income variables. This suggest that the average income voter is significant, but it is hard to distinguish between median and mean income.

In a different kind of study, Bruner and Ross (2010) analysed the voting patterns in two popular referendums on expenditure in public schools in California where both high and low-income households preferred low public expenditure and middle-income households favoured high expenditure. In both cases, a majority of voters favoured low public expenditure. However, in these cases, the voter with the median income was not decisive.

There are indeed several conditions under which the preferences of a median voter may not be decisive. First, people may vote for a representative based on party positions on several issues, not just a single issue. In this case, a candidate may win by attracting votes from swinging voters on several issues. This may include supporting some radical positions rather than one central position. Second, with optional voting, many people not vote, especially when there is little to choose between the parties and voting incurs time and travel costs.[3] Thus, candidates may move away from the centre to attract people to vote. Third, and most importantly, in 'safe-seat' electorates where one party is certain to obtain the required majority, the elected candidate will be the person who wins party selection. In these common conditions the successful party candidate must obtain the median vote of the party selectors. Thus, the views of many elected representatives may be quite different from those of the median citizen voter (as we saw in the same sex plebiscite in Australia in early 2018)! Fourth, and critically, to win the most seats in the country each major party will aim to win the most marginal seats. Therefore, each party targets the votes of swinging voters in marginal seats. This may encompass only a small number of voters in a few seats. Accordingly, the political

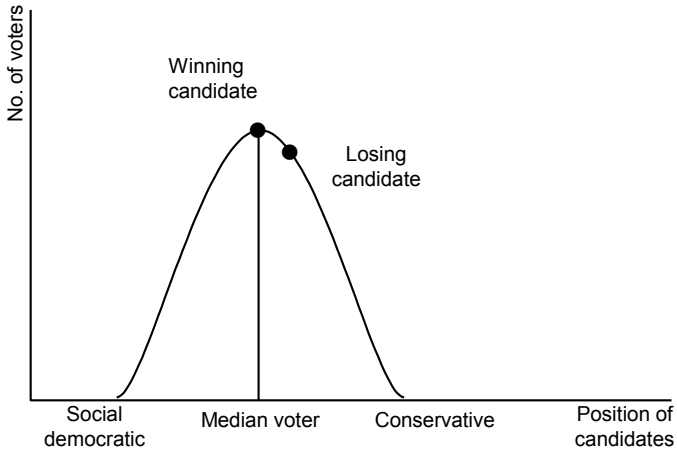


Figure 10.1 Median voter and candidate positions

[3] Voting is generally compulsory in Australia but optional in most other countries.

parties may design policies with a view to winning these votes, which may not reflect median voter preferences across the country.

A potentially more important perversion of the democratic process is the gerrymandering of seats. **Gerrymandering** is the setting up constituency boundaries for political gain. This typically involves establishing a few seats with large majorities of opposition party voters so that they cannot vote in other areas. By doing this, a party can win a majority of seats without a majority of votes. Where governments can determine political boundaries, as in the United States, this can lead to undemocratic outcomes (as has happened). In Australia, where political boundaries are determined by independent electoral bodies this is less likely.

More than two parties

As we noted above, even in traditional two-party legislatures, two dominant parties are not inevitable. In other cases, multi-party legislatures may be common, especially with a proportional representation voting system.[4] Under this method, elected places are allotted in proportion to the total votes cast. Proportional representation increases the number of parties in a legislature, causes candidates to spread out over the issues and results in representation of a greater number of views. However, even in this case some views may not be represented.

Even with multiple parties, outcomes may tend towards the policies preferred by the elected representative who represents a central position. To win government, and to win votes for major resolutions, parliamentary coalitions are required. Successful coalitions tend to be dominated by central parties because they can attract and hold majority support more easily than extreme parties can.

But on other occasions a party may win power or major votes with the support of a small minority group. In Australia, minority parties have often held and exerted the balance of power in the Upper House (the Senate). Following the 2010 election for the Lower House, three independent members with a wide range of disparate policies held the balance of power. In New Zealand, the small New Zealand First Party has held the balance of power on several occasions since 1998, when a form of proportional representation was introduced into the elections. In such cases a minority group, and sometimes only one or two individuals, has an influence over public decisions that is wholly disproportionate to their electoral support.

Another common consequence of coalitions is instability. Because of the constant pressure to negotiate the perks of power, coalitions are liable to break up and reform. This is akin to the unstable cycling over gains that occurs in votes over distributional issues. Mueller (2003) cites international studies that show that the duration of governments is negatively correlated with the number of parties in the parliament and that the probability of a government falling is positively related to the number of parties in the coalition that forms the government.

A politician's freedom of action

An elected representative often has an opportunity to pursue their own personal interests. Their freedom to do so depends on the power of the government to make discretionary decisions. This power depends in turn on the size of the government majority, which makes a government safe from losing confidence votes and facing a new election, and the length of time between elections. Also important is the ability of the voters to monitor government actions, the scrutiny of government actions and the freedom of the media.

The length of time between elections varies. National legislature elections are held within three years in Australia, four years in the United States and five years in the UK. In France, a president holds office for seven years. But even three years allows some freedom between elections. Political parties generally offer the public a package of policies. Voters have limited

[4] Members of the Australian Senate are elected by proportional representation.

choice and mandates are not clear. Government may decide which parts of its policy package it will implement and not infrequently it breaks election promises. After the 1996 election, the Australian Prime Minister, John Howard, distinguished between 'core' election promises that could not be broken and 'non-core' promises that were not real commitments! In contrast, directors of private companies can be sued for failing to provide accurate information and are accountable to shareholders formally once a year. More importantly, they are accountable daily in the capital market. If a company's share price falls, the company and the management are vulnerable to takeover and change. By comparison, the constraints on politicians are weak. Governments may even declare war without the consent of the governed.

This freedom from constraints invites corrupt practices. As Mill (1859) remarked,

> There is little doubt, that if power is granted to a body of men, called representatives, they, like any other men, will use their power, not for the advantage of the community, but for their own advantage if they can.

Recent major books by Rose-Ackerman and Palifka (2015) and Cockroft and Wegener (2017) document extensive corruption in government including in Western democracies.

Corruption arises partly because of the opportunities for politicians to enrich themselves at the public expense. While corruption can be constrained by frequent elections, a free media and rules and procedures for dealing with corruption (including imprisonment), opportunities for private gain in the exercise of public office occur in most countries. Moreover, the need to finance elections creates incentives for delivery of large favours to campaign contributors in any democratic system. When *Time* magazine asked Narasimha Rao (Prime Minister of India 1991–96), 'How significant was the corruption issue in the election?' Rao replied 'Corruption as an issue is there in all countries in all elections'.[5]

Conclusions

In representative democracy, citizens have only indirect influence on public decisions. Often there are only two or three main parties. When political issues can be represented by a one-dimensional right-to-left political spectrum, a party may secure power by winning the support of the centre and there is some evidence that parties focus on voters with preferences in the middle of the distribution of preferences (the median voter). Moreover, to win elections, parties often need to attract only a small proportion of swinging voters in marginal seats from other parties so voters in marginal seats tend to have disproportionate influence over policies. In addition, special interest groups (as discussed below) and politicians holding the balance of power also tend to have disproportionate influence over public policies. Finally, elected representatives may have scope to pursue their own interests unless there are strong legal or administrative constraints on them. Thus, overall there are many reasons why the preferences of voters may be poorly represented in the political process.

Public Servants and the Supply of Public Services

To introduce our discussion of the behaviour of public servants and the impact of their behaviour on the supply of public services, it is useful to distinguish three potential and actual roles of public servants. They are expected to:

1. serve the public interest and that is what they do;
2. serve the elected government and that is what they do;
3. serve the elected government and/or the public interest, but in practice they also pursue their own interests.

[5] *Time*, 19 August 1996.

Traditionally public servants in Australia, and other countries such as France and the UK, were expected to represent the public interest and to provide objective and independent advice to government. Indeed, the *Public Service Act 1999* declared that the Australian Public Service is 'apolitical' and the Act was designed in part to give public servants an obligation to the public as well as to the government.

An alternative view and probably the dominant one today is that public servants should implement the policies of elected representatives. Emy and Hughes (1991) express this view in rather extreme form: 'the public service is fundamentally a political instrument. There is no public interest above and beyond the government of the day'. Politicians are accountable to the electors, whereas public servants are not. In practice, in Australia public services are increasingly politicised. Ministers rather than public servants generally decide senior public service appointments and appoint people to carry out their wishes. Senior public servants no longer have tenure of office and the independence of view and action that this allows. Ministers also rely increasingly on political advisers rather than on expert advice.[6] Under this view of government, public servants are first and foremost government agents and only partially if at all independent servants of the public interest.

There remain some public agencies in which public servants have some autonomy. Although governments often appoint judges, once appointed the judges are usually independent. In Australia, many statutory authorities, such as universities, public trading enterprises and independent commissions such as the Australian Competition and Consumer Commission and the Productivity Commission, are established under legislation that provides for only limited directions from ministers. However, these are mainly executive agencies rather than policy-making bodies.

Be that as it may, how do public servants behave? The safest assumption is that the actions of public servants are influenced by a combination of their contractual obligations to government, their views of the public interest and, in part, by private interest considerations. It is certainly possible to point to examples of corrupt behaviour by public servants in many countries. As discussed below, some may pursue economic rents in the public sector or their own advancement and this may affect how public services are supplied. But I would have to say that from quite extensive experience, in state and local government especially, most public servants are conscientious and have high integrity, though they are often conflicted between carrying out political demands and their personal views of the public interest,

A public choice model of public supply

To model the impact of public servants on the supply of public services, assumptions are required about the objective of public servants and the constraints under which they act. In a classic article, Niskanen (1971) assumed that public servants have substantive private interests and are motivated by the '3 Ps'—pay, power and prestige. Further, in order to achieve these 3 Ps the public servant must acquire bureaucratic power. In particular, the 3 Ps are generally correlated with the size of the budget controlled and the number of employees managed. Therefore, a public servant has a strong incentive to maximise his or her budget and output. This increases all the key elements in the public servant's utility function, except perhaps a desire for a quiet life. Many public finance explanations of the supply of public services draw in some way on these assumptions.

Of course, these assumptions, like those in any model, are a simplification. As we have noted, many public servants hold altruistic social objectives. However, at the heart of the model is the insight from principal–agent theory that principals and agents may have different objectives. In this case, the principal is the elected government and the agent is the public

[6] For a detailed analysis of similar trends in the United Kingdom see Foster (2005).

servant. Elected representatives contract public servants to carry out certain functions. However, public servants may have different objectives, different information and some freedom of action and may not always act as the government wishes. It follows that there are many potential causes of inefficient output.

What are the constraints on public servants? Because they spend other people's money, they have no personal incentive to economise public expenditures unless constrained to do so. However, public sector budget constraints are soft ones. A hard constraint exists where a business must cease operating when it has run out of funds. A soft budget constraint is a constraint that may be breached without serious consequences. In the public sector, funds can be supplemented from consolidated revenue. Thus, the constraint on expenditure must be some form of government control. To understand how government controls public expenditure, we need to understand the environment in which government operates.

This environment has three main characteristics. First, most government departments are monopolies. Ministers must obtain most, if not all, services from their department. There is usually little direct competition within government or from outside. Government departments often fight fiercely for funds and territory but rarely compete directly in provision of services. This avoids wasteful duplication. But, without competition, there is little incentive to achieve efficiency and only weak measures of it. Second, there are rarely ready measures of the value of non-marketed public services. The scale of government activities, such as police or fire services, may be known but their value hard to measure. In this world, more inputs often count as more outputs. Third, information is often asymmetric, especially in large departments. Senior public servants generally know what the government wants. But the government has little idea of the real costs (or benefits) of the services. Ministers obtain most of their information from public servants. But public servants have no incentive to reveal the real costs to ministers, while they often have an incentive to exaggerate the benefits.

Economic models of government output based on these characteristics are illustrated in Figure 10.2. In this figure, the quantity of services is shown on the horizontal axis and their total value and cost on the vertical axis. Total benefit rises with output, but at a declining marginal rate. Cost is assumed here to rise linearly with output, implying a constant marginal cost of supply. The efficient level of output, where social net benefit is maximised, is Q_E. At this point marginal benefit equals marginal cost (the slopes of the total benefit and cost curves are equal). Higher output is inefficient because marginal cost exceeds marginal benefit.

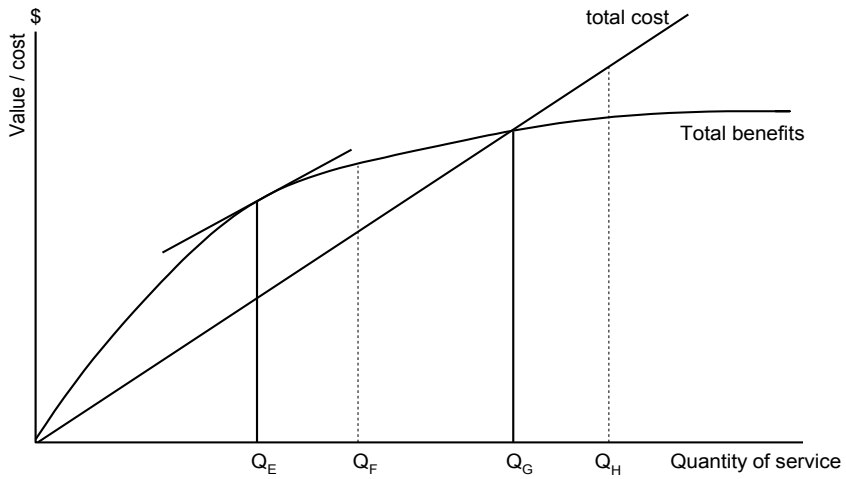


Figure 10.2 Non-marketed output in a public service model

Actual output depends on the behaviour of public servants and ministers. A common prediction is that output will rise to Q_G, which is the point where the total benefit of output equals total cost. Public servants are assumed to seek the highest level of output for which they can obtain funds. On the other hand, ministers are expected to be able to understand approximately the total benefit and cost of a program and to be unwilling to fund expansion beyond Q_G. Only public servants can measure the marginal cost and benefit of a program and they conceal this information from elected representatives. However, by controlling the flow of information to ministers, public servants can expand output from Q_E to Q_G.

Other behavioural assumptions and levels of output are of course possible. Ministers may not know even the total benefit and cost of a program. Public servants may exaggerate the benefits of programs; for example, teachers have an incentive to emphasise the benefits of smaller class sizes. If ministers believe that the benefit curve is higher than shown in Figure 10.2, the equilibrium quantity would move to the right from Q_G towards Q_H. On the other hand, public servants may favour a higher cost structure than that shown in Figure 10.2. This could reflect inefficient production methods, rent seeking (salaries in excess of opportunity costs) or a simple preference for a quiet life (the traditional perk of a monopoly). In any of these cases, the cost curve would shift upward and the equilibrium output, where total benefit equalled total cost, would fall below Q_G.

In this model, the actual supply of public services depends on the outcome of bilateral monopoly negotiations: the government is sole purchaser of a department's services and the department is sole supplier of the services. The outcome depends on negotiations between the two. Also, other departments (notably including the central Treasury agency) may support or oppose the department supplying the service. As in most bargaining models, various predictions of output are possible.

In summary, this model of public service output (often described as the Niskanen model) suggests that government is likely to oversupply public services, to supply the kinds of goods favoured by public servants and to adopt relatively high-cost methods of production. The classic TV series *Yes Minister* and numerous anecdotes suggest that the model contains much reality. Khursheed and Borcherding (1998) and Mueller (2003) provide useful summaries of empirical tests of these predicted outcomes. As shown later in Chapters 16 and 18, there is a fair amount of evidence that government supplies services at a higher cost than does the private sector. In the absence of user charges, it is harder to test whether public services are oversupplied or whether the wrong kinds of services are supplied. Empirical studies have produced a variety of results. For example, Staaf (1977) found that salaries in the US public school system were linked to education budgets and the size of school districts (supporting the Niskanen model). And it may be remarked that there appears to be a strong correlation between the growth of student numbers in Australian universities and the size of vice-chancellor and other managerial salaries. On the other hand, Johnson and Libecap (1989) found no evidence of a general relationship between agency growth and bureaucratic salaries.

The Role of Special Interest Groups

Special interest groups play an important role in most economies. A special interest group is any organised group of people which works to further the common interests of its members. Interest groups exist in many forms. Employer and employee groups were traditionally powerful interest groups. Examples include trade associations (e.g. for the coal, motor vehicle and pharmaceutical industries), farming associations, professional associations (such as for doctors and lawyers) and trade unions. Sometimes a company is sufficiently powerful to constitute a significant interest group on its own. In recent years, other influential interest groups have emerged, including consumer and environmental groups, and welfare associations to promote child care and the interests of the elderly.

Most interest groups aim to influence public policy. In some cases, the aims are idealistic such as protection of wildlife or provision of benefits to disadvantaged groups in the community. However, many interest groups seek to increase the income of their members, often at the expense of other sections of the community. They may do so by seeking government regulation of the market or by direct transfer of income to their group.

Lobbying the government to obtain higher than normal returns through regulations is known as rent seeking. Regulations may control entry into the industry, output, prices or advertising, competition from imports and so on. They may also require firms to employ only members of unions or workers with particular qualifications, for example locally obtained qualifications. Interest groups (employers and employees) aim to obtain such regulations to gain economic rent by reducing competition and increasing the market power of their members. Rent seeking typically has two economic costs: it leads to restrictions on output and it uses resources without providing any output.[7] Rent seeking may also result in inequitable distributions.

Special interest groups often influence public decisions in a way that is disproportionate to the size of their membership because of their ability to mobilise financial support for political parties. Political parties require very large funds for both their ongoing organisation and elections. Twenty years ago, *The Economist* reported that annual spending in political elections in the United States exceeded $3 billion. Election to a seat in the Upper House (the Senate) cost an average of $5 million.[8] Under the Citizens United (Supreme Court) decision in 2012 in the United States, political spending is protected speech under the First Amendment of the Constitution and government cannot restrict contributions to support candidates at elections.

Interest groups also derive influence from their ability to deliver votes. Some interest groups such as unions, environmental and seniors' groups can deliver large number of votes. The American Association of Retired Persons has 40 million members. When voting is optional, members of interest groups that are offered preferential policies have more incentive to vote than other citizens. A related factor is the ability of some interest groups, such as large firms and ethnic groups, to deliver a concentration of votes that may influence the outcomes of particular seats. In these cases, the votes of protected workers are likely to have more influence on policies than votes of consumers and taxpayers who are more widely dispersed.

Evidently special interest groups can affect both the quantity and price of goods. Industry protection for primary producers, manufacturers and professional services is common in many countries. As Mueller (2003) shows, protection tends to be greatest in concentrated industries and in labour-intensive industries which can establish effective and powerful interest groups. Another interest group that has been effective in many countries is the elderly who have obtained significantly increased pensions and health care benefits in recent years (Persson and Tabellini, 2002). Both traditionally, since it was established in 1871, and recently, the National Rifle Association of America has had extraordinary, and literally deathly, influence in promoting and retaining the gun industry in the United States.

General Causes of Government Failures

We have described some of the resource misallocations that are likely to arise from the political process, the public supply of goods and interest groups. However, this discussion, which is typical of the public choice literature, has not provided a systematic analysis of why government may fail to allocate resources efficiently similar to the analysis of market failure in Chapter 4. In this section we discuss some systemic features of the supply of non-market goods that are likely to result in a misallocation of resources—due essentially to the lack of a

[7] For further discussion of rent seeking costs, see Chapter 14.
[8] *The Economist*, 8–14 February 1997.

market. Some points have been mentioned above and other points are taken up in later chapters, especially Chapter 16, which discusses the supply of non-market goods.[9]

A key feature of non-market goods and a major cause of resource misallocations is the absence of property rights. In the absence of property rights over public revenue or expenditure, the whole government budget becomes effectively a common property resource. It presents an opportunity for rent seeking by those who can exert the greatest political strength. Most citizens want a share of the rent. But, because the rent is a common property resource, individuals have more incentive to exploit it than to conserve it. A major related problem with non-market goods is the absence of prices. Prices convey critical information about the value of goods. Government has to estimate the value of non-market goods without this information. When outputs are measured by inputs (as government output is measured in the national accounts), more inputs may be regarded as more output regardless of whether this is the case.

In the absence of prices, the demand for public expenditure is 'decoupled' from the cost. In markets, the benefits of consumption are coupled to expenditure. In the public sector, individuals can obtain benefits at no personal cost. This creates a 'free riding' environment in which individuals can demand that services be supplied without paying for them. In effect they are gaining rents. When a service is free, there is almost always excess demand for it and pressure for more supply. Similarly, with regulations, individuals or groups can lobby for regulations that raise their incomes without paying the costs associated with the regulations. On the other hand, the absence of a pricing mechanism creates a need to ration supply, which provides public servants with opportunities for corrupt behaviour. The decoupling of benefit and cost is most extreme with income redistribution programs, where one group of people can lobby for programs to be paid for by other people. Moreover, because of the high discount rate of politicians facing elections, politicians may satisfy the demand for services or other benefits now regardless of future costs that tend to be discounted.

On the supply side there are also major failures due to the lack of markets. First, single-source production is often inefficient because it is not subject to competition. In the absence of competition, evaluation of the cost and quality of the output is difficult. Second, when the revenues that fund a service are decoupled from the costs of supplying it, there is no clear benchmark for efficiency and the scope for misallocation of resources is greatly increased. Third, in the absence of direct financial constraints on the supply of a service, there is no termination mechanism.

A fundamental challenge is the wide variety of views of fairness and justice in society. Fairness may be viewed as equal total income or as equal wages per hour of work, as equality of opportunity or as fair rewards for work, as assistance for low-income earners or more generally as assistance for disadvantaged groups. Equity may mean horizontal equity (treating like individuals in the same way) or vertical equity (individuals should pay tax based on their capacity to pay tax) and so on. Citizens (voters) may well have differing views about fairness for these are ethical views on which individuals may differ. However, the variety of views about equity creates considerable potential for inconsistent government actions.

Predictions of Public Choice Analysis

Our analysis has indicated several general causes of government failures and inefficient resource allocation. Does public choice analysis provide specific predictions about government behaviour and economic outcomes? In this section we examine some effect of politics and bureaucracy on short-run macroeconomic policy, the size of government, the allocation and distribution of resources, and methods of production.

---

[9] This section draws on Wolf (1988), especially Chapters 3 and 4.

Elections, macroeconomic policy and the business cycle

A common public choice hypothesis is that elections influence the business cycle. The presumption is that, to win elections, government presents popular (high expenditure/low tax) budgets before elections. Fiscal and monetary policies are designed to maximise economic growth and minimise unemployment before elections. If there is spare productive capacity, increases in prices lag increases in output and any large inflationary effects are experienced after the election. Governments therefore increase spending, including transfer payments, and government deficits before elections, and increase taxes after elections. These policies create a business cycle.

This political business cycle hypothesis is based on one or more strong assumptions: that voters do not recall the overall performance of the government in office; that voters are not concerned about short-term budget deficits; or that voters are not concerned about possible inflationary consequences of short-run increases in aggregate demand.

Mueller (2003) reviewed empirical studies of the political business cycle hypothesis. Many studies cited support the political business cycle hypothesis, with increasing expenditures, lower taxes and increased deficits just before elections found in the United States, Canada and Japan, and more generally in OECD countries. See, for example, Bhattacharya and Wassmer (1995) and Reid (1998). Crosby *et al.* (1997) also found that government expenditure rose before elections in Australia. A casual review suggests that Australian government expenditure also increased above trend before each of the elections in 2001, 2004, 2007 and 2010.

Another prediction from public choice theory is that macroeconomic policies generally reflect the ideological preferences of parties. Right-of-centre parties are expected to favour price stability over reducing unemployment. Left-leaning parties to have reverse priorities. This prediction assumes that parties can follow their own preferences and promote the interests of their main constituencies rather than pursue the support of the median voter (or the marginal swinging voter), in which case their policies would converge. Support for this prediction is strong in the United States. Between 1952 and 1988 Republican administrations reduced the inflation rates and increased the unemployment rates that existed when they took office, whereas Democratic administrations increased inflation rates and reduced unemployment rates. Similar outcomes have been found in Europe (Mueller, 2003). However, Crosby *et al.* (1997) found no evidence of such partisan influences over the macroeconomy in Australia. And in recent times with close elections there has been considerable policy convergence especially in the 2010 election.

Total government expenditure

As we saw in Chapter 2, total public expenditure in Australia rose substantially in relation to GDP from 1960 to 1985 but has been broadly constant around 35% of GDP since then. We also discussed there how economic factors influenced these outcomes. We now consider the role of political factors. These factors include (1) the demands of voters and special interest groups for more free or subsidised public services, such as health services, and transfer payments and (2) the role of representatives and public servants in supplying more goods or income.

There are several demand-side political explanations of the growth of public expenditure. One is based on the political importance of the median voter. When income distribution is skewed with top income earners receiving very high incomes, the median income is less than the mean and the median voter prefers an increase in tax rates because their tax contribution will be less than the value of public services he or she receives. Providing that income redistribution does not create too many disincentives, politicians can win the support of the majority by promising to redistribute income to less well-off households. Also, in many countries low-income and other disadvantaged groups have become more powerful politically

as the vote was extended to more groups and eventually to all adult citizens. Increasing inequality in market incomes in many countries (see Chapters 20 and 34) has also increased demands for income transfers.

Another possible explanation for rising public expenditure is that special interest groups tend to seek increased public expenditure on goods and transfers (see Olsen, 1982). These groups include (1) producer groups of all kinds (farmers, miners, importers, artists and so on) who seek subsidies and (2) consumer groups who seek more public funds for education, health, child care and aged care and so on. These interest groups often have more effective power than the mass of taxpayers who may favour lower taxation, although we note below some exceptions in the use of democratic power in Switzerland and the United States However, the demand for more services and benefit transfers is not limited to interest groups. It is a widespread phenomenon and a natural corollary of the free-rider characteristic of many public services, such as local hospital, educational or transport services, that people lobby for services funded out of consolidated revenue and provided to them at no cost.

On the supply side, politicians gain by providing projects to their constituents funded from consolidated revenue. And, as we have seen, public servants may gain higher incomes from managing larger budgets. Chapters 16 and 18 provide some support for the view that in-house production of services increases project costs and hence government expenditure.

Another popular supply-side explanation for the growth of public expenditure is the **ratchet or displacement effect** generally attributed to Peacock and Wiseman (1961). Peacock and Wiseman argued that governments like to spend public money but are constrained by their perception of what the public will bear. In crises such as war the public will bear higher taxes. The expansion of government-funded services increases the tolerance of the public to higher levels of taxation. After the crisis, public expenditure continues to displace private expenditure, as it is racheted up to a permanently higher level. There is international and Australian evidence for this hypothesis, though the evidence is now mostly historic.

Fiscal illusion is another possible explanation for the rise in government expenditure. Fiscal illusion occurs if voters do not understand the impacts of government fiscal actions. For example, there may be fiscal illusion about the results of pump-priming expenditure to promote economic growth before elections. These extra goods and transfers may be paid for after the election by inflation or by lower government spending. More generally, whatever the cause(s) of inflation, unless tax rates are indexed to allow for price changes, inflation increases taxes as taxpayers move into higher tax brackets and in effect experience real tax increases. Australia, like most other countries, does not index tax rates.

Evidence. Drawing on Mueller (2003), various studies support the view that political factors influence the size of government spending. First there is evidence that public expenditure falls when voters have more control over the budget. In a study of public expenditure in 110 cantons/municipalities in Switzerland, Pommerehne and Schneider (1982) concluded that public expenditure was 28 per cent higher in the 62 cantons that were governed by elected representatives and had no direct voting on any major issues than in the 48 cantons which operated under direct democracy. Funk and Gathmann (2011) found likewise that direct democracy reduced canton spending, though by less than had been previously estimated. Public expenditure is also much lower in New Hampshire, where there are constitutional constraints on public spending, than in neighbouring states in the United States. Perhaps most famously, the popular plebiscite vote in California, Proposition 13, under which property taxes were capped and halved represents an example of the power of the mass of people rebelling against higher taxes and presumably, by implication, higher public expenditure.

Second, several studies have shown that interest groups affect the size of government (e.g. Mueller and Murrell, 1985 and 1986). Based on a cross-sectional study of 12 OECD

countries, Lybeck (1986) reported that government size was significantly influenced by the degree of unionisation, the number of public employees and by unemployment. He also found that the size of government in relation to GDP in Sweden varied over time with the fraction of employees who were members of interest groups. Plotnick (1986) reported that income support in the states of the United States depended on the size of the pro-welfare interest groups.

Third, public expenditure is generally a smaller proportion of GDP in federal states than in unitary states. The inference is that voters have more power in a federated country where there is also more competition between states. Blankart (2000) found that public expenditure in Germany rose as it became more centralised.

In Australia, Hackl *et al.* (1993) reported that, in addition to conventional economic variables such as income and population, several political factors explained changes in government expenditure. The political factors included interest group effects, the bureaucracy, the political complexion of government, the size of the budget deficit, tax share and price inflation. Moreover, the ratchet effect of war has had a major and sustained impact on the level of government expenditure, notably after the Second World War during which the Commonwealth obtained income tax powers.

The allocation of public expenditure and regulations

Much public choice literature suggests that both policy and the allocation of public expenditure can be predicted from models of the political process. It predicts, for example, that governments will deliver regulations and goods that are attractive to interest groups (especially to narrowly concentrated groups), to voters in marginal seats and to swinging voters especially, who can provide party finance or critical votes. Because of the strategic voting position of the median voter, government will provide programs of special benefit to the perceived median voter. More generally, government will redistribute income towards middle-income households.

In Australia, governments have regularly provided programs for special interest groups. Examples are tariff protection for the textile and clothing and motor vehicle industries, accelerated depreciation allowances for primary producers which reduces their tax liability, and licensing requirements for professional groups such as doctors, pharmacists and lawyers. Before elections, Commonwealth and state governments invariably announce major expenditure projects in marginal seats. At all times, the Commonwealth government provides substantial support for middle-income earners in health care, education, home ownership, child care subsidies and so on. Although overall government tax and expenditure programs reduce income inequality (see Chapter 20), there is considerable support for middle-income groups and marginal seats.

Worldwide, probably the most generous industry support goes to agricultural interests. Half the whole European Union budget goes to agricultural producers. This is striking testimony to their political power. Lopez and Pagoulatos (1994) show that industry rents from tariff protection in the United States are a significant positive function of the size of political donations. Goldberg and Maggi (1999) show that non-tariff protection is also a positive function of an industry's campaign contributions.

Methods of production

The public choice literature predicts that both politicians and bureaucrats will prefer public production of government output. Government ministers may favour public ownership of the means of production because financial transfers can be made to politically favoured groups in hidden ways, for example by cross-subsidies. Bureaucrats may prefer public production because it expands their area of responsibility, which in turn enhances the 3 Ps (pay, power

and prestige). The fact that public production is often preferred to private despite higher costs (see Chapters 16 and 18) provides some support for these predictions in the literature.

Corruption in government

The public choice analysis of government also predicts that some corruption is likely. The monopoly power that politicians and bureaucrats exercise over the provision of many services enables them, unless accountable, to extract a monopoly rent from supplying the service. Corruption is an extreme form of rent seeking and can result in an extreme misallocation of resources. In a survey of the literature, Gruber (2016, pp. 266-269)) shows that corruption is associated positively with electoral systems such as proportional representation where voters elect slates of representatives rather than individuals (thus reducing accountability) and the degree of red tape for business operations (which increases the benefits of bribery to business), and negatively with the wages of public servants (who have less to lose from being punished for corrupt practices). Rose-Ackerman (1999), Rose-Ackerman and Palifka (2016) and Cockroft and Wegener (2017) document large amounts of corruption in many countries.[10] Australian politicians are not immune from corruption. Over the last 20 years several Australian politicians have been jailed for corrupt practices.[11]

Policy Implications

Evidently, for various reasons government may misallocate resources and fail to provide the services that citizens want. We discuss below how government decisions can be made more responsive to citizen preferences. In Chapter 16 we discuss the related issue of how public services can be supplied efficiently.

Direct democratic measures. The most direct ways to increase citizen control over public policies and services are by constitutional controls or by various direct voting processes. Some writers, for example Milton Friedman, have argued for constitutional rules that would allow citizens to trade across national borders without government restrictions or that would enshrine limits on government expenditure or budget deficits. However, moves along these lines to date have been limited and had mixed success. The US Congress has made several attempts to reduce the US budget deficit by legislation that would bind both the President and Congress. However, Rosen and Gayer (2014) observe that Congress has been skilful at circumventing the budget caps. Recently Congress classified expenditure over US$100 billion spent on the wars in Afghanistan and Iraq as emergencies and thus not to count under the cap. Nearly all states in the United States have also introduced rules in their constitutions that forbid operating deficits. This requires clear distinctions between operating and capital revenues and expenditures, which is not always easy. There is also the issue of enforcement. How does a legislature deal with a situation where an operating budget surplus is planned but not delivered? It is hard to test whether the statutory caps have been effective because the capped outcomes may reflect more fiscally conservative legislators rather than the statutes themselves. However, Auerbach (2008) found that the fiscal rules did have some impact on deficit control.

Direct voting procedures include voting for major representative positions, recall elections, referendums and voter initiatives.[12] A recall election is a special election initiated by citizens

---

[10] See also the regular reports of Transparency International, a non-profit organisation established to promote clean government.

[11] Jailed state government politicians included Burke and Parker (an ex-premier and ex-deputy premier respectively in Western Australia), 'Buckets' Jackson, who had been Minister for Corrective Services in New South Wales, "Lunch-a-lot" Macdonald and mate Minister Eddie Obeid (also NSW) and Nuttall in Queensland (for 12 years).

[12] The US information in this paragraph is drawn from Gruber (2016).

with the aim of replacing a sitting elected representative. In the United States, 18 states allow the recall of state officials and 36 states allow the recall of local officials. A referendum allows citizens to vote on state laws. All US states allow legislatures to invite a popular ballot and 24 states allow citizens to collect enough signatures to require the legislature to take a popular ballot. Twenty-four states also allow voter initiatives that allow citizens to place their own legislation on the ballot for voters to accept. Many cantons in Switzerland also have a tradition of referendums. There is no such tradition in Australia. To be effective, a plebiscite must be binding on government, not simply advisory. There is no doubt that these various means of direct democracy have had a significant political impact in the United States, where among other results it led to the election of Arnold Schwarzenegger as Governor of California in 2003, and in Switzerland.

However, whether direct voting improves public resource allocation is an open question. According to *The Economist* (19 March 2011) California is 'now widely studied as an example of what to avoid'. Seventy-five per cent of the state's budget is outside of the government's control, the current budget deficit is US$25 billion and 'the roads and colleges are crumbling'. These are challenging observations. A fuller and wider review of the impacts of direct voting might provide a more balanced conclusion about the merits or otherwise of direct voting and more especially when it is beneficial and when not.

Indirect methods of increasing citizen control. There are many indirect ways to increase voter control over policies. One strategy would be to adopt political processes that encourage more political parties to emerge. This would allow more representation of voter preferences and introduce more checks and balances into the political process, as occurs in the Senate in Australia. On the other hand, an increase in the number of parties may reduce the stability of government.

Probably a more effective way to improve voter representation is by decentralising government and encouraging competition between governments. Several studies (for example Zax, 1989; Oates, 1989) have shown that public spending falls with greater fragmentation of political jurisdictions. A federal state encourages competition between the constituent states in taxation and the provision of services. Citizens can choose their preferred public package by moving to another jurisdiction (voting 'with their feet'). Competition provides an incentive to state legislators and bureaucrats to discipline their spending. Providing public services at the most local level of government compatible with efficient production helps to ensure that voter preferences are recognised in the provision of services.

As we have seen, the cost of elections is another major issue with implications for the power of special interests over resource allocation. The standard policy responses to this include increased public financing of election campaigns, greater disclosure of contributions, cost controls on elections and more enforcement of electoral rules. However, there are apparently no such limits in the United States.

Another electoral issue is the transparency of information provided by government to voters. Under the Australian *Charter of Budget Honesty*, government must provide forward spending and revenue estimates with each budget and a mid-year outlook. After an election is called, the Secretaries of Treasury and Finance must sign off on revised budget estimates and issue an economic and fiscal outlook. These requirements make the existing government more transparent and accountable. New governments no longer have the excuse that they 'did not know how bad things really were'. The National Commission of Audit (1996) proposed that the Secretaries of Treasury and Finance also provide independent economic reports at budget time and half yearly, but given the nature of the government/public servant relationship this is impractical and has not been implemented.

Tax hypothecation, the process whereby tax revenues are raised for specific projects, is another way to increase the transparency of public decisions. However, treasuries tend to argue that all tax revenues should go into consolidated revenue and then be allocated efficiently to priority projects (see Chapter 28).

Finally, two other strategies should be mentioned. One involves greater use of statutory agencies and officials. These are agencies and officials that are charged by statute to carry out various functions with given objectives, including acting in the public interest, and who once appointed are subject to limited ministerial control. This would reduce day-to-day political decision making in these designated areas.

A more general strategy for increasing voter control over public programs is to provide consumers with choices by subsidising use of services rather than supply. For example, government can provide parents of school children with educational vouchers that can be used to purchase school services, including services from private suppliers. People requiring health care can be given the choice of health care provider. This gives citizens more power over the use of resources. However, politicians and public servants often resist this strategy because they perceive that it dilutes their control.

In summary, there are many government failures and consequential resource misallocations —and no single solution. However, there are many possible strategies that, taken together, can increase voter control and mitigate the extent of the resource misallocation.

Summary

- In representative democracy, citizens have limited influence on collective decisions. Often there are only two or three main parties. Elected representatives rarely represent a broad cross-section of the voters.
- Median voters and swinging voters in marginal seats and special interest groups tend to have disproportionate influence over policies. Also, elected representatives often have significant scope to pursue their own interests.
- Public servants also have an interest in the supply of government goods. In the absence of competition, this is likely to lead to an oversupply of government goods, a supply of goods preferred by public servants and public means of production.
- Special interest groups often have a disproportionate influence on policies because of their ability to deliver funds and votes to politicians.
- Underlying these political issues is the systemic nature of non-market goods. Public income can be viewed as a common property resource. Individuals can obtain benefits at no personal cost.
- When the price of a service is zero, demand exceeds supply and there is pressure for an inefficient increase in supply.
- Government failures in supply also include the difficulty of measuring output, single-source production, weak measures of efficiency and soft budget constraints on the supply of a service.
- Empirical studies suggest that election politics influence macroeconomic policy and possibly the business cycle. Also, political factors strongly influence the size and composition of government expenditure. Public choice theory also predicts that corruption in government is likely.
- Government decisions can be brought under greater citizen control in various ways. Direct methods involve more citizen control over legislation. Indirect methods include decentralisation of public services, more limits to funding of elections, greater transparency of government actions, increasing the number and responsibility of statutory agencies and offices and providing more choice of services to users of government services.

Questions

1. Should public servants serve elected politicians or the public interest? Or is this an immaterial question because they will serve their own interests in any case?
2. How do public servants influence the supply of public services?
3. What are the main systemic causes of government failure?
4. Does the financing of elections give special interest groups too much influence? If so, what might be done to curb this influence?
5. How can the hypothesis that government's macroeconomic policies are influenced by political (re-election) considerations be tested? What is the evidence, if any, that electorally induced macroeconomic policies influence the business cycle?
6. What is the evidence, if any, that governments oversupply public services in aggregate or that they supply the wrong kinds of public services?
7. What evidence is there that direct voting affects government revenue or expenditure? What conclusions, if any can be draw about the effect of direct voting on social welfare?
8. Draw on public choice theory to explain why most governments fail to index tax rates.
9. Assume that a government department consistently provides its minister with an underestimate of the total costs of a program. Using the Niskanen model of public output, show how this is likely to lead to a level of output in excess of the efficient level.
10. Assume that government supplies a publicly financed good with a total cost function $C = 40 + 10Q$. The total community benefit is given by $B = 20Q - 0.5Q^2$. What is
    i. the maximum level of output that can be supplied without a net loss to the community?
    ii. the efficient level of output?
11. European Union (EU) countries frequently violate the EU rule that deficits be kept below 3 per cent of GDP despite the formal sanctions of large fines (which are not actually imposed). Is it feasible and desirable to have legislative rules that govern the size of government deficits?
12. Why are economic rents a feature of government? And why do economic rents encourage corruption?

Further Reading

Funk, P. and Gathmann, C. (2011) 'Does direct democracy reduce the size of government? New evidence from historical data, 1890-2000', *Economic Journal*, 121, 1252-1280.

Gruber, J. (2016) *Public Finance and Public Policy*, Chapter 9, MacMillan Education.

Khursheed, A.F. and Borcherding, T.E. (1998) 'Organising government supply: the role of bureaucracy', Chapter 2 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Mueller, D. (2003*) Public Choice III*, Chapters 10 to 22, Cambridge University Press, Cambridge.

Persson, T. and Tabellini, G. (2002) 'Political Economy and Public Finance', pp. 1551–1658 in *Handbook of Public Economics*, Vol. 3, A. Auerbach and M. Feldstein (eds), Elsevier, Amsterdam.

Rose-Ackerman, S. and Palifka, B.J. (2016) *Corruption and Government: Causes, Consequences and Reform*, Cambridge University Press.

Wolf, C. (1988) *Markets or Government*, Chapters 3 and 4, MIT Press, Cambridge, MA.

Part 5

Building Economic Foundations

Chapter

11

Public Goods

Government is a contrivance of human wisdom to provide for human wants.
Men have a right that these wants should be provided by this wisdom.

Edmund Burke, *Reflections on the Revolution in France*

Framework for Analysis of Public Goods ♦ Efficient Supply of Public Goods ♦ Revealed Preference Methods of Valuation ♦ Stated Preference Methods of Valuation ♦ Valuation Methods: Applications and Conclusions

A non-excludable good
People cannot be excluded from consuming the good

A non-rival good
One person's consumption of the good does not reduce the availability of the good to others

Public goods are the basis of the state. Public goods include all the services that are required for a state to function: defence, law and order, rules of property and contracts, social and economic infrastructure, basic education and health services, waste collection services and many other services that provide collective as well as private benefits that markets for one reason or another are unlikely to provide.

As discussed in Chapter 4, public goods are technically goods that are non-excludable or non-rival in consumption. Some public goods have both attributes. A good is **non-excludable** if, when supplied to one person, another person cannot be excluded from consuming it or, more broadly, benefiting from it. A good is **non-rival** when one person's consumption of the good does not diminish the supply of the good to others. These characteristics create market failures. Non-excludable goods are under-supplied. Non-rival goods are under-consumed.

Given the fundamental role of public goods in the economy, the provision of public goods is discussed in many parts of this book. In this chapter we start by setting out how the range of issues relating to the provision of public goods is dealt with in this text. We then focus on two prime issues: how to determine efficient quantities of non-excludable and non-rival goods and the related issue of methods for estimating the value of public goods.

Framework for Analysis of Public Goods

The major tasks in the analysis of public goods along with their places in this text are outlined in Table 11.1. The starting point is the definition of public goods and an analysis of market failures, which generates a potential role for government. Given that markets do not produce efficient quantities of public goods, we need set out how efficient quantities can be determined. As usual, the efficient quantity for any good depends on the social benefit and cost of an extra unit of the good.

Table 11.1 Analysis of public goods

Task	*Main features*	*Main text references*
1. Define public goods	Non-excludability or non-rivalness or both	Chapter 4
2. Identify market failures	Public goods are not supplied or supplied inefficiently	Chapters 4, 11
3. Set social objectives	Definitions of efficiency and optimal quantities (including considerations of equity)	Chapter 11
4. Demand estimation	Principles and issues in estimation of preferences	Chapters 6, 11
	Main methods of estimation of preferences	Chapter 11
5. Cost estimation	Principles of cost estimation	Chapter 8
	Methods for estimating costs of production	Chapter 16
6. Evaluation	Usually an application of cost–benefit analysis	Chapter 8
	Education, transport and health applications	Chapters 12, 19, 24
7. Assess supply options	Various ways to provide public goods	Chapter 16
	Pricing, financing and ownership issues	Chapters 17,18
	Central or local supply of public goods	Chapter 33

However, in the absence of a market there is no readily observable demand or willingness to pay for public goods and therefore no clear measure of benefits. When a good is non-excludable, individuals have an incentive to free ride and to misstate their preferences for the good. Accordingly, economists have done a large amount of work on ways to elicit the true values that people attach to non-market goods. These methods fall into two main categories, revealed and stated preferences, which we discuss below. These methods are critical for establishing the value all forms of public goods, including natural environmental goods.

Various issues also arise in estimating the costs of supplying goods. These issues include determining the inputs needed to produce varying levels of output, the allocation of joint costs, the estimation of fixed, variable and average costs and the allocation of capital expenditure over time. These issues are taken up in Chapter 16.

Once benefits and costs are estimated, an overall method of evaluation is required. Cost–benefit analysis (CBA, Chapter 8) provides a general evaluation method for determining an efficient output of public goods. Chapters 12, 19 and 24 describe how CBA can be applied in the education, transport and health sectors respectively. However, as we saw in Chapters 7 and 8, equity objectives may need separate assessment.

Given that an appropriate quantity of a public good is determined, the best way to supply the good must be determined. Public goods may be supplied by:

- public production of publicly financed goods;
- contracting private firms to supply publicly financed goods;
- subsidising private firms to increase their output of specified goods; and
- subsidising consumers of specified goods.

The best way to supply public goods may vary with the degree of market failure, the nature of the good (whether it is a pure public good or a mixed public–private good), the amount of competition among potential suppliers and the competence of government. Much depends on the excludability attributes of the public good. Therefore, there is no single prescription of the provision of goods with public good attributes. In any case, the possibility of government or regulatory failure should be weighed against market failures. General options for the supply of various kinds of public goods are discussed in Chapters 16 and 18 as well as specific issues in education and health (Chapters 12 and 24).

Another important characteristic of many public goods is their spatial dimension. Even the provision of internal security services will vary across the country. Services such as libraries, swimming pools and street lighting are local public goods. Residents of other areas are excluded by geography from enjoying these services. As discussed in Chapter 33, local government may supply these local public goods more efficiently than central government.

Finally, local private clubs may also provide non-rival local public goods. Private clubs can exclude people, but once someone becomes a member, he or she can usually use the facilities free of charge. This is efficient when use is non-rival. However, these solutions are not available when goods are non-excludable. They may also be considered inappropriate when the exclusion process rations access by fees that excludes low-income persons.

Efficient Supply of Public Goods

As we have stressed in this text, an optimal output of a good depends on efficiency and equity objectives. Efficiency requires that goods should be supplied up to the point where the social marginal benefit from an extra unit of output equals the marginal cost. Using economic valuation, the marginal benefit that someone obtains from a good is measured by what he or she is willing to pay for it. Equity issues are complicated because there is no single definition of, or agreement on, what is equitable. A low-income household may obtain as much or more benefit from a good such as a health service as does a high-income household but not be able to pay as much for it. In such cases government may decide to provide a service free or subsidised to all or selected households. We describe below efficient outcomes separately for non-excludable and non-rival goods. Later chapters discuss meeting equity objectives.

Efficient supply of non-excludable goods

The market demand for private (excludable) goods is the sum of the quantities of the good that individuals want to buy at various prices. In a competitive market the market determines the price and consumers vary the quantities that they purchase. The market demand curve is constructed by horizontal summation of the demand curves. Figure 11.1a shows Amy and Ben's demand for beer (say bottles of beer) as D_A and D_B respectively. At price P_M, market demand is $Q_M = Q_A + Q_B$.

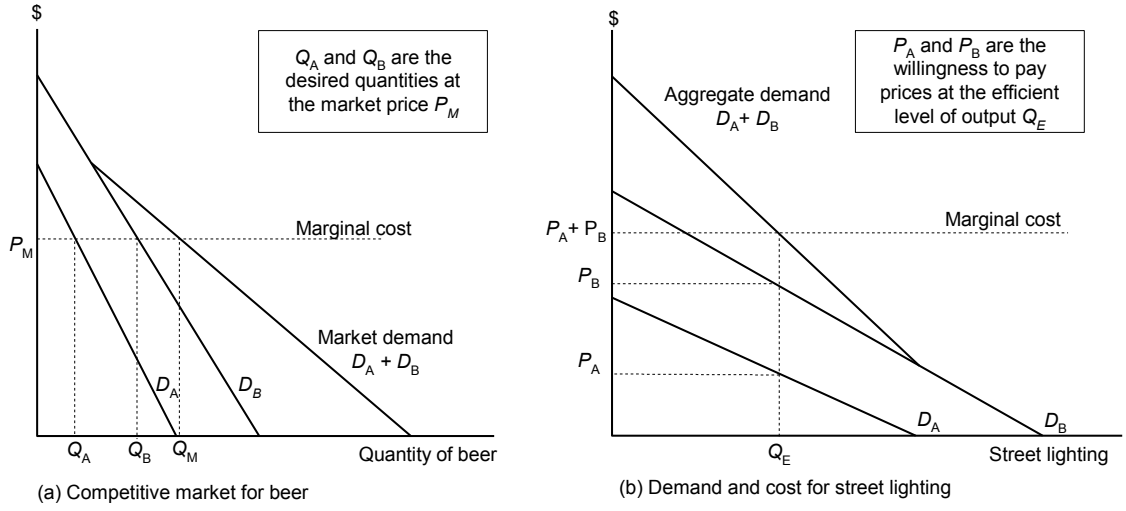


Figure 11.1 Demand and supply for private and public goods

Output Q_M is efficient. Competitive market equilibrium exists where price equals marginal cost (MC). Although Ben has more demand for beer than Amy has, they pay the same price, P_M. It follows that Amy and Ben obtain a similar marginal benefit (MB) from the purchase of one extra beer in that they are willing to sacrifice the same amount of other goods for the extra beer. Therefore, $MC = P_M = MB_A = MB_B$. This satisfies the requirements for Pareto efficiency described in Chapter 3.

On the other hand, when a good is non-excludable, once a good is provided to one household, all households have access to it, though not necessarily equal access and the amount they are willing to pay for it may vary. The collective demand curve is constructed by summing individual demand curves vertically. Figure 11.1b shows Amy and Ben's demands for various amounts of street lighting (D_A and D_B). For an extra unit of lighting at Q_E output, Amy is willing to pay P_A and Ben to pay P_B. Their total willingness to pay (WTP) amount for this extra unit of street lighting is $(P_A + P_B)$.[1]

Efficient supply requires that the amount that individuals are collectively willing to pay for an extra unit of lighting must equal the marginal cost of supplying the unit.[2] The efficient quantity of street lighting is therefore Q_E. This quantity maximises the net social benefit (NSB). The NSB is the area between the collective's demand (WTP) curve and the marginal cost curve. At any other level of output, NSB is lower and there is some inefficiency in resource allocation.

Using the terminology of Chapter 3, pure public goods are supplied efficiently when the sum of the marginal rates of substitution (MRSs) for the good for all individuals is equal to the marginal rate of transformation (MRT). An individual's MRS is the quantity of private goods that he or she is willing to give up for an extra unit of a public good. Total MRS is the total quantity of private goods that individuals are willing to give up for an extra unit of the public good. MRT is the quantity of private goods that must be given up to supply one more unit of the public good. Efficiency requires that the total quantity of private goods that individuals are willing to give up for the public good (ΣMRSs) must equal what they have to give up (MRT). Box 11.1 provides a proof for an economy with two persons and two goods.

Box 11.1 Condition for efficient supply of a non-excludable good[a]

Consider an economy with two individuals, Amy and Ben, and two goods (X is a private good, and G is a public good), where:

$X = X_A + X_B$ *and* $G = G_A + G_B$

A Pareto-efficient outcome exists when Amy is as well off as possible given Ben's level of utility subject to their total resource constraint.

Thus the aim is: Max $U_A\ (X_A, G)$ subject to

$U_B\ (X_B, G) = U^*$ and $X_A + X_B + cG = R$

where c is the cost of the supply of the public good and R is the resource constraint.

This gives a Lagrangian:

$$L = U_A(X_A, G) - \lambda(U_B(X_B, G) - \mu(X_A + X_B + c(G) - R)$$

Maximising L with respect to X_A, X_B and G gives the three first order conditions:

$$\delta L/\delta X_A = \delta U_A/\delta X_A - \mu = 0 \quad (1)$$

$$\delta L/\delta X_B = \lambda\delta U_B/\delta X_B - \mu = 0 \quad (2)$$

$$\delta L/\delta G = \delta U_A/\delta G + \lambda\delta U_B/\delta G - \mu\delta c/\delta G = 0 \quad (3)$$

From (1): $\mu = \delta U_A/\delta X_A$

From (3): $\mu/\lambda = \delta U_A/\delta X_B$

Dividing (3) by μ and rearranging:

$$(1/\mu)\,\delta U_A/\delta G - (\lambda/\mu)\delta U_B/\delta G = \delta c/\delta G$$

Substituting in for μ and μ/λ, we obtain:

$$(\delta U_A/\delta G)/(\delta U_A/\delta X_A) + (\delta U_B/\delta G)\ \delta U_B/\delta X_B = \delta c/\delta G = MRT$$

Note that $(\delta U/\delta G)/(\delta U/\delta X) = MRS_{GX}$

Therefore the Pareto-efficient outcome is where:

$$MRS_{GX}^{\ A} + MRS_{GX}^{\ B} = \Sigma MRS = MRT$$

(a) This proof is based on Connolly and Munro (1999, Box 4.1).

[1] It may be noted that Figure 11.1b has a similar structure to Figure 9.1.

[2] Strictly this assumes that the quantity of supply is continuous. If it is discrete, as for numbers of street lights, at an efficient level of supply the willingness to pay for the marginal extra unit may exceed the marginal cost. It also assumes that consumption is non-rival.

Efficient supply of non-rival goods

A key attribute of a non-rival good is that the marginal cost of consumption is zero or very low. Consider, for example, non-rival vehicle trips on roads with no congestion. The marginal cost of road use is zero, except for heavy vehicles that cause road wear and tear. The general principle for efficient pricing is that price should equal marginal cost (see Chapter 17). It follows that efficient use of a lightly trafficked road requires that there should be no charge for use other than for heavy vehicles that cause road wear and tear.

Figure 11.2 shows the demand for trips across a bridge as a function of bridge tolls. It also shows average cost (AC) per trip as a function of trips made. The AC curve shows the toll required at various trip levels to recover fixed construction expenditure (spread over the life of the bridge), assuming no maintenance expenditures. For vehicles with zero marginal user cost, the efficient price would be zero and the efficient quantity of trips would be Q_E. This would maximise the net benefit from use of the bridge. Note that this is below bridge capacity Q_C. If a road toll of say P_1 is charged to cover construction cost, the quantity of trips would fall to Q_1 and the bridge would be under-utilised. The deadweight welfare loss would equal area AQ_EQ_1. Note that with the demand curve shown, P_1 is the minimum toll that a private firm would charge to recover costs. A higher toll could generate economic rent.

We have assumed above that the bridge exists and sought to determine the efficient amount of use. What determines whether the bridge should be built? Using the net social benefit criterion, the bridge should be built if the sum of WTP amounts for use of the bridge equals or exceeds the cost of building and maintaining it in present value terms. Given the demand and cost curves in Figure 11.2 (and no maintenance costs) this condition is satisfied at any level of demand greater than Q_0.

Revealed Preference Methods of Valuation

Revealed preference valuations
Inferring an individual's valuations of a good from their behaviour

We now turn to methods of valuing preferences for public goods, or indeed for any non-market good, and start with revealed preference (RP) methods. Using RP, economists infer individual valuations of goods from observing behaviour. This includes inferring preferences from behaviour in product and factor markets even when there is no specific market for the good. Four main RP methods of benefit valuation are described: use of information from markets, hedonic price analysis, and analysis of travels and defensive behaviour.

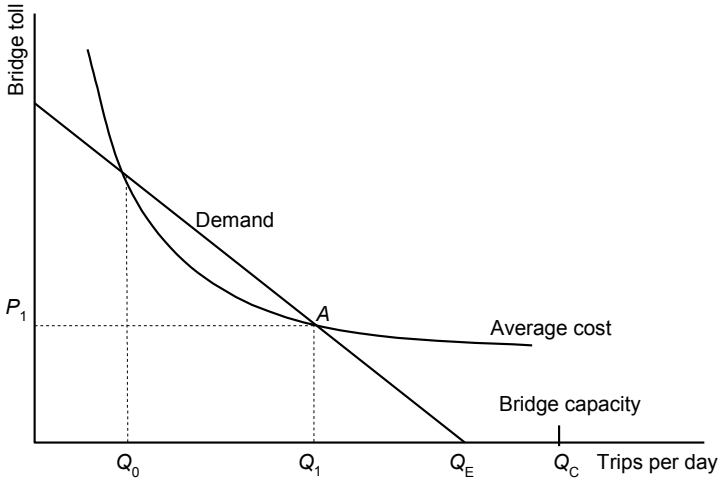


Figure 11.2 Efficient use of a non-rival lightly trafficked bridge

Markets and close substitutes

Markets are often an important source of information about values and costs even when products are not marketed. For example, in Chapter 8 we saw that the value of travel time saved in working hours can be inferred from hourly wage payments. Also, the value of travel time in leisure hours can be inferred from observing individual choices in the use of toll roads or travel modes. Although road users are not being charged directly for the time saved, the value of their travel time savings can be can inferred from the trade-off between the time saved and road toll or the extra cost of the faster mode.

Considering first the benefits of public goods to businesses, these may include savings in expenses, improvements in productivity or increased revenues from increased output. Investment in public goods, such as water or soil quality as well as transport infrastructure, may reduce industry costs and improve productivity. Programs to clean up waterways improve water quality and reduce the costs of water treatment to water supply companies and other users of water. Soil conservation programs improve soil productivity and farm output and reduce farm costs. In each case, the benefit is the sum of any savings from lower costs for existing output plus higher profits from increased output. The value of these benefits can usually be estimated from industry data or, more often, from surveys of the firms involved.

Markets can also provide information about the values of non-marketed goods to consumers. For example, the productivity benefits of a free or subsidised education or training program can be measured by the increases in earnings of the trainees. The productivity benefit component of health care programs can be estimated as a function of the increase in employment and incomes. However, as we will see below, care must be taken to identify the relationships between programs and outputs and sometimes a multivariate hedonic regression equation is required.

The value of other free or subsidised goods may also be inferred from the prices of close substitutes. For example, the value of public housing apartments may be inferred from private market rents. The value of public tennis courts or free public entertainment can be inferred from charges for private courts or prices for similar entertainment respectively. The benefits of reducing water pollution and producing drinking water from the source can be inferred from the price of (similar quality) bottled water. Note, however, that where goods are provided free, some consumers will value them at less than market prices.

In some cases, experimental studies can provide data on valuations. A key feature of experimental studies is that participants are randomly selected (see Rosen and Gayer, 2014, Chapter 2). For example, in the 1980s the US government provided vouchers for housing expenditures to a sample of low-income households to determine whether such vouchers would increase expenditure on housing or on other goods. In 1990 the state of Wisconsin in the United States conducted an experiment in which students who requested to attend private schools were randomly allocated to private and public schools with the aim of eliciting the added value if any of private schools (see Rouse, 1998). Boardman *et al*. (2011) describe several experimental projects designed to elicit WTP values for various policy changes. However social experiments may be expensive valuation methods.

Finally, brief mention should be made of laboratory experiments. These are typically artificially constructed markets in which participants are provided with cash or an endowment of tokens to be allocated to private and public goods under various scenarios. In a review article, Ferraro and Vossler (2010) conclude that ‘decision errors, confusion and noisy behaviour’ are common characteristics of these studies with the implication that findings are at best indicative. In any case, because experiments are usually one-off events in an artificial environment, care must be taken in drawing general conclusions from them.

Hedonic price analysis

Whereas market prices show the values of goods, hedonic price analysis shows the implicit values of the attributes of goods or of access to goods such as schools. Hedonic price analysis can be applied to any good or occupation with multiple attributes. The basic idea is that the value of something is the sum of the value of its attributes.

A common application of the hedonic price method is the valuation of environmental goods through analysis of house prices. This is usually done by a multiple regression study of house prices and their determinants. Consider a general relationship between house prices (P_h) and environmental and other variables,

$$P_h = p\ (S, A, E) \tag{11.1}$$

where S, A and E are sets of structural, access and environmental attributes of houses. If the equation has a linear functional form, the partial derivatives, dP_h/dS and so on, show the implicit price for a unit change in each attribute. If the equation is semi-log form, the estimated coefficient shows the percentage change in house prices as a function of a unit change in the environmental or other attribute. If the equation is log-linear, the estimated coefficients can be interpreted as elasticities.

Turning to examples, many studies (see Boardman *et al*., 2006) indicate that a one unit increase in traffic noise (measured by L_{eq}) causes house prices to fall by between 0.14 and 1.26 per cent per L_{eq}. Other studies show that a 1 per cent increase in suspended particulates (air pollution) causes house prices to fall by between 0.05 and 0.14 per cent. McNair and Abelson (2010) showed that 30 per cent of households in Canberra are willing to pay at least 2.9 per cent of their house values for underground electricity and telephone wires.

Such studies provide estimates of the implicit *marginal* WTP values of many goods. Thus the Canberra result just cited was based on a study in which 30 per cent of households were served by underground wires and 70 per cent with overhead wires. Results may also vary with differences in demand and supply conditions, so care has to be taken in transferring values from one situation to another. Estimated values can be transferred from research studies to policy areas if the environments have similar demand and supply characteristics.

Such methods can also be applied to public goods like schooling, public safety and public transport. Drawing on a house price study in Canberra, Davidoff and Leigh (2008) found that a 5 percentage point increase in test scores is associated with a 3.5 per cent increase in house prices, which is also in line with private tuition costs. They argue that estimating the effects of school quality on house prices provides a measure of the values that parents place on better educational outcomes. Abelson *et al*. (2012) used a house price study in Sydney to infer the values that households attach to lower crime rates and access to rail and bus services.

Hedonic wage studies relate earnings to the attributes of workers and working conditions. These studies can be used to derive the productivity value of education, which depends on the relationship between earnings and education, holding other factors constant (see Chapter 12). They can also be used to infer the value that workers attach to safety, which depends on the wage premium required for more risky occupations.

Hedonic wage studies have been used extensively to estimate the value of safety in the workplace and to infer the value(s) that individuals attach to life (see Box 11.2).

Travel cost studies

Analysis of travel expenditure can reveal the values of non-market goods that involve travel. Figure 11.3 (page 194) illustrates the method. This shows trips per capita per period to a park as a function of trip costs (with no entry fee). The residents of two zones (X and Y) face trip costs of C_X and C_Y and make T_X and T_Y trips per capita to the park. Given enough population zones, a relationship (V_1) between trips per capita and trip costs can be estimated. Note, however, that trip costs are not themselves WTP prices for park visits.

Box 11.2 The value of life

Traditional valuations of life equated the value of life with the discounted present value of output (income) or consumption forgone. However, this is not a revealed preference value. It is an ex-post value of life based on what is lost after the event of death. For most policy purposes, we want to know what individuals are willing to pay to reduce the possibility of early death.

Economists have developed the concept of value of a statistical life (VSL) because most policies reduce the risk of death rather than avert specific deaths. Supposing that people are willing to pay an average (mean) amount of $\$x$ for a one in 5000 reduction in the probability of their death, collectively they would be willing to pay $5000 \times \$x$ to prevent one statistical death. If $\$x$ is $1000, VSL would be $5.0 million.

Economists derive willingness-to-pay values for life in three main ways from: hedonic wage-risk studies, studies of household purchases and stated preference (SP) surveys.

In wage-risk studies, workers are assumed to require income compensation for taking on risk. The wage-risk equation is typically of the following kind:

$$Lnw_i = \alpha_o + \alpha_1 S_i + \alpha_2 X_i + \alpha_i Z_i + \pi F_i + \pi NF_i + \varepsilon_i$$

where w is the wage of worker i, S is years of schooling, X is years of work experience, Z is a vector of family or socio-economic variables, πF and πNF are the probabilities of a fatal and non-fatal injury respectively for worker i and ε_i is an error term. The wage-risk method presumes that workers understand risk differentials, which are often very small, that the model distinguishes between premiums for fatal and non-fatal accidents and that the results are not statistical artefacts of the way in which the model is specified. These strong assumptions, especially the understanding of very small risks, have led some analysts to question the inferences from the studies.

Studies of consumer behaviour infer values of life from decisions to purchase safety devices or choices between safety and travel time. For example, Blomquist *et al*. (1996) estimated the value of risk reduction implied by the use of safety belts, child restraint systems and motorcycle helmets. In order to estimate the true willingness to pay values for risk reduction, these studies may make adjustments for individuals' misperceptions of the true risks as well as estimates of time values.

In stated preference (SP) surveys, people may be asked what they are willing to pay for a reduction in risk or how they would choose between discrete alternatives that include money and safety trade-offs. However, there are concerns that respondents may not give accurate answers to questions involving small risk reductions and that answers may depend on how questions are presented. SP studies have produced a wider range of results than revealed preference studies. However, some recent studies represent sophisticated attempts to deal with these problems, for example Krupnick *et al*. (2000) in Ontario (Canada) and McNair *et al*. (2011) in Sydney.

The results of a large number of studies produce estimates of VSL that range from under US$1.0 million to US$15 million. As discussed in Abelson (2008), the most reliable results are probably in the range of US$3 million to US$5 million (in 2008 prices).

WTP values can be derived as follows. With a linear visitation relationship as in Figure 11.3, a visitor to the park from zone X has an average surplus per trip of $0.5(C_N - C_X)$, where C_N is the trip cost that deters all visits. A visitor from zone Y has an average surplus of $0.5(C_N - C_Y)$. The total value of the park is obtained by summing the surpluses over the populations in all relevant zones.

The trip data, generally obtained by a visitor survey, can also be used to derive a demand curve for the park. The key assumption is that visitors would respond to admission fees as they would to an increase in trip costs. For example, if a fee equal to $(C_X - C_Y)$ were charged, residents of Y would behave like residents of X without an admission charge and trips per capita from zone Y would fall from T_Y to T_X. Summing over all zones, the total number of visitors willing to pay a fee equal to $(C_X - C_Y)$ can be estimated. Other points on the demand curve can be obtained by other notional variations in the admission fee being interpreted as changes in trip costs. In this case, the total surplus for the park equals the area under the estimated demand curve.

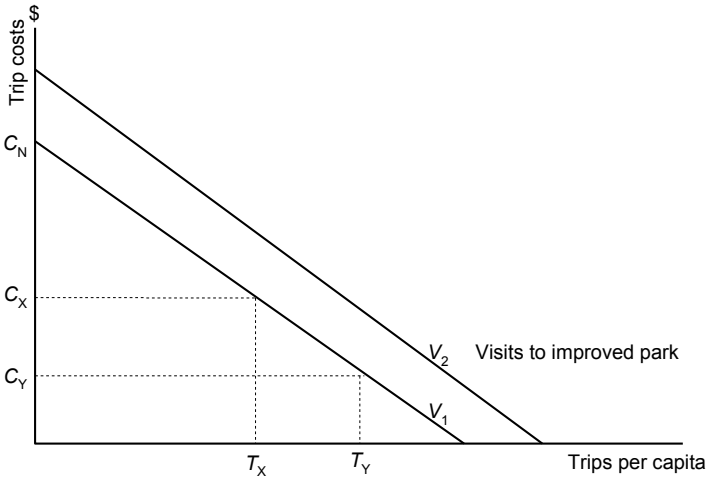


Figure 11.3 Park visits and consumer surpluses

Many studies have employed travel cost analysis to estimate the value of recreational activities. Drawing on over 200 recreational studies in the United States, Walsh *et al*. (1992) estimated that the average value of a recreation day ranged from about $30 for a low quality site to over $70 for a high quality site (in then prevailing dollar values). The highest values were for fishing and non-motorised boating. The lowest values were for camping, picnicking and swimming. Landsell and Gangadharan (2003) employed the travel cost method to estimate the value of urban parks in Victoria (Australia). Rolfe and Dyack (2011) used the method to value recreation in the Coorong (the estuarine region in the mouth of the Murray River in South Australia).

The travel cost method depends on accurate estimates of travel costs, including the value of travel times. Travel time on a recreational trip generally has a lower cost than commuting time. Another critical issue is whether the trip has one or more purposes. If a trip has several purposes, trip costs must be allocated between them. The results may also be sensitive to the functional form of the visitation equation. The size of the benefits is also sensitive to the availability of substitute sites.

In practice, policy makers are often concerned with changes in the quality of a recreational facility rather than with creating or losing a whole facility. Here the analyst must estimate the visits and associated surpluses that would occur with the improved facility (line V_2 in Figure 11.3) or draw on estimated values of activities in similar facilities.

Defensive expenditures

Defensive expenditure is expenditure that averts or mitigates the impact of an event *before* it occurs or expenditure that corrects or reduces damages *after* it is experienced. For example households or communities may construct defences against sea levels rising due to global warming. Alternatively they may accept the damages and incur correction costs after sea levels have risen.

Examples of averting expenditure or behaviour include the use of seat belts and smoke detectors to reduce the risk of injury and death (Blomquist, 1979 and Dardis, 1980 respectively); boiling water or purchasing bottled water to reduce the risk of illness from polluted water supplies (Harrington *et al*., 1989); and installing double glazed windows to

decrease exposure to road traffic noise (Garrod and Willis, 1999). However individuals may also adjust their behaviour to avoid damages. For example they may spend more time indoors on days when air pollutants are a major risk to health. There is no simple way to value this change in behaviour. Households may also be willing to undertake defensive expenditure, say against beach erosion, but be unwilling to do so because their neighbour is unwilling to do so. This is an example of what is known as a weakest link public good (Hirshliefer, 1983).

In principle the value of goods can be inferred from averting expenditures by assuming that a firm or household purchases inputs up to the point where the marginal benefit from a reduction of a unit of the unwanted effect equals the marginal cost. Consider for example double glazing of windows to reduce traffic noise. If each L_{eq} unit of noise reduces household welfare by \$1000, a rational household would be willing to spend \$1000 to reduce noise experienced by one L_{eq} unit. The marginal benefit of noise reduction can be inferred from the observed marginal expenditure of households on noise reduction.

However, the averting expenditure valuation method must be used with care. When expenditure, for example for noise insulation or for air cushions in cars, is lumpy (discrete), households may not be able to make optimising marginal adjustments. Thus the expenditure represents the *minimum* WTP value for the associated noise-free or safety benefit. The *marginal* expenditure associated with a *unit* reduction in pollution is often difficult to estimate. In addition, some expenditure provides joint benefits. For example, double glazing of windows may reduce noise and energy bills. Thus the value of double glazing may be evident but not the separate value of quiet and energy saving.

A generic example of correction expenditure is medical expenses to treat for illnesses arising from air or water pollution. This is often described as the *cost of illness method*. Harrington *et al*. (1989) discuss mitigating expenditures to measure the cost of polluted water supplies. Firms may also incur costs in the treatment of polluted water before using it in various industrial processes.

Watson and Kaksch (1982) estimated the costs incurred to clean buildings in the United States as a result of air particulates. This valuation method is sometimes called the *dose-response method*. The dose is the impact of an event and the response is the cost of restoring the capital value of the asset to its pre-dose state.

In so far as corrective expenditures quite closely restore the individuals or assets to their previous state they represent a WTP figure for the good involved. Of course the corrective expenditure may sometimes fail to restore the good, such as health, and at other times provide for an improvement. Care should be applied especially to hypothetical replacement expenditures, like nutrient replacement for soil loss, when the cost of the expenditure 'cure' may far exceed the cost of the disease.

Stated Preference Methods of Valuation

Stated preference valuations
Eliciting an individual's valuation of a good by asking him or her in various ways to state their preferences honestly

Stated preference (SP) methods of valuation represent the other major approach to estimating individual valuations of public goods. SP methods seek to elicit values for public goods by asking people to state their values or preferred choices.

Traditionally economists questioned whether people would give honest answers to WTP questions. As we saw in Chapter 9, when government proposes to finance a public good by a separate hypothecated (ear-marked) tax, individuals have an incentive to understate what they would be willing to pay for it. On the other hand, when a good is funded from consolidated revenue, potential beneficiaries are likely to overstate their WTP amounts. Either way, as Samuelson observed, 'any one person can hope to snatch some selfish benefit in a way that is not possible under the self-policing competitive pricing of private goods'.[3]

[3] Samuelson (1955, p. 389).

Despite these incentives to free ride (to lie or respond strategically), some economists argue that strategic responses are not a major problem. In early experimental work, Bohm (1972) concluded that most people answer questions honestly. Johansen (1977) and Freeman (1986) also contend that there is little evidence of strategic behaviour. However, most analysts believe that the possibility of strategic behaviour cannot be ignored.

As shown below, it is possible to devise a demand-revealing survey method that, in conjunction with tax payments, eliminates incentives to strategic behaviour. However, this method is not very practical. The contingent valuation (CV) method, in which people are asked in various ways to state their WTP amounts for various goods, is more practical. In recent years economists have also made increasing use of choice modelling (CM) techniques. CM elicits values from respondents' ranking of options, where the options include a monetary component.

Demand revelation without free riding

The demand-revelation method that eliminates strategic behaviour is based on asking individuals to state their WTP dollar values for a public good on two main premises: (1) government will provide the good (or an extra unit of it) if the sum of dollar values exceeds the marginal cost of provision and (2) each individual's payment will equal the difference between the marginal cost of the public good and the sum of the values that everyone else places on the good.[4]

To see how the method works, consider an example. Suppose that a unit increase in the supply of a public good costs $100 and that, apart from Amy, all citizens are willing to contribute $90 for that amount. If Amy is willing to pay $10 or more for the extra unit of the good, it is provided, and Amy will contribute exactly $10. If she is not willing to pay $10 for the extra unit, it will not be provided, and Amy will not pay anything.

We now show why honest revelation is Amy's best strategy. Suppose that Amy's WTP value is $10 but that she does not know the collective sum of other WTP values. Table 11.2 shows seven scenarios in which other people are willing to pay between $80 and $100 for the public good and three response strategies for Amy, giving her true WTP value of $10 and giving false values of $5 and $15. The table also shows the amount that Amy would be required to pay and her net gain in each case. The net gain is the difference between her WTP amount and her actual payment. Honest revelation of $10 is the dominant strategy. If Amy states a WTP amount of only $5, she will be less well off when the collective WTP dollar amounts of others are between $90 and $95, but she obtains no offsetting gain when the

Table 11.2 Gains from honest preference revelation (if Amy's real WTP value = $10)

Sum of others'	*MC ($100)*	*Amy states $10*		*Amy states $5*		*Amy states $15*	
WTP amounts	*less column 1*	*Pays*	*Net gain*	*Pays*	*Net gain*	*Pays*	*Net gain*
100	0	0	10	0	10	0	10
95	5	5	5	5	5	5	5
93	7	7	3	0	0	7	3
90	10	10	0	0	0	10	0
88	12	0	0	0	0	12	−2
85	15	0	0	0	0	15	−5
80	20	0	0	0	0	0	0

[4] This is sometimes described as the Clarke-Groves mechanism or tax after two economists who developed this survey method. Mueller (2003) shows that several other economists contributed to its development.

collective amounts are less than $90 or more than $100. If she states a WTP amount of $15, she risks paying more than the value of the good to her when other WTP amounts total between $85 and $90, but again receives no offsetting benefit.

Rosen and Gayer (2014) show more generally and more formally that this method should generally elicit honest answers (and true preferences). However, it is not necessarily Pareto efficient because the contributions that arise do not guarantee a budget balance. In any case, governments do not finance public goods in this manner and are unlikely to do so. Even if they did, it would be impractical to run regular, complex surveys of this kind. The surveys would be neither easy to understand nor cheap to run.

Contingent valuation methods

In contingent valuation (CV) surveys, individuals are asked to state what they would be willing to pay for a specified good if it were provided. The term 'contingent' denotes that the valuation is based on hypothetical provision of the good.

To elicit answers, a CV survey must first establish the nature of the good to be provided and the 'bid vehicle'. The bid vehicle is the way in which payment would hypothetically be made, for example in user fees, higher local taxes, contributions to a non-profit environmental fund and so on.

The actual question eliciting WTP values can be asked in various ways.

- The most direct way is by **open-ended question**. People are asked simply what they would be willing to pay for a good. However, many respondents find it difficult to answer such a direct question, especially if they lack previous experience of the issue. Analysts tend therefore to use other elicitation methods.
- Using the **payment card** method, interviewers present respondents with a range of WTP values from which to select. However, payment cards may provide implied cue values, including minimum, average and maximum values. This is called **starting point bias.**
- Using the **bidding game** technique, individuals are asked to respond to a specified dollar figure. If the respondent is willing to pay this amount, it is increased until the respondent reaches his or her maximum WTP amount. Conversely, if the respondent declines the initial amount, it is reduced until an acceptable figure is reached. In this case, responses may again be influenced by the starting figures. Potential biases with starting points, as with payment cards, may be overcome at a cost by administering questionnaires with different starting points or payment ranges to separate households.
- With the **close-ended bidding** or **referendum** model, individuals are usually presented with a single payment, which they are asked to accept or reject. This approach is designed to eliminate biased responses, although there is some concern that people are more inclined to answer 'yes' than 'no'. However, it reduces the information provided by respondents and increases the sample size necessary to generate useful information.

As with any survey, there are statistical issues. First, the required sample size may be several hundred households. The size depends on the expected variation in responses, the desired degree of accuracy and the expected non-response rate. Extensive pre-testing of the questionnaire is usually necessary. The second issue is treatment of outliers. Analysts often eliminate extreme bids on the grounds that they are likely to reflect untruthful or erroneous responses. Extreme answers may be culled by employing an arbitrary 5 or 10 per cent cut-off point, using regression techniques to estimate a bid curve or by drawing on subjective judgement more than on statistical science. Third, mean WTP values are required. Estimating mean values from a continuous distribution of WTP values is straightforward. Estimating mean values from data on the proportions of people that would pay particular WTP values requires more complex econometric analysis (Bateman *et al*., 2002). Fourth, there is the issue

of aggregation. Total WTP may be estimated simply by multiplying the estimated mean WTP by the relevant population. If the sampled population is not representative of the total population the relationships between WTP bids and various independent variables, such as income, age or household size, should be estimated and used to calculate the total WTP of the population.

Other issues with CV surveys. As we have seen, strategic bias (exaggeration or lying) is a chronic issue in CV surveys especially when respondents are not required to pay for a good. Respondents may be especially prone to exaggerate when asked what they would be willing to accept as compensation for losses of goods.

Information bias arises when answers depend on the information provided about the environment. Box 11.3 describes a CV study of values for preserving an area in the Kakadu National Park in Australia from mining. The results varied significantly with the information about the possible damages to the park provided respectively by the mining company and the Australian Conservation Foundation.

Part-whole bias arises when respondents fail to distinguish between the value of the whole good and part of it. Kahnemann and Knetsch (1992) reported no significant difference between respondents' mean WTP to protect all lakes in Ontario and one per cent of the lake areas. A related issue is the sequence problem—respondents may state a high WTP for the first good and lower values for subsequent goods.

Box 11.3 Contingent valuation of conservation value of the Kakadu Conservation Zone

The largest CV study in Australia (Imber *et al*., 1991) aimed to estimate the conservation value of the Kakadu Conservation Zone, an area of 50 km^2 surrounded by 20 000 km^2 of the Kakadu National Park for which mining was proposed. Respondents were presented with two scenarios: expected minor and major impacts that reflected the differing views of the mining company and the Australian Conservation Foundation respectively. The impacts related to mine traffic, chemicals used to extract minerals, mine process water and waste rock material.

The sample was 502 respondents from the Northern Territory (NT) and 2034 respondents from the rest of Australia. The latter were a selection of eight people over 18 years of age from a stratified random sample of 256 areas across Australia.

The study employed a two-stage referendum model to elicit WTP values. Individuals were asked: 'Would you be willing to have your income reduced by about $\$X$ a week, that is, $\$Y$ per year, for the next 10 years to add this area to the Kakadu National Park rather than use it for mining?' The starting values used were \$5, \$20, \$50 and \$100 a year, with different numbers used for different respondents. If a respondent answered yes/no, he or she was presented with the same question with a higher/lower figure. Whatever the response to the second question, there were no further WTP questions.

The mean WTP answers (in \$ per annum) were as follows:

	Scenario	
Area	*Minor damages*	*Major damages*
Rest of Australia	\$53 p.a.	\$124 p.a.
Northern Territory	\$7 p.a.	\$14 p.a.

The estimated value of preserving the Kakadu Conservation Zone was between \$0.6 billion and \$1.5 billion.

Imber *et al* (*ibid*) suggested that the study results were validated by the strong relationships between WTP values and scenario damages, between WTP values and respondent attitudes toward the environment and level of education, and between WTP values and respondent income.

However, the gap between the responses of local residents and others caused problems of interpretation. Some commentators explained that NT residents were well informed; others viewed them as 'rednecks' who did not value the environment. Curiously, respondents from the rest of Australia who had visited the Kakadu National Park also gave lower WTP values than respondents who had never visited the area. This seems to raise the question of information bias.

Diamond and Hausman (1994) reported that when respondents were asked how much they would pay to save whales and seals (in that order) seals were worth (US) $142 and whales $195. When the order was reversed whales were worth $172 and seals only $85. However, surveys can be designed to take account of the budget constraint. Willis and Garrod (1991) identified the respondents' total recreation budget before asking WTP values for the Yorkshire Dales national park.

Hypothetical bias may occur simply because individuals are asked to respond to hypothetical questions. Inaccuracy increases when respondents are asked to value unfamiliar and non-marketed goods. In such cases, an individual cannot learn from the experience of regular purchases.

A major issue in any CV study is the willingness of respondents to accept the premise that they should be willing to pay for something. For example, elderly people eligible for medical benefits in some European countries have been unwilling to state WTP amounts for health services that they receive free of charge. Many responded to CV questions with zeros indicating what they thought they should pay. CV studies are of limited use when respondents perceive that their property rights are threatened.

Together, these biases create significant valuation issues. To some extent, the biases can be minimised by survey design and implementation. Biases such as information and part-whole bias can be reduced by extensively tested survey design and by administering different questionnaires to separate groups. Hypothetical bias may be reduced by larger sample sizes. Strategic bias is harder to deal with.

Importantly, tests are required to assess the validity of any CV study. Validity refers to the degree to which the study measures the intended quantity. Smith (1993) outlines seven tests. These include comparison with outcomes in markets in which goods are actually purchased, consistency with demand theory (e.g. stated WTP should rise at a plausible rate with household income), stability of CV results in test/retest comparisons, laboratory experiments and comparisons of purchase intentions revealed in surveys with actual purchases. The main practical criteria for the acceptability of a one-off CV study are internal consistency of responses, consistency with the predictions of economic theory and consistency with the results of other credible studies dealing with a similar topic in a not dissimilar environment. Box 11.3 reports on some tests of the validity of the CV study in the Kakadu Conservation Zone.

Choice modelling

Choice modelling, like hedonic price analysis, is based on the idea that a good can be defined in terms of its attributes and the levels that these attributes take. For example, a bus service can be defined in terms of its cost, frequency and comfort. Values of these attributes are inferred from these choices.

In CM studies, respondents are asked to choose between alternatives with different attributes including a monetary amount in each case. Typically, respondents are presented with a set of six to twelve choices where each choice set has two options as well as the status quo. An option may contain four or five attributes including a cost attribute. Including more options in a choice set or more attributes in an option may be confusing. Box 11.4 overleaf shows an example of three options in a choice set with each option having five attributes.

Given the choices of the respondents, a CM study models the probability of choosing an option as a function of the utility of that option as measured by its attributes, including cost, relative to the utility (attributes) of other options. In most applications the utility of an option is modelled as a simple linear combination of costs and attributes. The choice of model depends on assumptions about random, non-measured, attributes. If the distribution of random elements is the Gumbel distribution (which is similar to a normal distribution), and the choice is between two options, the estimation model is a binary one.

Box 11.4 Valuing quality using choice modelling

Morrison and Bennett (MB, 2004) used the choice experiment method to estimate the environmental value of rivers in New South Wales. The following is one example of the choice set. Respondents were asked to choose between the following three options for the Bega River. In each case 'not sure' was also a permitted response.

Option (as in MB paper)	*One-off levy on water rates*	*Recreational uses*	*Health of riverside vegetation*	*Native fish*	*Waterbirds and other fauna*
A	None	Picnics Boating No fishing No swimming	Along 30% of river	15 native species present	48 species present
D	$50	Picnics Boating Fishing Swimming	Along 80% of river	21 native species present	59 species present
E	$50	Picnics Boating No fishing No swimming	Along 80% of river	25 native species present	88 species present

Based on the answers to such choices, Morrison and Bennett estimated the following environmental values ($ per household) for the five rivers:

Within catchment values	*Vegetation (per cent of river covered with healthy native vegetation)*	*Fish (per species)*	*Fauna (per species)*	*Fish (across whole river)*	*Swim (across whole river*
Bega	2.32	7.37	0.92	53.16	50.14
Clarence	2.02	0.08*	1.86	47.92	24.73
Georges	1.51	2.11	0.67*	48.19	27.28
Murrumbidgee	1.45	2.58	1.59	53.43	20.35
Gwydir	1.49	2.36	1.89	51.31	60.21

* Insignificant coefficients in model at the 5 per cent level.

If the dependent variable takes three or more values, a multinomial logit model is employed. Taking a choice between two alternatives (Q), the binary logit model may be written as:

$$\text{Log}\,(P_{IQ}/I - P_{IQ}) = \sum \beta_K X_{KQ} \tag{11.6}$$

where the left-hand side is the logarithm of the odds that a representative individual will choose alternative 1, the X_K are the values of the variables (K) relative to the alternative choice and β_K are the parameters to be estimated.

The marginal willingness to pay (MWTP) for any attribute is given by:

$$MWTP = -B_K/\beta_C \tag{11.7}$$

where β_C is the coefficient on cost and B_K is the coefficient on attribute K. Dividing B_K by β_C gives the monetary value of an extra unit of the attribute K. In general, in a discrete choice model that is linear in the attributes, the marginal rate of substitution between two attributes is

given by the ratio of the coefficients of the attributes. This result is frequently used to derive estimates of WTP for an improvement in a given attribute.

CM can be viewed as a natural extension of discrete binary choices between proposals and monetary values in a CV study. In a CM study the analyst has more scope to vary the trade-offs between levels or attributes of a good and money and can thus obtain more detailed estimates of the components of value. Also, because the monetary component is less explicit in CE studies, respondents may respond less strategically to the choices offered.

However, respondents do not always find it easy to respond consistently to the complexity of choices offered in CM studies and may make inconsistent choices. Day and Prades (2010) show that there are ordering effects, with answers depending on the order of the choices presented. Also, some of the problems encountered in CV studies, such as hypothetical bias, information bias and the willingness of respondents to accept the choices on offer, may arise again in CM studies.

Valuation Methods: Applications and Conclusions

Table 11.3 shows examples of public goods and relevant valuation methods.[5] The goods include educational and health services, safety, transport services, recreational facilities and environmental goods.

As can be seen, RP valuation methods of one kind or another are available for all the benefits shown. The main strengths and weaknesses of the methods are shown in Table 11.4 overleaf.

Table 11.3 Examples of public goods, benefits and valuation methods

Nature of public good/benefit	*Revealed preference method*	*Stated preference*[a]
Educational skills/training	Hedonic wage analysis	
Value of life	Hedonic wage analysis Expenditure on safe products	CV/CM
Value of health	Increased earnings Defensive expenditures Savings in medical expenditures	CV/CM
Police protection/safety	Hedonic property price analysis	CV/CM
Travel time savings (work)	Value of increased output	
Travel time savings (leisure)	Analysis of travel choices	CV/CM
Reductions in vehicle operating costs	Savings in expenditures	
Active recreational facilities	Prices of substitutes	CV/CM
Passive recreational areas	Travel cost analysis	CV/CM
Environmental inputs to production	Reduced costs of production Profits from increased output	
Environmental amenities	Hedonic property price analysis	CV/CM
Flood/fire protection	Hedonic property price analysis Savings in expenditures	CV/CM

(a) CV = contingent valuation; CM = choice modelling.

[5] In recent years, other methods have been developed to value "subjective well-being" (OECD, 2013). These methods apply principally to provision of social welfare services and are discussed in Chapter 22.

Table 11.4 Summary on willingness-to-pay valuation methods

Valuation method	*Main strengths*	*Main weaknesses*
Revealed preferences		
Market data	Easily observable Provides data on productivity	Does not measure non-market goods like quality of life
Hedonic wage method	Provides main market-based method of valuing education, training and safety	Wage premiums are not always a reliable indicator of risk
Hedonic property prices	Has many environmental and other applications and is a reliable method	Requires extensive data on variables affecting prices
Travel cost analysis	Produces reliable answers if site is accessible and study well done	Has to deal with multi-trip purposes and the value of travel time
Defensive expenditure	Provides a useful lower bound to values	Defensive expenditure may have multiple benefits and may not indicate marginal valuations
Stated preference methods		
Contingent valuation	Flexibility in a variety of applications	Answers may be biased if the study is not carried out properly
Choice modelling	Ability to generate a rich data set	May be difficult for respondents to make required comparisons

Stated preference surveys (CV or CM) can also be applied to most kinds of benefits for consumers and indeed to any attribute of any public good. Such surveys can provide information about individual values for most goods. However, because of their hypothetical nature and other issues such as information bias, respondents may not always give accurate or honest answers.

SP surveys require careful design and interpretation. Where possible, the results of such surveys should be supplemented with analyses of revealed preferences in markets or other behavioural situations.

Of course, it is often not feasible to conduct primary research for an economic evaluation. Analysts must then adopt values, modified if necessary, from other studies, especially research studies, rather than undertake a large amount of primary data collection and analysis. However, these values vary with local demand and supply conditions.

Ideally, an analysis of research studies, sometimes called meta-analysis, would provide reasons for any differences in the results between studies, so that the most relevant values can be selected and transferred to the study being undertaken. In the absence of a meta-analysis, it is common practice to adopt mean estimated values from studies that are considered broadly similar (a process known as 'benefit transfer').

In future chapters we will consider how to apply these valuation principles and practices. We will also consider how to deliver public goods. Policies are only as good as their delivery. We will discuss how the best way to supply public goods may vary with the nature of the good (whether it is a pure public good or a mixed public–private good), the degree of market failure, the amount of competition among potential suppliers and the competence (or otherwise) of government.

Summary

- Public goods are goods that are non-excludable or non-rival in consumption. Some public goods are non-excludable and non-rival.
- Public goods include most goods that are necessary for the functioning of the state as well as basic health, education, transport, environmental and many other services.
- This chapter described the efficient provision of non-excludable and non-rival public goods and methods for estimating the value of public (and other non-market) goods.
- Other chapters describe the causes of market failure, methods for estimating costs, the evaluation of public good supply and methods for supplying public goods.
- The supply of a non-excludable public good is efficient when the amount that individuals collectively are willing to pay for an extra unit of the public good equals the marginal cost of supply of that unit.
- Efficient consumption of a non-rival good requires that the price for the good should equal the marginal cost of consumption of the good.
- The valuation of public goods is a critical step in the efficient provision of public goods.
- Revealed preference methods of valuation are based on observations of behaviour in various situations. The methods include the use of relevant market and industry data, hedonic price analysis in product and labour markets, and analyses of travel expenditures and defensive expenditures.
- Stated preference methods obtain individual valuations of goods by survey methods, including contingent valuation and choice modelling methods.
- Between them, revealed and stated preference methods provide the means to value any good or any attribute of a good.
- Future chapters discuss how best to supply public goods. As we will see, the optimal provision of public goods varies with the nature of the good (whether it is a pure public good or a mixed public–private good), the degree of market failure, the amount of competition among potential suppliers and the competence of government.

Questions

1. Consider the services provided by: the Australian Broadcasting Commission, a local police station, public schools, underground power lines, waste collection services and the Tour de France bicycle race. Are these public goods?
2. For the public goods identified in question 1, what methods would you employ to determine the value of a marginal increase in provision of these services?
3. How does efficient provision of a non-excludable and non-rival public good differ from efficient provision of a non-rival but excludable public good?
4. A local government authority wants to determine how many times to clean a residential street per month. It costs \$150 to clean a street. There are 20 houses in the street. Ten households each have a monthly demand for street cleaning equal to $Q^d = 5 - 0.25P$. The other 10 households each have a demand equal to $Q^d = 5 - 0.5P$. Convert these demand curves into estimates of willingness to pay and estimate the efficient number of street cleanings per month.
5. Are strategic responses likely to be a problem in stated preference surveys of individual values for public goods? Do laboratory experiments avoid the problem of strategic responses?
6. Economists often assert that 'there is no such thing as a free lunch'. But they also identify free riding as a major problem for the provision of public goods. Can these positions be reconciled?
7. Suppose that several hedonic house price studies give different values for the cost of noise varying from, say, 0.3 per cent to 1.5 per cent of house price per average dBA above 70 dBA. Would you conclude that the estimates were faulty for some reason (if so, what would the reasons be) or explain the differences in some way?
8. A popular park is located 15 minutes from city *A* and 30 minutes from city *B*. Each city has 100 000 residents with similar incomes and preferences of park amenities. The cost of getting to the park is \$0.50 per minute including vehicle and time cost.

   There are no other users of the park. Residents of city *A* average 10 visits to the park per annum; residents of city B average 5 visits a year. Also suppose that each city pays \$250 000 per annum to maintain the park.

   i. If a linear visitation curve is derived from the two observations of the number of visits, what is (a) the consumer surplus per resident of each city and (b) the total surplus for each city?

ii. Suppose that an urban developer offers $5 million to each city for the park for the purpose of redevelopment; should the park be sold? For the purpose of this question, assume that the park has a potential permanent life and that the real discount rate to apply is 7 per cent per annum.

9. There are three households in a community. Their demand for hours (H) per week of public television is given by $P_1 = 30 - H$; $P_2 = 60 - 2H$; and $P_3 = 70 - H$. Suppose that public television is a pure public good that can be produced at a constant marginal cost of $100 per hour. What is the efficient number of hours of public television?
10. The Department of Health wishes to value improvements in health and asks you to advise on the potential use of contingent valuation and choice modelling techniques for valuing health. What advice would you give?
11. Can public goods, like surf life saving groups, be funded by tax concessions?
12. Is equality of income distribution a public good?

Further Reading

Bateman, I.J. *et al*. (2002) *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar, Cheltenham. UK.

Boardman, A.E. *et al*. (2011) *Cost–benefit Analysis, Concepts and Practice*, 4th ed., Chapters 12–15, Prentice Hall, New Jersey.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapter 7, Worth Publishers, New York.

Hindricks, J and Myles, J.D. (2006*) Intermediate Public Economics*, Chapter 5, MIT Press, Cambridge, MA.

Rosen, H.S. and T. Gayer (2014), *Public Finance*, Chapter 4, McGraw-Hill.

Samuelson, P.A. (1954) 'The pure theory of public expenditure', *Review of Economics and Statistics*, 36, 387-389.

Samuelson, P. (1955) 'Diagrammatic exposition of a theory of public expenditure', *Review of Economics and Statistics*, 37, 350–356.

Chapter

12

Education

I hold it to be indisputable, that the first duty of a state is to see that every child born therein shall be well housed, clothed, fed and educated, till it attain years of discretion.

John Ruskin, *Time and Tide*, letter xiii

Introduction ♦ Market Failures, Equity and Role of Government ♦ Returns to Education ♦ Funding Education ♦ Producing Education

There is longstanding acceptance that government should be involved in the provision of education. In ancient Greece, Plato argued in *The Laws* that education is the most important single activity in society and that the prime minister should also be the Minister for Education. Today, most governments in developed economies provide free primary and secondary education. The public good nature of education and fundamental equity considerations provide strong reasons for government involvement in education.[1] However, the extent and form of public involvement are critical and debatable issues.

In the first section below, we provide some information about expenditure on education and discuss some basic issues in the provision of education. We then consider reasons for government involvement in education, the returns to educational expenditure and issues associated with the funding and production of education.

Introduction

Overall, Australia spends about 5.6 per cent of GDP on education (see Table 12.1 overleaf). Primary, secondary, and other non-tertiary education account for about 70 per cent of expenditure on education. Universities and technical and further education account for the rest. As of 2011, average annual expenditure per student was about $12 000 in a public primary school and $14 500 in a public secondary school (Productivity Commission, 2011).

Average expenditure across OECD countries is 5.2 per cent of GDP. The range includes low rates of about 4.5 per cent of GDP in Italy and Germany and rises to nearly 8 per cent of GDP in Denmark and the United States. **CHECK** There are also differences in the composition of spending on education. Compared with Australia, on average, OECD countries allocate a higher proportion of educational funds to pre-tertiary education and a lower proportion to tertiary education.

[1] "In the old days class warfare was between the rich and the poor… These days it is clearly between the educated and uneducated", Joe Bageant, 2009, p.26, *Deer Hunting with Jesus: Dispatches from America's Class War*, Scribe Publications, London.

Table 12.1 Expenditure on education by purpose as % of GDP in 2016

	Public[a]	*Private*[b]	*Total*
Australia			
Primary, secondary and other non-tertiary	3.3	0.7	3.9
Tertiary education	0.7	1.0	1.7
Total[c]	3.9	1.7	5.6
OECD average			
Primary, secondary and other non-tertiary	3.4	0.2	3.6
Tertiary education	1.1	0.5	1.6
Total[c]	4.5	0.7	5.2

(a) Includes public subsidies to private (and religious) schools as well as direct spending on educational institutions.

(b) Net of public subsidies to private educational institutions.

(c) The totals include pre-primary spending not shown here. The OECD average for pre-primary spending is higher than Australian spending.

Source: OECD (2017) *Education at a Glance*, Table B2.3.

In Australia, in 2016 governments funded about 70 per cent of education, with the balance provided by private funds. Governments contributed 85 per cent of the funds for pre-tertiary education, but only 41 per cent for tertiary education (and this proportion is falling).

On average, OECD governments fund a higher percentage of educational expenditure, funding over 90 per cent of pre-tertiary education and 67 per cent of tertiary education.

In Australia, state governments fund and manage public primary and secondary schools and colleges for technical and further education. The Commonwealth provides funds to the public universities, although they are state-based statutory authorities. The Commonwealth also provides substantial subsidies to private schools (mainly secondary schools, and especially to religion-based schools) and some funds for technical and further education.

Issues in educational economics

Many of the problems of resource allocation that occur with non-marketed public goods arise in education. Outcomes are hard to measure and value and there is considerable disagreement about the relationship between inputs and outcomes. Figure 12.1 shows the basic relationships between educational inputs, outputs and outcomes. Inputs include market inputs of labour, capital and materials, but also household inputs, which can have an important influence on educational outcomes. Educational outputs are typically classes, course completions, exam passes and university degrees. But these outputs are only indirect measures of real outcomes: cognitive achievement, knowledge and employment skills. Some general questions can be inferred from Figure 12.1. What are the most cost-effective ways of producing educational outputs? Do these outputs achieve desired educational outcomes? What are the benefits of these outcomes? Do they justify the costs?

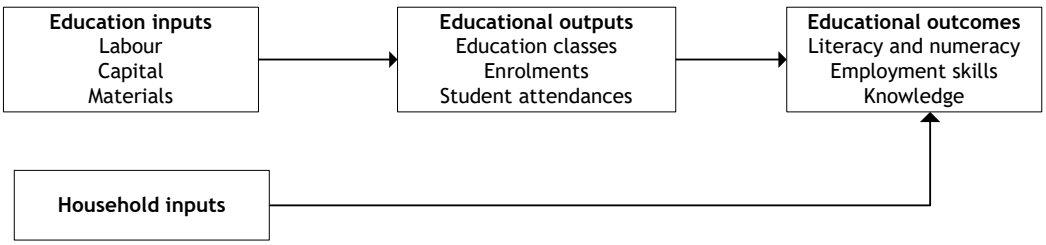


Figure 12.1 Education: inputs, outputs and outcomes

In pragmatic terms, to determine the efficient quantity of education, we want to know the net benefits (or returns) to education. To determine how that education should be financed or paid for, we want to know to what extent these net benefits accrue to the community as a whole or to private persons (to what extent education is a public or private good). We also want to know how educational outcomes can be produced most efficiently. Following a review of the role of government in education, each of these issues is addressed below

Market Failures, Equity and Role of Government

The main market failures in the education sector are the public good (positive externality) attributes of education and capital market imperfections. Also, economies of scale may limit competition. In addition, arguably some households underestimate the value of education.

Education provides substantial non-excludable and non-rival positive externalities.[2] Basic public benefits arise when people learn to read and write, communicate, understand laws and participate fully in the life of the community. These are basic requisites of an effective democracy and commercial system. Education also contributes to the social capital of a community, reducing crime and anti-social behaviour, and to the skill base of the economy. Universal primary education, especially, has long been viewed as a public good. At higher education levels, knowledge and skills pass between people in numerous ways. More educated workers make other workers more productive and in imperfectly competitive conditions make employers more profitable. Also, government gains increased tax revenues from the higher earnings. The development of European countries such as Germany, Switzerland and Sweden has often been attributed to national investment in advanced technical disciplines in the universities.

Not all economists accept that education is a public good. For example, Blaug (1989) argued that 'we cannot specify, much less measure, the externalities generated by educated individuals'. He criticised writers who simply list the various types of external benefits and infer with confidence that there is a strong case for state subsidies. However, since Blaug wrote this, there has been considerable research into the nature and size of externalities in education (see the discussion of the social return to education below).

Capital market imperfections arise from borrowing constraints. Typically, students cannot purchase education because they have neither the current income nor the capacity to borrow against future income. Few lenders accept human capital as a security against a loan. If the return from education exceeds normal lending rates allowing for risk, borrowing constraints indicate that the capital market is imperfect. In effect, children from low-income households would be excluded from market-based education. Research by Belley and Lochner (2007) in the United States indicated that credit constraints could be quite large. On the other hand, in Australia, Cardak and Ryan (2006) found no evidence of credit constraints deterring higher achieving school students from entering universities. However, they attribute this to the government's income-contingent loan scheme described below.

The literature on the economics of education pays less attention to imperfect competition in the supply of education services. However, there are substantial entry costs and economies of scale in provision of education services because of the high fixed costs. It is also hard to provide quality specialist teaching in small schools. Thus, there are significant constraints on competition in many areas of education.

The idea that some households undervalue education is a merit good argument (see Chapter 4 for introduction to merit goods). It is often suggested that households with less educated adults undervalue the benefits of education and under-invest in education. The alleged undervaluation of education by children and parents is employed to justify compulsory

[2] Of course individuals may be excluded from education, but educational providers cannot appropriate all the social benefits derived from their education.

education, for example minimum periods of primary and secondary schooling. Merit good arguments may also be used suggest that professional educational suppliers rather than parents should determine what students learn.

The merit good argument is hard to assess. Doubtless some early school leavers would benefit from more education and the increased earnings that would follow. However, given the strength of demand for education in most countries it is not clear that the benefits of education are widely undervalued. Also, the problem of early school leaving, at least over a certain age, may be addressed better by increasing student choice in educational subjects than by compelling students to attend prescribed schools or classes.

Equity issues. Equity is central to public provision of education. Many people would agree with Ruskin's dictum that the state should ensure that all children receive a basic education as of right. Normal market operations will not provide even a basic education for all and will certainly not provide an equal quality education for all. The often-observed geographical segregation of society into distinct socio-economic groups exacerbates the uneven supply of market-provided education. Accordingly, government is generally viewed as having an overriding responsibility to provide a basic education to all citizens. In developed countries, this responsibility is generally interpreted to provide free schooling at least to age 16.[3]

These equity arguments for government involvement in education may be taken further. At one level, it may be argued that the principle of equality of opportunity implies that all children should have access to equal educational resources. However, it may be contended further that children with less ability or less family support should have access to additional educational resources. These various points of view are influential in educational policy. It is the role of decision makers (politicians) to arbitrate between these different views of equity. Economists may assess the resource costs and outcomes of these alternatives.

Conclusions. Market failures and equity considerations justify substantial public funding of education. However, the precise form of public intervention needs to be determined. Importantly, the funding and delivery of education are separate issues. Arguments for funding and monitoring education do not themselves justify government production of education services. Educational services can be funded and supplied in various ways to meet efficiency and equity objectives.

Returns to Education

Education provides private and social returns. The private return is the estimated net benefit of education to an individual after allowing for any private costs incurred. The social return from education is the net benefit to society, inclusive of private and third-party benefits. In this section, we first discuss methods of estimating private and social returns and then summarise some results.

Private returns

Human capital model
An individual invests in education or training to increase the present value of their lifetime income

The human capital model provides the basis for analysing returns to education. Human capital is the present discounted value of the productivity of people with skills and training. Education increases knowledge and skills and thus human capital and earnings. In the human capital model, an individual invests in education to maximise the present value of their lifetime income. A student forgoes income now and incurs out-of pocket expenses (for learning materials, travel and fees where applicable) for more income in the future. He or she makes the investment if the present value (PV) of the increase in income over time exceeds

[3] Arguably this should include pre-school assistance as this can be critical to school performance.

the PV of the total cost to the student. Of course, this model simplifies the education decision. It assumes that the income gains from education are reasonably assured and that the individual is indifferent between studying and working. Also, individuals may invest in education for lifestyle reasons, but these benefits are rarely estimated.

The income benefits of additional education can be estimated by simply comparing the incomes of similar age persons with different educational qualifications. This would allow for any age effect. However other factors may also affect wages so the standard way to estimate the income benefits of education is to estimate a full wage equation as developed by Becker and Chiswick (1966) and Mincer (1974). This is typically of the following general form:

$$\ln w_i = b_0 + b_1 S_i + b_2 X_i + b_3 X_i^2 + e_i \qquad (12.1)$$

where $\ln w_i$ is the natural log of wage earnings for individual i, S_i is years of schooling, X_i is years of work experience and e_i is a disturbance term. The quadratic expression for X_i allows for declining earnings in later working life. Other factors, notably parental educational qualifications, can also be included. Equation 12.1 can be estimated using ordinary least squares regression and cross-section data across individuals at a point in time. Holding other factors constant, an increase of one year in an individual's schooling would increase $\ln w$ by b_1. In other words, the estimated value for b_1 represents the percentage increase in earnings for one extra year of schooling.

As with any regression model, the results may be biased by model mis-specification or omission of an important variable. For example, Equation 12.1 does not allow for hours of work. Wage rates may be a better measure of the impact of education than total earnings. An explanatory factor that is often important is parental level of education, omission of which may bias results especially if years of schooling are associated with parental learning.

Generally, the estimated return to education is biased upwards if length or type of schooling is correlated with an unobserved measure, such as parental level of education or student ability, which causes an increase in earnings. This is sometimes described as **endogeneity bias**. This occurs when the level of schooling is itself a function of other variables that may cause an increase in earnings. If more intellectually able and motivated individuals choose to undertake higher levels of education, an econometric study that does not control for these factors will give a biased (upwards) estimate of the importance of education.

Empirical studies generally attempt to deal with omitted ability bias in one of three ways: by direct controls, natural experiments or instrumental variables (see Ashenfelter *et al*., 1999; Leigh and Ryan, 2008a). The first approach is to control explicitly for ability by introducing a proxy for ability (such as results in IQ tests) into the equation to be estimated. However, these ability proxies may themselves be influenced by or related to education, leading to a downward bias in the estimated return to education. Natural experiments use natural events that factor out the effect of ability on earnings. For example, Ashenfelter and Krueger (1994) used sample sets with twins who received different kinds of education.

The third approach is estimate the education-earnings relationship by using an instrumental variable (IV). Generally, to estimate the effect of a variable x on another variable y, an instrument is a third variable z which will affect y only through its impact on x. In our present context, an IV is a variable that is related to the amount of education but not to ability. The analysis of the education–earnings relationship first models the relationship between the amount of education and the IV and then estimates the relationship between earnings and the modelled amount of education. For example, Angrist and Krueger (1991) used date of birth as an instrument. The date of birth influences the length of time that children spend in school in the United States but is unrelated to ability. They inserted the predicted values of a regression of education on date of birth into the earnings equation, thus removing the ability element. However, IV studies give biased results if the IV is itself not truly independent of earnings (that is, if it affects earnings other than via its effect on education).

Table 12.2 Social cost and benefits of public education

Party	*Costs*	*Benefits*
Students	Forgone earnings Out-of-pocket costs (transport, books etc)	Increase in private earnings Improvement in lifestyle
Government	Provision of public education Subsidies to higher education and private sector	Increase in tax revenues Lower social security payments Lower costs of health care, prisons etc
Employers		Increased profits
Other parties	Displacement of existing workers	Productivity of other workers Improvement in social capital Improvement in health and mortality Reduction in crime

Social rate of return

The main costs and benefits of publicly-funded education are shown in Table 12.2. The costs include all the public costs of education, inclusive of any subsidies to households or businesses. On the other hand, government may gain from increased taxes, lower social security payments, better qualified employees, and savings in social expenses, such as health care or corrective detention costs. Employers may profit from a more skilled workforce. Other benefits may include productivity spillovers to other workers, improvements in public health, and reduced crime.

Many of these benefits are hard to value. Wolfe and Haveman (2001) surveyed the variety of benefits from education including improvements in health, reductions in crime and an increase in social interactions and contributions to the community and concluded (p. 245) that the value of non–labour market influences is conservatively 'of the same order of magnitude as estimates of the annual marketed, earnings-based effects of one more year of schooling'. Fu (2007) provides a good quantified account of local market externalities. The OECD (2010) reports quantitative estimates of the increased proportions of people reporting good health, an interest in politics and interpersonal trust as a result of increases in education across OECD countries. Lochner (2011) provides a useful analysis of the empirical issues in sorting out the impacts of education on crime, health and political participation. There appear to be particularly high returns to completing secondary school education.

Some analysts employ macroeconomic models to estimate the full economic value of education. These models are typically cross-country regressions of the sort described in Chapter 5 in which GDP is the dependent variable and investment in education an explanatory variable. The difference between the impact of education on total income (growth of GDP) and on individual earnings is attributed to externalities. Kreuger and Lindahl (2001) argue that such macroeconomic models can be used to evaluate the social returns to education although they also consider the microeconomic studies to be the more econometrically robust. However, analysis of causation presents a major difficulty in macro studies. Hanushek and Woessmann (2010) note that there remains 'considerable controversy' over the causal interpretation of any statistical association between education or skills and output or growth—for example, whether higher cognitive abilities lead to higher growth or whether higher growth leads to higher cognitive abilities. Krueger and Lindahl (2001) also note that it is difficult to separate the causal effect of education from the positive income demand for education in cross-country data over long time periods.

Screening model of education. It is sometimes argued that the social return to education is less than the private because education simply reallocates jobs among workers. In the screening model proposed by Arrow (1973), education separates high- from low-ability people but does not necessarily improve skills. Educational qualifications are signals to employers about the likely productivity of employees. These signals minimise employers' search costs and ensure a productive workforce. In the extreme version of this model, education has no effect on productivity. Educational qualifications would provide private benefits. But the social benefits would only be savings in search costs. The case for public funding of education would be greatly reduced.

Screening model of education
Educational results act as a screening device to indentify pre-existing ability rather than the benefits of education

Analysts have tried to test for screening in various ways. Most tests attempt either to control for underlying ability in a human capital wages equation, so that remaining wage effects reflect productivity differences due to education, or to find control groups that are similar except for the amount of schooling that they receive. Gruber (2016) concludes that these tests generally support the human capital model rather than the screening model. However, some studies show a significant return to obtaining educational credentials.

Quiggin (1999) discusses two other tests. First, the screening model would predict that earnings differentials due to education would decline over time as employers could directly observe productivity differentials, but this does not occur. Second, people who plan to run their own businesses should invest less in further education than individuals looking for employment, but this does not appear to be the case. Thus, both these tests suggest that the screening model has limited applicability.

Empirical results

Table 12.3 provides an overview of estimates of the private and social returns from education for OECD countries as a whole and for Australia in 2016. The private returns are based on the discounted stream of after-tax earnings less the private costs of education. The public returns are based on the full costs of education and include government contributions, tax effects and other savings in social contributions. A low real discount rate of 3 per cent is applied to the earnings stream to estimate net present values. The estimated internal rates of return are perhaps easier to understand and more useful as a guide to investment. But whichever criterion of value is used, the average returns are high.

Table 12.3 Some key results in OECD countries in 2016

		Net present value (US$s)[a]		*IRR (%)*	
		Male	*Female*	*Male*	*Female*
Private net present value for an individual obtaining upper secondary or post-secondary non-tertiary education as part of initial education	Australia	116,600	51,900	16	9
	OECD	112,400	64,600	12	8
Private net present value for an individual obtaining tertiary education as part of initial education	Australia	209,600	147,100	9	9
	OECD	258,400	167,600	14	12
Public net present value for an individual obtaining tertiary education as part of initial education	Australia	128,700	89,900	10	10
	OECD	143,700	74,100	10	8

(a) Net present values are calculated using a real discount rate of 3%.

Source: OECD, *Education at a Glance*, 2017, Tables A7.1, A7.2, A7.3 and A7.4.

Table 12.4 Estimated rates of return by school type and area

	Social (public)			*Private*		
Region	*Primary*	*Secondary*	*Higher*	*Primary*	*Secondary*	*Higher*
OECD	8.5	9.4	8.5	13.4	11.3	11.6
World	18.9	13.1	10.8	26.6	17.0	19.0

Source: Psacharopoulos and Patrinos, 2004.

International microeconomic studies produce similar results. Using a Mincer-type equation Card (1999) found that each year of schooling increases earnings in the United States by between 6 and 15 per cent. Controlling for ability using instrumental variables or by natural experiments, he found an average rate of return on education of 10 per cent in the United States. The meta-analysis by Ashenfelter *et al*. (1999) found similar results. These are purely private returns.

On the other hand, Psacharopoulos and Patrinos (2004) provided OECD and worldwide estimates of average public and private rates of return from education (see Table 12.4). Their estimated rates of return are similar to those shown in Table 12.3. Their estimates suggest that the average private rates of return are nearly everywhere higher that the social returns. They also indicate that returns to education are significantly higher in less developed countries than in the OECD.

Early macroeconomic studies, for example Barro (1997), found that an average extra year of schooling would increase GDP growth by as much as one per cent per annum. However, Hanushek and Woessman (2010) argue that this growth effect is principally associated with increase in cognitive skills and that the quantity of schooling has no statistically significant additional effect once cognitive skills are included in the model. This does not mean that education does not increase growth. Rather, it would imply that education is important to the extent that it is responsible for building cognitive skills.

Australian studies have produced similar but slightly lower results. Using a study of twins, Miller *et al*. (2006), found a mean return to schooling of 5–7 per cent after controlling for genetic and family effects. Leigh and Ryan (2008a) used two different natural experiment techniques (month of birth and changes in compulsory schooling laws) to estimate a return to schooling of around 10 per cent after correcting for ability bias, which the authors argued accounted for between 10 per cent and 40 per cent of the gross differences. Barrett (2012) estimated that the mean return from an additional year of education is 6.2 per cent, but when cognitive ability is allowed for the return to education falls to 4.4 per cent. Barrett's interpretation is that 29 per cent of the return to an additional year of education is due to the higher levels of cognitive skills associated with an extra year of education. Barrett also found that credentials give rise to additional returns beyond those from the accumulation of years of education, especially for higher education.

Leigh (2008) estimated the return to various specific educational attainments, finding the greatest per year returns at the level of high school completion (as high as 30 per cent depending on the correction for ability bias) and Grade 10 completion (20 per cent). Bachelor and higher degrees also give significant returns as do diplomas/advanced diplomas and, for high school-dropouts only, Certificates III or IV. Drawing on extensive Australian census data, Wei (2010) also found high returns to a university education in the order of 15 per cent for males and 17 per cent for females but he acknowledges that these estimates did not take account of possible ability bias.

Finally, it should be remarked that, important as these high-level results are, government needs to know the return to *additional spending at the margin* and whether this will produce equivalent or greater benefits than on average. Thus, government needs to know, or at least

estimate, the marginal social rates of return from a larger or smaller education program and from investments in the many different levels and kinds of education.

Funding Education

As we have seen, government's role in providing funds for education arises from the public good features of education (notably positive externalities), the merit good nature of education, capital market imperfections and, above all, for reasons of equity or social justice.

In principle, the implications of positive externalities for funding education are exact. Government should provide a subsidy equivalent to the marginal external benefit of the service. If the cost of a year's education is $12 000 and the estimated external benefits are $5000, the subsidy should be $5000 and the student (or their family) would pay the $7000 balance. The aim is to produce an efficient allocation of resources to education. Students (or their families) will purchase a year's education if the expected private benefit is worth at least $7000 but will not do so if the expected benefit is worth less than $7000. Of course, this implies both that government can accurately estimate the external benefits of different kinds of education and that students understand and can afford to fund the private benefits.

There are other limitations to this approach to educational funding. One is that the subsidy must be received by, or passed on, to the student. A subsidy to educational suppliers will be passed on fully to consumers only when the suppliers are fully competitive. However, governments in Australia and elsewhere usually provide subsidies to suppliers of education who operate in an imperfectly competitive market, so that subsidies may not be fully passed on.

Secondly, students may undervalue their education. Arguably, from an efficiency perspective government would then fund the difference between its valuation of an individual's private benefit and the individual's own valuation. This would establish the appropriate private incentive to invest in education. However, there are obvious practical problems associated with eliciting student or family valuations of education when respondents have an incentive to minimise their true valuations. Accordingly, on merit good (and equity) grounds governments commonly make education compulsory for certain age groups (up to school year 10, about age 16, in Australia) and pay for most education expenses. There may remain some merit good effects beyond age 16 which would warrant a public subsidy (in addition to the externality subsidy) at older ages. However, identifying those students who need to be encouraged and subsidised is hard and community-wide subsidies are expensive.

On the other hand, borrowing constraints on private educational investment do not necessarily justify educational subsidies. The efficient policy response to this form of market failure is public loans. Australia has been a world leader in providing income-contingent loans to tertiary students and this policy is discussed further below.

But overwhelmingly, public funding of education at least for children up to the age of 16 or 17 is based on equity considerations. Children cannot pay or borrow for education. In most countries, there is widespread agreement that the state should finance education for those who cannot afford it. However, critical questions remain. Should families who can afford to pay for their children's education be required to do so? How much education should the state pay for? The answers to such questions require ethical judgements beyond the scope of this discussion. Below we briefly discuss some equity–efficiency trade-offs in the provision of education.

Who should receive subsidies: producers or consumers?

Governments usually provide subsidies to public suppliers of education. Government may also subsidise private schools or universities. The Australian government provides substantial subsidies to private schools based on the number of students they attract and, more

contentiously, on the socio-economic attributes of the local population—which may have little correlation with the school population.

The traditional system of funding state-owned institutions was based on the concept of a vertically integrated supply model. The vertical system of control facilitated administration by limiting transaction costs and the number of agencies that received a subsidy. Government could more easily control the supply of education services, how education is supplied and the prices charged. Also, it can ensure that less advantaged communities receive equitable access to educational services.

However, providing subsidies to education suppliers in an imperfectly competitive market may create cost padding and permit inefficiencies rather than improved educational services. This applies especially to government schools which provide free education and where there is limited competition between schools and limited choice for families.

Crowding out
Free public schooling may crowd out private educational expenditure and so reduce total resources allocated to education

Another issue is **crowding out**. This occurs when subsidised public schools crowd out private schooling. Suppose (*A*) that a family has to choose between a public school with an annual fee of $12 000 or a private school at $18 000 a year and chooses the latter. This family has judged that the extra $6000 expense provides a net benefit. Now suppose (*B*) that a family has a choice between a free public school and a private school at $18 000 a year, it may choose the public school. Now the subsidy has crowded out the private schooling and reduced total expenditure on education by $6000 (unless the family purchases supplementary private tuition). There is also a net social cost. True, the family has saved $18 000 under (*B*). But the taxpayer has lost $12 000 and the family has forgone an educational benefit that it valued at over $6000.

In this example marginal crowding out reduces total expenditure on education. However, other families that have low demand for education may consume more of it when it is free than when priced. Therefore, the overall effect of providing a free education service on total expenditure on education is indeterminate.

Educational vouchers

Providing subsidies to consumers avoids many of these problems but raises others. The most common proposal is that government would provide vouchers worth a given amount, say $12 000 per annum per child of school age, which can be spent at any type of school (subject to accreditation). The parents would choose the form of education, including education that requires private funding top-ups. There are several variants of this strategy (Johnes, 1993). For example, vouchers could be larger for disadvantaged children. Also, the dollar value of the vouchers could be subject to income tax. Other related issues are whether schools should be permitted to charge different fees or vary entry requirements for enrolments. If they cannot vary the number of enrolments, they have to deal with excess or insufficient demand by varying entry conditions. While educational vouchers are still rare, Sweden and various communities in the United States have some form of educational voucher system in place.

While the pros and cons of vouchers depend on the nature of the scheme adopted, some general points can be made. The main arguments in favour of education vouchers are that they promote user choice and competition among education suppliers. This competition would promote an efficient product mix (the kind of education that families want) and lower production costs. Vouchers would force public schools to compete with other public and private schools and be more cost effective. Methods such as differential vouchers can be devised to mitigate equity concerns. Marlow (1997) shows that increased competition among schools, including competition in the public sector, significantly increases student achievement. In a study of the effects of schooling vouchers on poor pupils in Columbia, Angrist *et al*. (2002) found that voucher recipients were more likely to attend private school, to attend school for longer and to have a lower drop-out rate than those who did not receive vouchers. In a follow-up study in Columbia, Bettinger *et al*. (2010) found similar positive

attendance results for voucher recipients—those who chose vocational schools instead of main-stream schooling were 25 per cent more likely to complete high school.

Turning to some contrary points, in an imperfectly competitive market where there are significant scale economies in school services, vouchers may create an inefficient outcome. Suppose that a community contains 1000 students, that school *A* can provide education at a cost of $12 000 per annum per student, and that with a voucher of say $5000 per student the families would be willing to pay $13 000 per year per student. If the voucher reflects external benefits, the net social benefit would be $1.0 million per annum (1000 students × $1000). Now suppose that a new school *B* could provide a high quality education for $18 000 per annum per student and that 300 families, with the $5000 subsidy, are willing to pay $20 000 a year at *B*. Also suppose that school *A* would lose scale economies and the cost of its services would rise to $14 000 per annum per student. Although school *B* would generate a net social benefit of $600 000 per annum (300 students × $2000), school *A* would now generate a net social loss of $700 000 per annum (700 students × –$1000). The net social surplus of $1.0 million has turned into a net social loss of $0.1 million. A simple voucher system could result in a cost-inefficient two-school solution with a lower overall net social benefit and an increase in the public subsidy for students at school *A*.

Perhaps more important, voucher schemes have potential equity problems. First, if vouchers are provided universally without means testing, they are expensive and poorly targeted. Second, when schools can vary fees and entry requirements, a hierarchy of schools from excellent to poor may emerge, with weaker students tending to attend poorer schools. Segregation over schools could accentuate prevailing social disadvantages. Third, some schools by consequence of their location may be more expensive to run.

Some educational authorities would add the merit good argument that they are better judges of a child's educational needs than are the parents.

Doubtless some of these issues may be resolved by effective scheme design. However, voucher schemes illustrate a general point about public policy. A scheme's effectiveness and acceptability often depend as much on the way in which the scheme is designed and implemented as on general theoretical principles.

Income-contingent loans

An important argument against charging fees for education is that, because students cannot readily borrow from financial institutions against future earnings, fees discriminate against poor students. One response to this could be means testing. However, it is not clear whose income (the family's or the student's) would be means tested. Nor is it satisfactory that someone should be deprived of education because their wealthy parents (or partner) will not pay for their education. Another possible response to discrimination against poor students would be government guarantees for private loans to students. But this raises moral hazard issues because financial institutions would not have full incentives to recover loans. Public loans to students do not have these drawbacks. The loans allow students to invest in education. Government has the incentive and the means through the tax system to recover loans.

An income-contingent loan is a loan for which repayment depends on the borrower's income. In a world first, the Australian government introduced income-contingent loans for university students in 1989. However, questions arise again in scheme design. Should the scheme be self-financing? Should repayments depend on income? If so, at what level of income should students start to repay their loan? What interest rate, if any, should be charged? At what rate should the loan be repaid? Chapman (2005) provides an excellent discussion of these issues. Here we make a few observations on the first two questions.

First, a self-financing scheme would require some students to cross-subsidise other students who do not repay their loan with appropriate interest. This would be actuarially unfair and, if

the premium were significant, it would lead to some students opting out of the scheme and finding capital from other sources. It would be more efficient, and probably viewed by most people as fairer, for the community to subsidise those students who do not repay their loans.

Second, there are arguments for and against making repayments depending on income. Loan repayments that depend on income are fairer in that they reflect both some of the benefits received from education and ability to repay the loan. On the other hand, they create disincentives to earning and may have adverse efficiency implications for labour supply.

In his review of income-contingent loans, Chapman (*ibid*.) found that they have two major advantages compared with private financial arrangements. They provide default protection and consumption smoothing for students. They also increase the funding available to finance higher education. Moreover, Chapman found that the introduction of higher university fees combined with income-contingent loans did not reduce enrolments in higher education. Enrolments have increased significantly. Although low-income households remain under-represented in higher education, the proportion of disadvantaged households in universities has not fallen. Further, the administration costs of higher education fees and income-contingent loans are modest. However, as Chapman notes, 'the operational and design features of such schemes are of fundamental importance to their potential efficacy'.

Equity and the allocation of education services

Equity would be uncomplicated if all children had similar abilities and opportunities. The principle of horizontal equity (that similar individuals should be treated alike) implies that each child should receive an equal share of resources devoted to education. In this case, outcomes as well as inputs would be equal.

In practice, children have different abilities and capacities to learn. The efficient use of resources requires that educational resources should be applied until the value of the marginal outcome equals the marginal cost. If an able student can learn more from an hour of teaching than can a less able student, he or she would receive more teaching hours. This maximises the outcome that can be achieved from educational resources. However, this increases any inequity due to differential ability. Not only do the able students achieve more with an equal input of educational resources, but they would also receive more educational inputs.

Compensatory education
Slow learners should receive more educational resources than fast learners even though value added per dollar of education is lower

On the other hand, the principle of vertical equity requires that children who need more educational assistance should receive it. This is called **compensatory education**. Children who are slow learners or disadvantaged in any way should receive more educational resources than would able students. Only then do individuals enjoy real equality of opportunity. In the United Kingdom the government allocates funds to local authorities based on pupil numbers in various age bands, weighted by socio-economic group. Thus, more funds and inputs are provided per capita to less advantaged groups. In Australia the Gonski (2011) review of school funding recommended additional funding per child for children with learning disabilities.

This is another example of the trade-off between equity and efficiency in public policy. Compensatory education requires that government allocates more resources per capita to children with low marginal gains from education than to children who would achieve higher marginal gains.

Subsidies to education for equity purposes are provided not only via funding differentials for slow learners but also more generally through pricing subsidies, particularly for post-school education. Again, such subsidies may not be a simple zero sum income transfer. Subsidised prices for post-school courses increase the demand for education and the allocation of resources. If prices are set below the cost of services, students may enrol in courses of low value to them and in which they apparently make little application. Unless there are significant social benefits, the misallocation of resources can be considerable.

Producing Education

Efficient production of education means achieving quality educational outcomes at least cost. In this discussion of production issues, we first discuss the use of tests of outcomes and the cost-effective size for a school. We then discuss how school management, class size (educational inputs) and teacher quality may affect educational outcomes.

Education tests. Central to most discussions about education is whether educational outcomes are improved more by increasing educational inputs or by more effective use of educational resources However, as is often the case in the public sector, it is hard to measure quality of outcomes. The outcomes that matter most are cognitive and creative skills. To a large extent, these skills must be measured by tests.

Yet the creation of tests may itself distort the behaviour of teachers and the provision of education. In addition, many factors, most notably home environment, contribute to the development of skills. To measure the impact of different educational inputs, such as differences in class size, it is necessary to account for these other factors. This may be done by assessing the change in achievements by different groups of students over a period such as a year (where the home environment is a constant) rather than by the relative levels that they achieve at the end of the period. Concern about misinterpretation of the results of tests, with excessive focus on levels rather than changes in levels, has been a major issue in the recent introduction of national student tests in Australia. It is also an issue when we consider matters like class size and teacher quality below.

School size. To provide education cost-effectively, educational authorities need to know the costs of schools of alternative sizes. They will want to avoid establishing or maintaining small schools with high unit costs. On the other hand, parents often prefer small schools that are perceived to provide more personal services than large schools and value small local schools that minimise journey-to-school travel distances and times. It is important therefore to determine the cost penalties, if any, of small schools. In particular cases, the costs of running a school may be determined from its accounts. However, for general planning purposes it is desirable to have a model of school costs in relation to size. To determine such costs, we need to estimate a cost function like:

$$C = a + bQ + cQ^2 + dZ_i \qquad (12.2)$$

where C is total cost, Q is the number of students and Z_i represents a vector of other factors. The constant term a captures the fixed cost and the quadratic function allows for possible diseconomies of size. Other cost factors may include rural locations or a high proportion of culturally and linguistically diverse students.

Johnes (1993) reviewed several estimated cost functions, mainly in the UK. He concluded that primary schools can realise substantial scale economies up to 800 students, but that unit costs level off thereafter. On the other hand, the cost-minimising size for secondary schools is about 1200 students in the United Kingdom, but higher in the United States. In the tertiary sector, the optimal size of an institution is likely to be over 10 000 students. Colegrave and Giles (2008) estimated that the efficient secondary school size in the United States is about 1543 students. These school sizes are much larger than are typically found in Australia.

School management. Public school management can be classified broadly into three models: a hierarchical centralised government control model, a decentralised public service model and an outsourced community or privately-run model. Under the first model, the education authority controls not only the curriculum in detail but also the allocation of staff to schools. School principals cannot select staff and have little control over the allocation of their

budgets. This model has been the traditional one in Australia. In New South Wales, school principals have not even been aware of the salaries paid to individual staff. The traditional model has been justified by both quality control and equity arguments (that this ensures an even quality of staff in schools around the state).

There are currently some attempts to move towards the second, more decentralised model, which would allow principals some control over curricula, allocation of budgets and staffing decisions. Indeed, this is the practice in Victoria. A more decentralised model may enhance management of both schools and staff, allow schools to be more responsive to local conditions and generally enhance efficiency by introducing an element of competition between schools. Under the 2011 Commonwealth-State Teacher Quality National Partnership, principals in 47 schools across New South Wales were given a greater say in staffing mix, budgets and other areas; those principals have generally claimed positive benefits for their students including improved attendance, behaviour and results.[4] This appears to be an area where some robust research is required.

Under the third and more radical model, public schools operate under special government charter. These schools are often called charter schools. They are publicly funded schools and held to state standards, but within the limits set by their charters they have some freedom in methods of education and in hiring and expenditure decisions. Rosen and Gayer (2014) reported that 41 states in the United States support charter schools. The UK also has charter schools. There is some evidence that charter schools improve educational outcomes. Hoxby and Rockoff (2004) reported on a case study in Chicago where there was excess demand for admittance to charter schools and the students were selected by lottery. Because admittance to charter schools was by lottery it could be assumed that students gaining entrance to charter schools were similar to those not gaining entrance (there was no self-selection bias). Hoxby and Rockoff found that those in charter schools scored higher grades in both maths and reading tests. Rosen and Gayer also observe, citing the example of Arizona, that charter schools increase diversity of choice. This is generally assumed to increase efficiency (by varying the product package according to preferences) but takes us again into the position where there may be some inequity in the provision of education.

Class size. Staff make up two-thirds of costs in Australian public schools. Therefore, government is naturally concerned about the number and pay of teachers. Currently in Australia the student to total staff ratio in public schools is 16.5, compared to 15.5 in private schools. On the other hand, teachers often argue that educational outcomes would be improved by smaller class sizes (and therefore more staff)

Formally we need to estimate production functions, as distinct from cost functions, to analyse the effectiveness of educational inputs. Following Hanushek (2002), a production function typically has the following general form:

$$O_{it} = f\ (F_i^{(t)}, P_i^{(t)}, S_i^{(t)}, A_i) + v_{it} \tag{12.3}$$

where O_{it} is the performance of student i at time t, $F_i^{(t)}$ represents family inputs cumulative to time t, $P_i^{(t)}$ is cumulative peer group inputs, $S_i^{(t)}$ is cumulative school inputs, A_i is innate ability and v_{it} is a stochastic term. Of course, all the variables must be well defined. This formulation allows student performance to be a function of cumulative factors, including non-school factors. In some models, teachers are distinguished from schools, thus allowing the effectiveness of individual teachers to be estimated.

Hanushek (2002) reviewed the results of 376 estimates from 75 studies. He concluded controversially that adding resources to schools has little effect on performance principally as measured by test scores. He found little support for beliefs that employing more teachers,

[4] Minister for Education, Employment and Workplace Relations, Media Release, 6 December 2011. However, some principals claim that the government has also used the reforms as a shield for cost-cutting.

paying teachers a higher salary or increasing overall spending will improve student performance, principally as measured by test scores. Only 27 per cent of the estimated coefficients for per pupil expenditure were positive and statistically significant. Only 14 per cent of the estimated coefficients showed lower class size has any significant positive effect. Hanushek found that teacher quality and peer effects were more important than class size.

Other studies have reached different conclusions about class size. Hedges et al. (1994) conducted a meta-analysis of a subset of early studies assessed by Hanushek, in which they took account of the precision of the estimates and found that a positive relationship between expenditure and performance was likely. Card and Krueger (1994) found that a 10 per cent increase in school spending is associated with 1–2 per cent higher annual earnings for students in later life. However, this study used earnings rather than test scores as a measure of educational output. Krueger (1998) analysed 11 600 students and their teachers who were randomly assigned to different classes from kindergarten to third grade. He concluded that, on average, performance on standardised tests increases by 4 percentile points in the first year that students attend small classes and that the test score advantage of students in small classes expands by about one percentage point per year in subsequent years. Dewey et al (2000) found that although school inputs were often used ineffectively, overall an increase in school inputs improved student performance. Krueger (2003) contended that his 1998 study represents the gold standard in research methods. He critically reviewed Hanushek's conclusions and found that Hanushek relied on many estimates from small samples within a larger study and that these had disproportionate weight in Hanushek's results. He also found that several of the studies were weak. When he allowed for study quality in Hanushek's sample, he concluded that class size is a statistically significant determinant of student performance.

Teacher quality. On the other hand, there appears to be widespread agreement that teacher quality is a key input into educational performance. Hanushek (2002) found that, of all the school measures, higher student standards were significantly related to stronger teacher test scores. Hanushek (2010) reinforced this finding. He estimated that a teacher one standard deviation above the mean effectiveness generates marginal gains of over US$400 000 per annum in present value of future student earnings with a class size of 20 and proportionately higher with larger class sizes. He also estimated large gains to the US economy from replacing the bottom 5–8 per cent of teachers with average quality teachers.

There is less agreement on how to raise teaching quality. Hanushek (*ibid.*) notes that, in the United States, there is little attention to teacher contributions and that there are few rewards for them. Leigh and Ryan (2008b) estimate that the quality of new teachers in Australia has fallen significantly in recent years. Drawing on longitudinal literacy and numerical tests of students at school and afterwards, they estimate that between 1983 and 2003, the average percentile ranking of new teachers fell from 70 to 62. Leigh and Ryan attribute this to a more than 10 per cent decline in the mean pay of teachers relative to other professions and the increased variance in salaries outside teaching. Another major factor has been the increase in non-teaching employment opportunities to women. This trend towards lower aptitude teachers represents a challenge to public sector employment practices that are based on compressed wage differentials and that tend to resist performance testing and incentive payments. Drawing on Queensland experience, Leigh (2010) estimated that the top tenth of teachers were twice as effective at adding value to students as the bottom tenth. He also noted that teacher performance was not correlated with experience or qualifications. Work by Lavy (2004) in Israel suggests that rewarding teachers for strong value-added results improves student outcomes without biasing the teaching process. Unless government responds to these market challenges, teacher quality will continue to fall in the public sector and there will be

increasing demand for private school education. In Canberra over half of all students now attend private schools.

Efficiency and equity. Finally, it is necessary to remind ourselves that productive efficiency is only part of the aim of education. For example, student streaming by ability either in separate schools or within the same school may be efficient in that it maximises the beneficial impacts of peer groups. But streaming also tends to increase inequality among children. A major challenge for public policy in education is how to use resources to maximise the return to educational expenditure along with providing equal opportunities for all and special help for slow learners.

Summary

- Educational spending accounts for 5.6 per cent of GDP in Australia. Of this, government contributes about three-quarters.
- Educational spending in OECD countries overall is slightly lower than in Australia and the share of government funding is also higher.
- Equity objectives are core to government involvement in education. The public good nature of education (extensive positive externalities) along with merit good views also justify substantial government funding of education.
- The human capital model provides the basic framework for analysing the benefits of education. Using this model, most studies find that private and social rates of return to education are high.
- For efficient use of resources, government funding would reflect the public share of benefits. Public income-contingent loans help to make this feasible for tertiary students.
- For children, equity issues are dominant and usually warrant free education up to at least 16 years of age. However free education may crowd out private education and reduce the total resources allocated to education.
- Educational vouchers provided to consumers enhance consumer choice and competition in supply. However, there are significant design issues such as means testing or taxation of the vouchers.
- Economic analysis has many applications in the supply of education. For example, cost functions can provide evidence on the cost efficiency of different school sizes.
- Production functions can show the factors, such as class size, that contribute to student performance. However, analysts differ on the impacts of alternative school management systems, smaller classes and increased educational resources on student performance.
- Teacher quality is widely considered to be an important contributor to student outcomes but there is little agreement about how this may be achieved.
- Educational programs must also meet equity objectives. Equality of opportunity or social needs may require that more resources be devoted to slow learners or to children from disadvantaged backgrounds rather than to students who will learn most from education.

Questions

1. How can the benefits of education be measured?
2. What is the screening model? How can we test whether it is a valid model?
3. It is sometimes argued that the provision of free education can reduce the amount of education supplied and consumed. How could this happen?
4. Suppose that the full cost of a year of university education is $18 000, that a student pays $12 000 and a university course takes three years. Also a university student forgoes $30 000 a year to study. Suppose further that a university education increases a student's gross income from $50 000 to $65 000 a year for 25 years, but the student's income rises after tax from $40 000 to $50 000 a year.

   What are the private and public rates of return to the student's university education?
5. Teachers often argue that smaller class sizes will improve educational outcomes. How would you test this claim?
6. What are the main arguments for and against a voucher system for primary and secondary schools? How can these arguments be resolved?
7. How might a researcher attempt to measure teacher quality?
8. What are the arguments for and against paying teachers on merit?
9. To what extent, if at all, are student literacy and numeracy tests an indicator of school quality?
10. What efficiency issues arise, for and against, income-contingent loans for education expenses?
11. How may equity considerations alter otherwise efficient allocations of educational resources?

Further Reading

Barrett, G.E. (2012) 'The return to cognitive skills in the Australian labour market', *Economic Record*, 88, 1-17.

Card, D. (1999) 'The causal effect of education on earnings', pp. 1801–1863 in *Handbook of Labour Economics*, O. Ashenfelter and D. Card (eds), Vol. 3A, Elsevier, Amsterdam.

Hanushek, E.A. (2002) 'Publicly provided education', pp. 2046–2041 in *Handbook of Public Economics*, A.J. Auerbach and M. Feldstein (eds), Vol. 4, Elsevier Science, Amsterdam.

Hanushek, E.A. (2010) *The Economic Value of Higher Teacher Quality,* Working Paper 16606, National Bureau of Economic Research, Cambridge, MA.

Johnes, G. and Johnes, J. (2007) *International Handbook on the Economics of Education*, Edward Elgar, Cheltenham, UK.

Krueger, A.B. (2003) 'Economic considerations and class size', *Economic Journal*, 113, F34–F63.

Leigh, A. and Ryan, C. (2008) 'Why and how has teacher quality changed in Australia?', *Australian Economic Review*, 41, 141–159.

Meghir, C. and Rivkin, S.G. (2010) *Econometric Methods for Research in Education,* Working Paper 16003, National Bureau of Economic Research, Cambridge, Ma.

Miller, P.A. and Voon, D. (2011) 'Lessons from *My School*', *Australian Economic Review*, 44, 366-386.

Organisation of Economic Cooperation and Development (2017) *Education at a Glance*, OECD, Paris.

Quiggin, J. (1999) 'Human capital theory and education policy in Australia', *Australian Economic Review*, 32, 130–144.

Rosen, H.S. and Gayer, T. (2014*) Public Finance*, Chapter 7, McGraw-Hill, New York.

Chapter

13

Externalities and Environment

The charming landscape which I saw this morning is indubitably made up of some twenty to thirty farms, Miller owns this field, Locke that, and Manning the woodland beyond. But none of them owns the landscape. There is a property in the horizon which no man has but whose eye can integrate all the parts, the poet. This is the best of the men's farms, yet to this their land deeds give them no title.

Ralph Waldo Emerson, *Nature*

Efficient and Equitable Resource Allocation ♦ Property Rights and Private Solutions ♦ Policies for Externalities ♦ Common Property Resources

An externality occurs when an agent's actions affect the production or well-being of third parties without compensation.[1] Such events are common. Many activities affect the environment, which in turn affects the production of other goods or the welfare of individuals. Externalities are also common in urban areas.

The basic cause of externalities is absence of property rights. With property rights, there is no externality. Without, property rights, agents of third party impacts can ignore the impacts and resources are likely to be used inefficiently. Negative externalities are more common than positive ones because the agents have no incentive to avert them whereas agents creating an external benefit may try to collect the benefit. Whatever the nature of the externality, there are potential income or welfare gains from correcting them. Negative externalities are also widely viewed as inequitable. But what are the best outcomes and best policies?

This chapter seeks to answer these questions. We first discuss efficient and equitable outcomes. We then examine the role of property rights in creating and resolving externality problems. There follows a discussion of the main policy instruments for dealing with externalities—market-based instruments and regulations. The last part of the chapter discusses policies for common property resources.

Efficient and Equitable Resource Allocation

In the absence of externalities, resources are used efficiently when the marginal private benefit (MB) from an extra unit of output equals the marginal private cost (MC). When there are externalities, the equivalent efficiency condition is that the marginal social benefit (MSB) from an additional unit of output should equal the marginal social cost (MSC), where social

[1] As discussed in Chapter 4, it is important to distinguish between physical externalities where no compensation occurs and pecuniary externalities where people are affected through price changes but there are compensating effects (winners from the changes as well as losers). This chapter deals with physical externalities.

benefit and cost includes the benefits and costs of third parties. If the level of output is such that MSB is greater or less than MSC, someone can be made better off and no one worse off by altering output to the point where MSB = MSC.

Consider first a negative externality. Figure 13.1a shows industry demand and supply curves for coal-powered electricity. In a competitive industry, the supply curve can be interpreted as a marginal cost curve. The MSC curve includes the external damage costs of air pollution. Given a competitive market, the equilibrium output of electricity is Q_M, where MB = MC. Allowing for external costs, the efficient output would be Q_E, where MSB = MSC. A reduction in output from Q_M to Q_E would reduce external damage costs by area *ABCD*. But the area ACD represents a loss of consumer and producer surpluses. So there is a net welfare gain equal to area *ABC*. The increase in price of electricity from P_M to P_E for output Q_E is a transfer payment.

Now consider a positive externality. Suppose that a property owner improves their pastures and reduces soil runoff on adjacent properties. Figure 13.1b shows the marginal private costs and benefits and the marginal social benefit from each hectare of soil improvement, where MSB includes the benefits of the property owner and their neighbours. The equilibrium market output is Q_M. However, the socially efficient level of pasture improvement is Q_E hectares. If the property owner increases his or her pasture improvements from Q_M to Q_E hectares, there is a net welfare gain equal to area *ABC*.

These analyses of efficient resource allocation are accurate, but not complete. Consider the electricity example. Our analysis suggests that the efficient policy is to reduce output of electricity. This is correct if this is the *only* way to reduce air pollution. But if air quality can also be improved by using cleaner coal or by substituting other fuels as the energy source, one or other of these actions may be more efficient than reducing output of electricity.

To obtain the optimal solution, it is necessary to focus on the real issue. Pollutants, such as sulphur dioxides, are themselves commodities. The core efficiency question is: 'What is the efficient quantity of that commodity?' Equivalently, given the quantity of sulphur dioxides in the atmosphere, what is the efficient amount of reduction? An important related question is how best to achieve this outcome?

Figure 13.2 overleaf shows two sets of costs associated with pollution: the marginal damage cost of pollution and the marginal cost of reducing pollution. The damages may

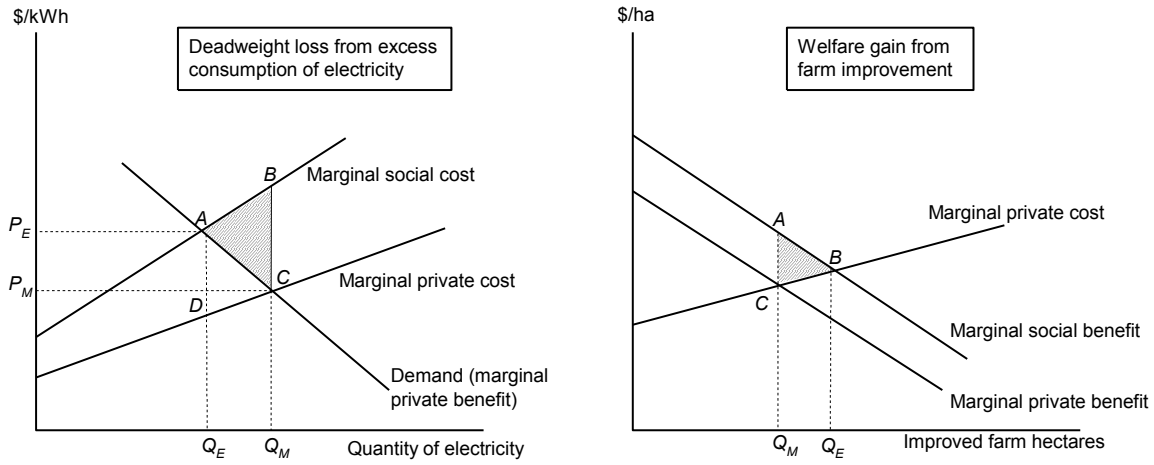


Figure 13.1 Efficient outcomes with external costs and benefits

include losses of productivity, health or amenity.[2] As shown, the marginal damage cost is small at a low level of pollution but increases with pollution. The marginal cost of a small reduction in pollution is also usually low but getting rid of all pollution is expensive. The marginal damage and abatement cost curves may have other shapes, but this does not affect the principle of efficient pollution control. Note also that the areas between these marginal cost curves and the horizontal axis represent total damage and control costs respectively.

Defining the efficient solution as the output that minimises the sum of pollution damage and abatement cost, the efficient amount of pollution is Q_E and the efficient amount of pollution abatement is $Q_F - Q_E$. At Q_E level of pollution, the marginal cost of pollution abatement equals the marginal cost of damages.[3] At higher levels of pollution (i.e. points to the right of Q_E), the marginal damage cost of pollution exceeds the marginal cost of pollution abatement and there is a net benefit from lower pollution. At lower levels of pollution (i.e. points to the left of Q_E), the marginal cost of abatement exceeds the marginal benefit of pollution reduction and it would be socially efficient to have higher levels of pollution.

The solutions in Figures 13.1 and 13.2 are efficient partial equilibrium solutions. They assume that there are no distortions in other related parts of the economy. In the examples above, it is assumed that the prices of the relevant energy resources reflect the marginal social costs of supplying these resources to electricity producers and that there are no unpriced externalities in complementary or substitute markets. If prices in related markets do not reflect marginal social costs, the efficient level of electricity output or the efficient level of pollution abatement may need to be adjusted to allow for these distortions.

In Figure 13.2, is Q_E fair as well as efficient? Any move towards Q_E is potentially Pareto efficient because the gainers can compensate the losers. But, in the absence of compensation, Q_E (and moves towards Q_E) may not be regarded as fair. If the community believes that households have a right to clean air and that firms should pay for polluting the atmosphere, government may decide that more pollution abatement and lower damages are desirable. On the other hand, if pollution controls are likely to result in unemployment, government may

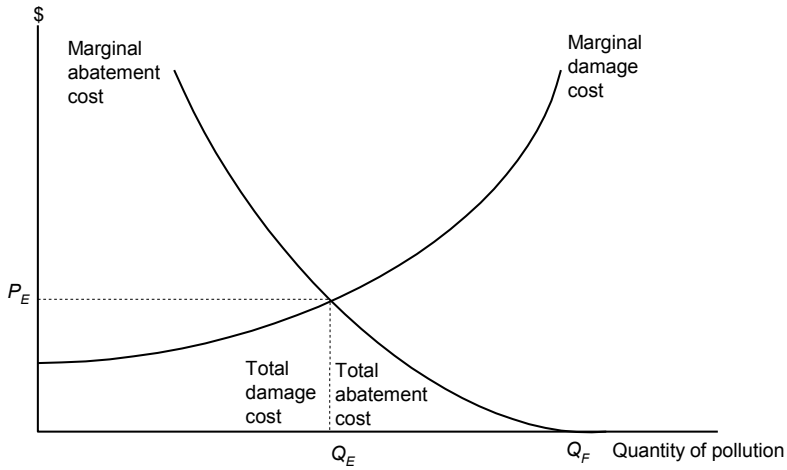


Figure 13.2 An efficient quantity of pollution

[2] The analysis assumes that these losses can be valued. Valuation principles are discussed in Chapters 6, 8 and 11.

[3] If the total cost from pollution $TC(P) = C_D(P) + C_A(P)$ where $C_D(P)$ is the damage costs which rise with pollution, and $C_A(P)$ is the abatement costs which fall with pollution, $TC(P)$ is minimised when $\Delta TC = 0$, which requires that: $-\Delta C_D(P) = \Delta C_A(P)$.

decide to allow more pollution than Q_E. Whereas for any given income distribution and set of property rights, and no distortions in related markets, the efficiency outcome is determinate at Q_E, views about an equitable outcome vary with social perspective.

Finally, we note that in the examples above the choice set is continuous and the objective is to choose the optimum quantity of a commodity. In practice, many choices are discrete. Consider, for example, private ownership of guns. The key policy issue is not how many guns an individual can own, but should he or she own any at all, and if so which kind of gun? Or consider a child care centre that may locate in either a residential or a commercial area. The residential location may minimise trip costs of parents but have more negative externalities due to the impacts of motor vehicles and child noise on local residents. In such cases, there is a limited choice set. Of course, decision makers may apply cost–benefit analysis to determine which option maximises net social benefit (although this may have limited application in extreme cases like gun controls). However, when the choice set is discrete, governments typically employ regulations to implement decisions rather than market-based instruments.

Property Rights and Private Solutions

Property right
An entitlement defining an owner's rights over resources or assets of various kinds

Property rights are central to the analysis of externalities. A **property right** is an entitlement, or bundle of entitlements, defining the owner's rights for use of a resource and any limitations on use. For a property right to be effective, it must be well defined, exclusive, transferable and enforceable. In principle, when property rights are comprehensive, there are no externalities. A property owner can obtain compensation for any damage (providing enforcement costs are moderate). However, if enforcement is costly, a formal property right may not prevent property damage.

Given that markets are likely to produce inefficient outcomes when there are externalities, it seems inconsistent to argue that private agents can deal efficiently with externalities. But, in a famous article, Coase (1960) argued that, in the absence of transaction costs, private agents will achieve an efficient outcome (and the same outcome in each case) regardless of the prior distribution of property rights or even without any prior allocation of property rights. This has become known as the **Coase theorem**.[4]

The theorem can be illustrated by an example. Suppose that Cyrus and David own adjacent properties and that Cyrus' farming practices inflict $2 million damages to David's property. Suppose also that by introducing conservation measures Cyrus could reduce damages on David's property as shown in Table 13.1 overleaf. The efficient outcome (the outcome that minimises total damage and control costs) is $0.8 million of conservation expenses, which would leave $0.5 million of damages.

Would David and Cyrus negotiate this outcome? And would the outcome be independent of the allocation of property rights? The answer to both questions is 'yes' if they act to maximise their personal advantage and negotiation is costless. Suppose that neither Cyrus nor David has any formal property rights. If he wished, Cyrus could cause $2.0 million damage to David's property. However, David would be willing to pay $0.8 million for conservation expenditure on Cyrus' property as this would save him $1.5 million in damages. David would also be better off accepting the residual damage cost of $0.5 million rather than pay for more conservation on Cyrus' property. Thus, the outcome would be efficient even if no property rights were specified. Clearly the same outcome would occur if Cyrus was entitled to damage

[4] Coase demonstrated his theorem by examples but did not provide a formal proof. Because of his informal treatment of the subject, several interpretations of the theorem exist. Another way that the theorem may be stated is that, in the absence of transaction costs, all allocations of property rights are equally efficient because the parties will bargain privately to correct any externality.

Table 13.1 Farm damages and conservation expenses ($m)

Total conservation expenses (costs)	*Marginal conservation costs*	*Total damage costs*	*Marginal savings in damages*	*Total conservation and damage costs*
0.0		2.0		2.0
0.2	-0.2	1.5	+0.5	1.7
0.4	-0.2	1.1	+0.4	1.5
0.6	-0.2	0.8	+0.3	1.4
0.8	-0.2	0.5	+0.3	1.3
1.0	-0.2	0.4	+0.1	1.4
2.0	-1.0	0.2	+0.2	2.2

David's property. Now suppose that David is entitled to compensation for any damage to his property. Cyrus must now pay David $2.0 million or adopt conservation practices. Cyrus would minimise his total cost by adopting $0.8 million of conservation and paying compensation of $0.5 million to David. In this example, *the efficient (minimum total cost) result emerges regardless of how property rights are allocated*.

However, the Coase theorem holds only in limited circumstances. These are when (1) the number of parties to the externality is small and (2) the impacts are small and there are no real income effects.

Small numbers are required because an efficient solution will be reached only if the relevant parties are well informed about the costs, do not gain by strategic behaviour or free riding and can reach a solution with limited transaction costs. In practice, externalities often involve several parties, sometimes thousands. Without an organising agency, such as government, it is not feasible for all interested parties to join in negotiations. Free riding occurs because it is in the interest of most individuals to let others organise and pay for the negotiation. Free riding is especially likely when property rights are unclear. Thus, when an externality affects many parties, private exchange generally fails to achieve an efficient outcome.

Income effects can also influence outcomes. As we saw in Chapter 6, when someone holds a substantial property right and expected real income changes are large, the amount that an individual requires as compensation for loss of the property right is often higher than the amount he or she is willing to pay to retain it. This reflects the decreasing marginal utility of income. Loss aversion may also add significantly to compensation requirements. Thus, valuations depend on the distribution of property rights. When income changes are large, the efficient outcome from a property dispute can depend on the distribution of property rights.

For example, suppose that Roger lives in a quiet street and that the local airport authority plans to build a new runway that would cause aircraft to fly low over his house. Suppose further that Roger has no right to the airspace over his house but would be willing to pay $20 000 to stop the planned runway. Although there are many other residents like Roger, their combined payments would not cause the airport authority to change its plan. On the other hand, if Roger has a right to the airspace over his house, the airport authority must compensate Roger for his loss of quiet. Roger may require a larger figure, say $30 000 as compensation, which is the amount that together with the noise overhead would make him no worse off than at present. Given the compensation payable to Roger and his neighbours, the airport authority would now look for another (lower cost) way to enhance airport capacity.

In summary, in simple cases involving only a few parties private exchange can achieve an efficient use of resources even in the absence of property rights (see Box 13.1). Also, when changes are small the outcome does not depend on the allocation of property rights. However, the allocation of property rights may affect the outcome because of income effects. In the

Box 13.1 Summary of conclusions on private exchange

No. of entities	*Size of change*	*Outcome of private exchange*
Very few	Small	May be efficient and independent of property rights
Very few	Large	May be efficient but outcome depends on distribution of property rights
More than a few	Small or large	Outcome indeterminate; likely to be inefficient and may depend on distribution of property rights

more common cases, when more than a few parties are involved, it is difficult to obtain an efficient outcome by private exchange because of high transaction costs, free riding and other strategic behaviour.

Policies for Externalities

Well-defined enforceable property rights increase the range of externality issues that can be resolved efficiently by private actions, if necessary supported by the courts. Advantages of a general property rights strategy are that a harmed party has an interest in obtaining compensation (and so ensuring that harm does not occur) and more knowledge of the damage sustained than government does.

However, reliance on property rights has two main limitations. First, legal processes are not necessarily efficient. Legal processes are often costly and may deter individuals from taking legal action. And, even if the criteria dealing with property damage appear to be clear, judicial interpretations are often unpredictable. Second, if many parties are involved, free riding remains a problem. It pays some people to let others file suit first and to await the outcome. In complex situations, like urban land use planning decisions where many parties are involved, it is unrealistic to expect private exchanges to produce an efficient outcome. Given these limitations, other instruments for dealing with externalities are needed.

The two main strategies for dealing with negative externalities are market-based instruments (MBIs) and regulations.[5] MBIs work through market prices or quantities. For example, a tax raises prices and reduces consumption. MBIs are also called economic instruments. Regulations control the behaviour of firms or households directly. Most governments rely more on regulations than MBIs. However, most regulations have to be supported by financial penalties so that the distinction between a regulation and an MBI is not always a clear one. We focus below on policies for negative externalities, with a brief addendum on positive externalities.

Market-based instruments

MBIs aim to align individual and social objectives by internalising externalities. Internalisation occurs when an agent bears the costs that he or she causes. Thus, MBIs are implicitly a property right strategy. If damage costs are internalised there are no externalities and competitive markets can operate efficiently. The most important MBIs are taxes on products that create externalities or on externalities themselves and quantity controls.

A corrective tax
A tax levied on a market activity that generates a negative externality

Taxes and subsidies for externalities. A tax levied on a market activity that generates negative externalities is generally called a **corrective** or **Pigouvian tax**.[6] To produce an efficient outcome, the tax should equal the *marginal* damage cost *at the efficient level of*

[5] Government may also exhort firms and households to behave in a socially responsible manner, but this tends to be a complementary rather than a primary strategy.

[6] Arthur Pigou (1920) is credited with initiating the concept of such a corrective tax.

output of the good or externality. For example, in Figure 13.1a, firms would be charged a tax equal to *AD*, which is the marginal damage cost at the efficient Q_E level of output of electricity. In Figure 13.2, there would be a tax (emission charge) of P_E per unit of pollutant emitted. However, as we have seen, it is generally more efficient to tax the pollutant itself rather than the good that produces it.

A Pigouvian tax (levied at the appropriate level and on the appropriate target) is efficient for two reasons. First, it encourages firms to produce *a socially efficient level of the pollutant*. The marginal damage charge becomes part of a firm's marginal costs. Firms will reduce pollution so long as the cost of reducing pollution is less than the charge. Thus, in Figure 13.2, firms will reduce pollution to Q_E because, at higher pollution levels, the cost of pollution abatement is less than the emission tax. A lower tax produces too little pollution abatement. A higher tax encourages too much abatement.

Second, marginal damage charges encourage firms to achieve the efficient level of output *at least cost*. For example, a tax on emissions allows each firm to determine how to reduce emissions in the least-cost way. Firms that can reduce pollution most economically do so. A Pigouvian tax provides a continuous incentive to firms to adjust their output or methods of production wherever adjustment costs are less than the tax. Of course, this assumes, and this is a big assumption, that the corrective tax is well designed. Box 13.2 discusses some aspects of an efficient tax on greenhouse gas emissions.

In addition, it is sometimes argued that corrective taxes have a double dividend. Nearly all taxes affect behaviour and have a deadweight loss. This means that the real cost of a marginal dollar of public revenue collected is greater than $1.00. If government revenue is fixed, revenue from corrective taxes can lower other marginal tax rates. In this case, a Pigouvian tax

Box 13.2 A tax on greenhouse gas emissions

The Stern Report (2006) described climate change as 'the greatest market failure the world has seen'. How then should governments deal with greenhouse gas emissions?

One approach is to levy a tax on greenhouse gases, such as carbon dioxide (CO_2), which reflects the marginal damage per unit of emissions. However, estimates of the present value of the stream of future global warming damages per tonne of CO_2 vary widely. According to Nordhaus (2007) the median estimate of marginal damages is between US$30 and US$40 per tonne of carbon emitted, which is equivalent to US$7 to $10 per tonne of CO_2 emitted. Note that one tonne of carbon is equivalent to about four tonnes of CO_2. But these estimates appear lower than those of other experts. Estimated damage costs are very sensitive to assumptions about climate change.

An efficient carbon tax would be levied at the same rate in all countries because all unit emissions of CO_2 cause equal marginal global damage. This would create an incentive to reduce emissions when the marginal control cost is less than the marginal damage. Such a uniform global CO_2 tax would reduce worldwide CO_2 emissions at least cost because the reductions would be made by firms and households when the costs of doing so were less than the carbon tax.

A fully tradable set of carbon permits would also be cost-effective because firms that could reduce CO_2 emissions most cost-effectively would do so and firms that could not do so would purchase permits.

On the other hand, direct government investment in carbon reducing technology is likely to be fiscally costly and inefficient without competitive market disciplines.

However, there are problems associated with a tax on CO_2 emissions. Practical problems arise because countries have widely different taxes on fuels so that it is hard to identify a CO_2 tax element and equalise it over countries.

On the other hand, a unilateral CO_2 tax may have little effect on worldwide emissions and global warming and risks putting a country at a competitive disadvantage.

In Australia, the government legislated that from July 2012 facilities emitting more than 25 000 tonnes of CO_2 would pay $23 per tonne of CO_2 or the equivalent amount of nitrous oxide or methane emitted. This applied mainly to miners and electricity producers with agricultural activities largely exempt. Some fiscal measures were introduced to compensate low income households from rising costs.

However, the government repealed the scheme as from 1 July 2014 and replaced it with The Emission Reduction Fund in December 2014, which was designed to fund public investment in renewable energy with only modest CO_2 emission reduction targets.

would both correct a market distortion and reduce distortions from other taxes. But this ignores the possibility that a corrective tax may also have a distorting effect on labour supply.

Subsidies for pollution reduction can also produce an efficient level of emissions! A subsidy of $\$X$ per unit of emission reduced has a similar short-run incentive effect to a corrective tax of $\$X$ per unit of emission. Indeed, governments sometimes employ subsidies to encourage firms to introduce pollution controls, especially when old plants may not be viable with high new expenditures and employment is a concern.

However, subsidies for pollution reductions have two main disadvantages. First, they may encourage firms to enter the industry and under-invest in pollution control technology to gain a subsidy. Second, subsidies change the distributional effects—taxpayers bear the cost of controlling the damages rather than the firm. This may well be viewed as inequitable.

Quantity controls and market creation. With Pigouvian taxes, government sets prices and the market provides efficient quantities. Alternatively, government may set a total tradable quantity (or quota) for a commodity and markets establish the price per unit. For example, in Figure 13.2 a regulator would establish the efficient ambient quantity of the pollutant (Q_E) and the market would set the price per unit of pollutant. In some cases, this may mean creating a market. Like an appropriate Pigouvian tax, setting an efficient tradable quantity will produce an efficient outcome (i.e. an outcome that maximises net social benefit).[7]

Tradable emission permits are an example of this approach. The regulator determines the total quantity of emissions to be permitted. Firms are allocated, freely or by sale, permits that control the quantity of their emissions. Importantly, firms can trade these permits.

In the absence of trading, the permits are essentially a form of regulation which, as shown below, tends to be a more expensive policy. The demand for permits in conjunction with the regulated supply determines the price of the permits (see Figure 13.3). A tradable permit

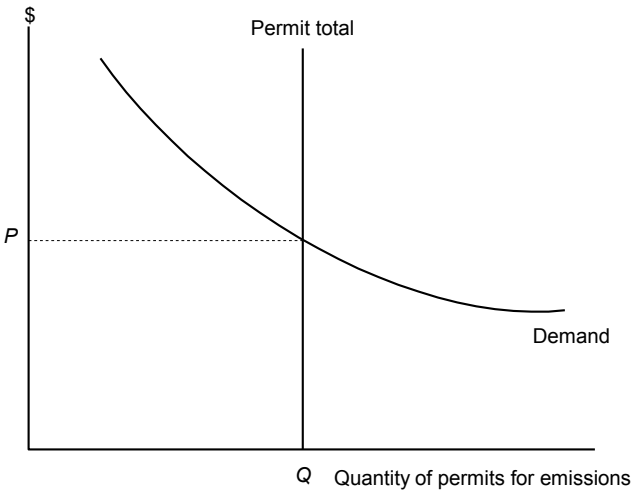


Figure 13.3 Creating a tradable permit market

[7] This assumes that government is certain of the marginal benefits and costs of the policies. Rosen and Gayer (2014) show that if there is uncertainty, the preferred instrument depends on the circumstances. When the marginal social benefit (MSB) of pollution reduction is inelastic and the marginal cost (MC) of pollution reduction turns out to be higher than expected, a tradable quota is more efficient. If the MSB is pollution reduction is elastic and the MC of pollution reduction is higher than expected a Pigouvian tax is more efficient. However, this seems to assume that the policies cannot be adjusted to meet changes in knowledge.

system has several advantages. First, unlike a corrective tax, it directly controls the quantity of pollution. Second, like an emission charge but unlike regulations, in a competitive market tradable permits produce a given reduction in pollution at least cost. Firms with low control costs introduce control first and sell permits to firms with higher control costs. Firms with high control costs can operate by purchasing permits without being subject to costly regulations. Third, there is a continuous incentive to search for ways to reduce pollution. Tietenberg (2006) reports that trade in emission permits reduced the costs of complying with the *US Clean Air Act* by over $10 billion compared with a regulatory approach in which all plants would have been required to achieve the same standard of output for emissions.

However, tradable permits work well only if the traded unit is well designed and standardised for marginal damage effects. Problems arise if the units have different damage effects because, for example, emissions of sulphur dioxides create more damage in some areas than in others.[8] Also, a sophisticated monitoring process, often involving significant transaction costs, is required to validate trades and ensure that conditions in the permits are observed.

Examples of market-based instruments. Governments employ a variety of MBIs. These include Pigouvian taxes on negative externalities such as waste emissions or effluents and trading permit schemes. Taxes on polluting products, for example petrol and diesel, energy products, fertilisers and pesticides, batteries and non-returnable containers, are more common than emission taxes. Product taxes are easier to administer but may produce less efficient outcomes. Tax differentiation is a related strategy. In some European countries cleaner cars attract lower sales taxes or registration fees. They also tax unleaded petrol more highly in order to encourage use of unleaded petrol. Deposit refund schemes impose a surcharge on the price of products with waste costs. The surcharge is refunded when the waste product is returned. Deposit refunds have applied to beer and soft drink bottles, other containers, batteries and even cars. Box 13.3 provides examples of these and other MBIs employed in OECD countries.[9]

Box 13.3 Examples of market-based instruments in OECD countries

Sweden is a leader in environmental charges. It levies taxes in the energy sector on oil, coal, natural gas and liquid petroleum gas, and in the transport sector on gasoline, diesel and domestic air traffic. In 1992 Sweden introduced a nitrogen oxides emission charge on large energy sources, which covered some 120 heating plants and industrial facilities with 180 boilers. Today, among OECD countries, Sweden has the highest tax rate on NO_x emission, followed by Norway and Italy. Few other countries have a significant tax on NO_x emissions.

Sweden also has a sulphur tax and a carbon tax. And Sweden levies special charges on nitrogen and phosphate fertilisers and another on acreage of land treated with pesticides.

France has introduced charges on the sulphur content of fuels and announced plans for nitrogen oxide and sulphur dioxide taxes.

Most countries have landfill charges for household waste. Austria, Denmark, the Netherlands, Norway, Sweden and the United Kingdom have especially high landfill charges.

Germany, France and Austria have introduced charges, levied on producers, to control packaging wastes. Many countries employ deposit refund schemes for bottles and aluminium cans. Norway, Sweden and Germany have extended the system to car hulks.

In 2002 the Irish government introduced the Plastic Bag Environmental Levy. Retailers were charged 15 cents per plastic bag, which they were required to pass on to consumers. The number of plastic bags used fell by over 90 per cent in one year.

[8] The same is true for Pigouvian taxes that should reflect marginal damages rather than emissions regardless of location.

[9] For further details of environmental taxes in OECD countries, see <www2.oecd.org/ecoinst/queries/index.htm>.

Regulating externalities

Governments employ numerous regulations to deal with negative externalities. Examples include prohibiting firms from discharging toxic waste into public places, setting fire codes for buildings to protect public safety and regulating driving speeds and amounts of alcohol that may be drunk before driving. Government may also prohibit individuals from owning certain types of guns or from smoking in aeroplanes and pubs. Local government typically regulates land use in some detail.

Regulations are of three main kinds. First, government may **regulate the activity** that creates the negative externality. Examples are restrictions on uranium or sand mining or use of sensitive wilderness areas for tourism or hunting. Such controls are appropriate when the damage is severe and all levels and methods of exploiting the environment are likely to be harmful. But when an activity has multiple outputs, it is generally preferable to focus on the output that creates the damage rather than to regulate the whole activity. Activity regulation is generally inappropriate if the problem is not severe and other methods of regulation are available. For example, it would generally be inefficient to regulate output of electricity in order to control air pollution if alternative pollution abatement actions were available.

Second, government may **regulate production processes or inputs**. Such regulations prescribe how output is produced. Examples are requirements that power-generating plants use only prescribed grades of coal, that industrial plants have effective scrubbers and other pollution control devices, or that smokestacks be a certain height. Controls over the location of industry (land use zoning) are another common form of input control.

Third, government may **directly regulate the externality**. A common example is an emission standard. This is a legal limit on the amount of a pollutant that an individual source may emit. Emission standards are widely used for industrial discharges to water or air and for automobile emissions.

Regulation of the externalities themselves has several advantages compared with regulation of production processes. First, it directly addresses the problem. For example, emission standards deal directly with the quantity of industrial wastes. Second, it allows each firm to decide how to deal cost-effectively with the problem and the firm is likely to know how to do this better than the regulator. Third, controls over production processes work only if plant and equipment are maintained adequately. This is often not the case. However, governments often think that it is easier to monitor production processes than emissions. Accordingly, they often adopt process controls.

Regulations compared with MBIs. Regulations have several advantages compared with MBIs. They are usually practical. They deal well with discrete problems like gun ownership or protection of endangered species. They can deal specifically with the problem that needs to be addressed. Moreover, in principle they can provide an efficient outcome. For instance, an efficient level of some environmental attribute. In addition, for some externalities, such as conflicting land uses, disposal of toxic wastes and control of gun ownership, there may not be a feasible MBI alternative.

However, regulations also have significant weaknesses. First, to determine an efficient outcome the regulator must be aware of marginal costs and benefits with limited or no access to price information. This is unrealistic. Consider, for example, emissions to the atmosphere. An efficient overall level of environmental quality requires an efficient emission standard for each plant which may vary by plant and location. To achieve this by regulation the regulator must know the marginal savings in damage costs and the least-cost method of abatement for each plant. But the regulator cannot know all these data. Moreover, firms have no incentive to disclose their production costs. A regulation that requires all firms to achieve the same environmental standards is highly unlikely to be cost-efficient. Second, regulations are usually a crude instrument and provide crude incentives. If a firm satisfies a regulation, such

as an emission standard, it has no incentive to adopt more cost-effective technologies and reduce damages any further. Jorgenson and Wilcoxen (1990) estimated that environmental regulations on 35 industry groups reduced the national product in the United States by 2.6 per cent. As shown in Box 13.4, regulations often fail to produce a cost-effective reduction in pollution.[10]

Moreover, regulations must be supported by penalties, most often by financial penalties. For most breaches of regulations imprisonment is too severe a penalty. However, if the fines are low, regulations are disregarded and ineffective. To be efficient, the fines must reflect at least the marginal damage cost at the efficient level of output. But these fines are equivalent to efficient emission charges! In practice, many countries impose low fines for infringing emission standards (China being a prime example). Firms then choose to pay the fine rather than meet the standard. Occasionally penalties are much higher than marginal damage costs. For example, factories may have to close if they do not meet an emission standard although closure costs exceed marginal damage costs.

Equity issues

Policy makers often want to know the equity implications of alternative policies. These implications are often difficult to determine. Here we mention three significant complications.

First, there is a common presumption that one or more parties create the externality and that other parties bear the costs. However, this implies that there are established property rights. Suppose that there are two parties, *A* and *B,* where *A* is perceived to be the polluter and *B* the damaged party. The damages occur only because of *B*'s claims of a property right. Externalities are bilateral. Without *B*, there would be no externality. Suppose that A is farming some land without neighbours and that *B* moves into part of the area damaged by A's farming practices. Arguably, *B* has created the externality. Or suppose that aircraft are flying harmlessly over farmland, which is turned into urban housing. Arguably the new residents

Box 13.4 Cost of regulations for air quality

The standard policy-making approach to stationary sources of air pollution is (1) to select an appropriate ambient standard for each major pollutant, for example sulphur dioxide, averaged over a specified period and (2) to select emission standards for each major source that will satisfy this ambient standard. The emission standards are generally imposed on several specific emission points, stacks, vents, etc. The time may be defined over a long period such as a year or a short period like three hours.

Usually similar ambient standards must be met everywhere although monitoring is done in specific locations. Emission standards may vary with plant, with allowance for age. New plants are often required to meet the lowest achievable emission rate (regardless of benefits) but this is not a fixed amount as 'achievability' depends on the costs of new technology that may not be proven.

The regulatory approach used to control air pollution in most countries is emission standards. But emission standards often fail to produce the required ambient standards at least control cost. Studies in the United States have found that abatement costs are 70–80 per cent higher under the mandated emission standards than they would be if the least-cost control methods had been adopted (Tietenberg, 2006, p. 379). This is because the emission standards for each plant do not vary sufficiently. Plants where it is costly to abate pollution must achieve similar standards to those that can reduce pollution at low cost. Also, standard emission controls set only a maximum level of emissions. There is no incentive to do better than the regulated emission level.

[10] Despite these disadvantages of regulations governments often prefer to regulate because regulations are easier to understand and appear easier to implement. A public choice explanation for the preference for regulations is that regulations create more work for public servants.

have created the noise nuisance externality. Evidently it may be unclear as to who creates and who bears a negative externality. Attribution of responsibility may be a political judgement and may have a strong bearing on how policy makers decide to deal with any individual situation.

Second, although externalities are not traded directly, markets often create implicit trades in externalities. For example, environmental externalities are traded indirectly when people buy houses. House price differentials reflect environmental qualities. Households pay a quantifiable premium for the environmental attributes of a house and its location. This is efficient because households buy the amount of the externality (e.g. air quality) that they want. It is equitable because households buying into environmentally disadvantaged areas are compensated by lower prices.

Third, externality policies often have many economic effects, again through market mechanisms. For example, policies that require firms to introduce controls to reduce pollution generally increase production costs and reduce output. These costs may be borne partly by consumers due to higher prices, by the shareholders owing to lower profits and by employees in the form of lower wages or employment owing to the lower demand for labour. The outcome will depend, *inter alia*, on the elasticity of demand for the relevant commodity and the ease with which the factors of production can move to other employment.

To develop equitable policies, it is necessary to establish who is creating and bearing the negative externality, whether those bearing the costs are compensated by market mechanisms and how policies dealing with externalities may impact on various social groups. There is no simple prescription for dealing with these equity effects. They have to be established on a case by case basis.

Positive externalities

Goods that have widespread positive externalities, such as public health, are treated in this text as public goods. When positive externalities are more limited, they may be treated more specifically as externalities. Examples of such goods include specialised training courses, in-firm research and development activities, farm conservation practices and property renovations. Each of these may provide beneficial third-party effects. Of course, the difference between a public good and a good with a more limited positive externality effect is one of degree rather than a firm distinction.

We have seen that goods with positive externalities are usually under-supplied in the market. Efficiency requires that output should be expanded to the point where the marginal social benefit of the last unit of output equals the marginal cost. There is therefore a *prima facie* case for public support for more output of these goods.

The most common way to achieve this increase in output is by subsidy, either by way of a grant or a tax concession. To be effective, the subsidy must relate to the firm's marginal operations or a household's marginal choice. To achieve an efficient level of output the subsidy should equal the marginal external benefit at the efficient output level.

In practice, the marginal impact of a subsidy is difficult to assess. The Australian government has provided substantial subsidies to industry research and development (R&D) activities, usually by way of tax concessions. However, R&D is difficult to define. It also seems likely that a high proportion of the R&D expenditure that received tax concessions would have occurred in any case without the concessions. Similar issues arise with assistance in any sector, for example with the provision of assistance to farmers for soil conservation. A subsidy may simply support ongoing soil conservation activities rather than generate new ones.

In summary, positive externalities justify public support for certain activities by firms and households. However, efficient and effective public support means identifying and targeting the marginal external benefit. This is not easy to achieve.

Common Property Resources

Common property resources (or common-pool resources) are another example of poorly defined or defended property rights. The term 'common property resources' denotes resources to which any party has free and unrestricted access.[11] Because of the unrestricted access, the users do not have adequate incentives to look after the resource. Consequently, the resources are over-exploited.

Common property resources can occur in conjunction with various legal concepts of ownership. The main relevant legal concepts are: *res nullius*, *res communes* and *res publicae.*

Res nullius denotes a regime in which no one owns or exercises control over the resources. Access is unrestricted and on a first-come, first-served basis. This occurs more often in traditional societies than in modern economies. However, surfing waves might be regarded as a modern example of *res nullius*. ***Res communes*** means formally private property owned by more than one individual. In economic and legal literature, common property tends to mean property jointly owned and managed by a group of people rather than by a few individuals. The co-owners may have formal legal or traditional property rights to the land. Common property regimes encompass a variety of access entitlements. For example, in Switzerland the Alpine meadows have been treated as common property for hundreds of years. An association of users has established members and agreed grazing rights and responsibilities. In effect, there are exclusive property entitlements. However, in other communities there is largely unrestricted access to community-owned land. For example, in the Western Highlands in Kenya, there is generally free access to the forests owned and controlled by the traditional groups. ***Res publicae*** refers to property owned by the state on behalf of the public. Public parks and beaches along with maritime resources are examples. In the United States, over 40 per cent of the land is owned by the state. The state is also a large landowner in Australia. When the state owns the land, it can restrict access. But, typically, access will be restricted only if government ensures this through policy making.

Free (unrestricted) access to resources occurs most often with *res nullius* or *res publicae* but can also occur with *res communes*. A key attribute of common property resources is that use of them is rival. One agent's use of the resources reduces the availability of the resources to other agents. Thus, the economic literature generally treats common property resources as an externality issue rather than as a public good.

As we saw in Chapter 4 with overgrazing of the village commons, common property resources tend to be misused because private interests diverge from social interest. The exploitation of scarce animal resources such as the African elephant or rhinoceros is a popular example of this problem. In Australia, surfers crowd each other out on good surf waves. Similar issues arise with any valuable common property resource unless there is some form of group ownership and control. Without appropriate policies, firms or individuals have little incentive to use these resources in a socially responsible way. The main issues and policy responses are illustrated below in the discussion of common fishery resources.

Common fishery resources

Two main problems occur with common fishery resources and both lead to over-exploitation of the resources. The first problem arises because the private revenue of the marginal fishing group exceeds the social revenue. Suppose that 10 fishing groups catch 10 000 tonnes of fish in a season and catch, on average, 1000 tonnes. Now an eleventh group, with similar technology, enters the fishery and the total catch rises to 10 500 tonnes. The average catch falls to 955 tonnes. This allows each group, including the new one, to cover their costs.

[11] In the economics literature the term usually applies to resources employed in production. But it could also apply to environmental amenities like air and water.

However, the marginal revenue from 500 tonnes would not cover the marginal costs of the eleventh group and there is excessive use of resources in fishing.

This scenario is illustrated in Figure 13.4a. Note that the unit on the horizontal axis is fishing hours (a measure of labour input). The marginal cost of a fishing hour is assumed constant. Given a constant price for fish, the average revenue curve reflects changes in the average product per fishing hour. Average product and revenue per fishing hour fall as fishing hours increase, as more labour chases a fixed stock of fish. Marginal revenue is less than average revenue because marginal product is less than average output. This occurs because fishing is a rival activity. The introduction of an extra fishing group reduces the catch of the existing fishing groups. Total fishing hours will be Q_M, where average revenue equals marginal cost. But the efficient quantity of fishing hours would be Q_E, where marginal revenue equals marginal cost. As fishing hours increase beyond Q_E, marginal revenue is less than marginal cost and resources are employed inefficiently. The deadweight loss is the excess of marginal cost over marginal revenue.

The second cause of overfishing is the lack of property rights over the future fish stock. Each fish has a breeding value as well as a current consumption value. The breeding value of the whole stock is the present value of the net benefit from the future catch. For small fish, breeding value exceeds consumption value. Each fishing group obtains the consumption value of fish that it catches but only a small part of the breeding value of fish it leaves behind. Thus collectively they over-exploit the fish stock. This is illustrated in Figure 13.4b, which allows for loss of fish stock. The social cost of fishing is the cost of fishing plus the loss of breeding value of the fish stock. The efficient level of fishing is Q^* rather than Q_E. At this point, the marginal revenue from the last fish caught just equals the marginal social cost of the catch.

The efficient level of fishing can be achieved using one of the two major instruments previously described: corrective taxes or quantity controls. A tax on the fish catch would raise the price of fish and reduce the consumption of fish and therefore the amount caught. Referring to Figure 13.4b, to reduce the fish catch from Q_M to Q^* government would need to impose a tax equal to $(P^* - P_M)$.

Alternatively, government can establish a total fish catch quota of Q^* to be allocated to fishermen. If this allocation is auctioned, the price for a permit to catch fish would be similar

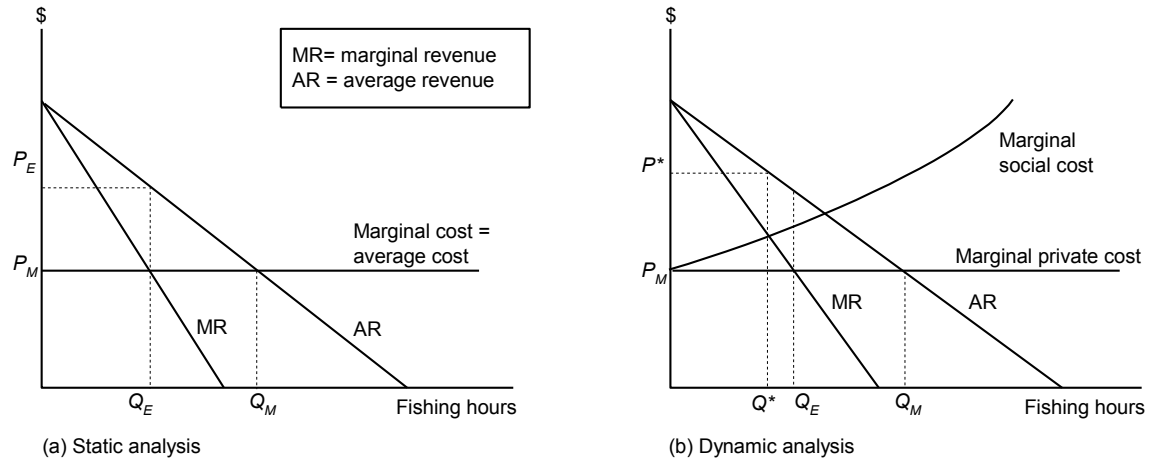


Figure 13.4 Efficient exploitation of a fishery

to an efficient tax, and the revenue would again accrue to government. From an efficiency perspective, the method of allocation is not important; individual quotas can be given away. What matters is that the quotas can be traded and the efficient price established. This ensures that the quotas are owned by the most efficient fishing groups, who will achieve the total quota at least cost.[12] Several countries, including Australia and New Zealand, have adopted quota controls on the fishing industry.

The outcomes of the tax and quota instruments are similar. The quantity of fish caught is reduced by similar amounts. Both policies are more efficient than regulatory policies that control the fishing process such as shortening the fishing season or limiting the size of fishing nets. These regulatory policies can reduce the quantity of fish caught, but they require the fishing industry to adopt high-cost and inefficient fishing practices.

Conclusions

In some cases, commonly owned resources are used harmoniously and efficiently. In 2009 Elinor Ostrom was awarded the Nobel Prize in economics for her case studies showing that many communities have adopted rules and enforcement mechanisms that enable them to manage common-pool resources cooperatively and sustainably. For theoretical support, Ostrom has drawn on game theory which shows that participants tend to be more cooperative in repeated games (Mailath and Samuelson, 2006). Evidently group ownership and management of common property resources can work in some cases. However numerous exceptions can also be found in under-managed game parks in Africa, in the forests of Amazonia and in many fisheries around the world.

When common property resources are over-exploited, government intervention is often desirable. This generally means establishing either property rights, or at least entitlements, to access the resources. When property rights are fully established there are no externalities. The owners have an incentive to use the resource efficiently, by producing up to the point where marginal revenue equals marginal cost, and by conserving the capital value of the resource when the discounted value of future output exceeds present output value. This may mean privatisation of the resource—converting common property into individual title, as occurred with the widespread property enclosures in England in the 17th century. However, dividing large common property resources into private ownership may not be practical or equitable. Therefore, the English enclosure movement allegedly made the workers worse off because the capitalists appropriated more than the gains that the enclosures generated.

When privatisation is not appropriate, the government or the community needs to establish clear rules for access to the resource. Tradable permits provide an efficient way of controlling use of a common property resource. Taxes on output are also an efficient way to control output. Other policies are possible but are generally less efficient than these market-based instruments.

However, as has been stressed throughout this book, market failure is a necessary but not sufficient condition for government intervention for efficiency objectives. Policies are only as good as their implementation. And most policies have transaction costs. These factors need to be considered in assessing government policies for externalities as for other situations.

[12] Of course the method of allocating quotas has significant distributional consequences.

Summary

- An externality occurs when an activity creates a cost or benefit without compensation to an external party that is not actively involved in the activity. Such an externality occurs because of an absence of property rights.
- Externalities almost always represent an inefficient use of resources.
- Resources are allocated efficiently when the marginal social benefit of an action equals the marginal social cost inclusive of any third-party effects. However, optimal resource allocation may also depend on equity considerations.
- When only a few parties are involved, private exchange can deal with externalities and produce efficient use of resources even in the absence of property rights. Private exchange does not produce an efficient outcome when there are many affected parties, high transaction costs and incentives to free ride.
- The two main kinds of policy instruments for dealing with externalities are market-based instruments (MBIs) and regulations.
- MBIs are based on the concept of creating and monetising property entitlements. The main MBIs are corrective taxes (or subsidies) and quantity controls. The efficient tax is equivalent to the marginal damage cost at the efficient level of output. Well-designed MBIs achieve an efficient outcome at least cost.
- Regulations may control activities, production or unwanted externalities themselves. Regulations are practical. However, they are usually a relatively costly way to achieve any given objective.
- Common property resources are resources to which anyone has access. Such resources are often over-exploited. Policies again include establishing property rights, setting prices to reflect marginal damage costs and quota systems.
- However, here as elsewhere, market failure is a necessary but not sufficient condition for government intervention for efficiency reasons. Policies require careful construction and may have significant transaction costs.

Questions

1. Why may the allocation of property rights not affect the allocation of resources even when there appear to be significant externality effects?
2. What should determine the efficient level of output in an industry that produces a polluting product? How would deadweight losses be estimated when this output is not achieved? Explain how distortions in other markets might affect the efficient outcome in the original market.
3. Explain why the optimal level of pollution is generally not equal to zero.
4. Why are market-based instruments generally more efficient policies for dealing with externalities than are regulations?
5. Should governments tax the use of four-wheel-drive cars in urban areas? If so, why and how?
6. Why is a tax on an externality at the existing marginal damage cost likely to be inefficient?
7. The marginal cost of a plant producing chemical materials is represented by the equation $MC = 100 + 25Q$. However, including the costs of wastes, the marginal social cost is $MSC = 100 + 50Q$, where Q is units of output. Assume that in the competitive market the chemical product can be sold for \$250. How much will the firm produce? What is the socially efficient level of output?
8. Suppose that a firm operating in a competitive market produces a negative externality by way of toxic emissions.

   The marginal cost of production is represented by $MC = 5Q$. Also each unit of output produces an externality equal to $2Q$. The firm faces the demand curve of $Q^d = 420 - P$. Determine the following:

   i. The equilibrium level of output and equilibrium price, where the firm ignores the effects of the externality.

   ii. The tax to be levied if government chooses to impose a per unit tax on each unit of output of the firm to account for the pollution costs involved in production.
9. Two firms are located next to a river: an upstream chemical firm and a downstream hotel. The chemical firm obtains water from the river and disposes of waste into it. The hotel gains from the amenity of the river. The net income of the chemical firm is $Y_F = 32 + 12C - 0.25C^2$, where C is the amount of pollution pumped into the river. The net income of the hotel is $Y_H = 200 - 8C$.

   i. Derive the schedules for the marginal benefit of pollution to the chemical firm and the marginal cost to the hotel. What is the efficient level of pollution?

ii. If the hotel owns the property rights to the river and no bargaining takes place, how much pollution occurs?

iii. If the two parties bargain, does the chemical firm have any incentive to make an income transfer to the hotel?

If so, what is the minimum it would need to offer? Will it enjoy any surplus after this transfer?

iv. Now suppose that the chemical firm could acquire a technology to prevent any pollution and would obtain any surplus from any bargain with the hotel to reduce pollution. How much would the chemical firm be willing to pay for the technology?

10. The aggregate daily demand function for visits to an aquarium is $P = 100 - Q/10$.

i. What would be the total number of visits per day with free admission?

ii. If the marginal cost (MC) for serving each visitor is \$15, how much should the aquarium charge and what would be the number of visits?

iii. Suppose that when visits per day exceed 500, the aquarium gets crowded, visitors block the views of each other and maintenance cost rises. The MC rises by \$0.10 per additional visitor over 500. For example, for the 510th visitor, MC = \$15 + (10× \$0.10) = \$16. In this case, what is the equilibrium number of visits and how much should the aquarium charge?

11. Why, for any given target reduction in CO_2 emissions, is government direct action to invest in green technology likely to cost the average Australian household more than a carbon tax?

12. What is the difference between a common property resource and a public good?

13. Does the method of allocating permits to use a common property resource, such as a fishery, affect the economic outcome?

Further Reading

Bovenberg, A.L. and Goulder, L.H. (2002) 'Environmental Taxation and Regulation', pp. 1474–1545 in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Vol. 3, Elsevier Science, Amsterdam.

Clarke, H. (2011) 'Some basic economics of carbon taxes', *Australian Economic Review*, 44, 123–136.

Coase, R. (1960) 'The problem of social cost', *Journal of Law and Economics*, 3, 1–44.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed. Chapters 5 and 6, Worth Publishers, New York.

McWhinnie, S.F. (2009) 'The tragedy of the commons in international fisheries: An empirical examination', *Journal of Environmental Economics and Management*, 57, 321–333.

Ostrom, E. (2000), 'Collective action and the evolution of social norms', *Journal of Economic Perspectives*, 14, Summer, 137–158.

Rosen, H and Gayer, T. (2014) *Public Finance*, 10th ed., Chapter 5, McGraw-Hill, New York.

Tietenberg, T. (2006) *Environmental and Natural Resource Economics*, 6th ed., Chapter 4, Addison Wesley, New York.

Chapter

14

Competition and Industry Policy

No one can argue that a monopolist is impelled by an 'invisible hand' to serve the public interest.

R.H.Tawney

Costs of Imperfect Competition ♦ Regulating Industry Structure ♦ Regulating Competitive Conduct ♦ Regulating Dominant Firms ♦ Competition Policy in Australia ♦ Benefits of Competition Policy ♦ Concluding Observations

Most real world markets depart from perfect competition. Some markets are naturally monopolistic because of technical innovation or economies of scale. Other markets contain only a few firms. Many others are monopolistically competitive—they contain several competing firms selling slightly differentiated products. In all such markets, firms have some market power. The firm can set the price at which it sells its' goods. It does not have to sell strictly at the price set by a competitive market. Thus, imperfect competition (market power) is a common feature of most economies.[1]

Market power can have significant efficiency and equity costs, including high prices, restrictions on output, inefficiency in production and weak innovation. Thus, the policy challenge is to create and sustain an efficient amount of competition in markets when this is often not the natural state of markets. Where competition is not feasible because of the natural monopoly features of the market, efficient regulation of the monopoly supplier is required.

In this chapter, we start by describing the costs of imperfect competition in product markets. Reducing these costs is the basic rationale for competition policy. The next three sections discuss the core competition policy strategies: regulating industry structure, industry conduct and performance. Industry structure regulation is designed to create a competitive industry, especially to deal with anti-competitive acquisitions and mergers. Conduct policies deal with the general competitive behaviour of firms, including prevention of anti-competitive agreements. Performance policies regulate dominant (private or public) firms to prevent abuse of market power. We then describe competition policy in Australia and the benefits of competition policy in Australia and elsewhere.

Finally, we briefly draw attention to two important, but unrelated, matters. One is the increasingly important role of digital companies. The other is the role of government in promoting productivity described by the Productivity Commission (2017).

[1] This chapter discusses market power in product markets. Chapter 27 discusses market power in labour markets.

Costs of Imperfect Competition

The economic costs of imperfect competition may include a deadweight loss due to reduced output, inefficient production, a lack of innovation and rent-seeking costs. To simplify the exposition, we illustrate the costs associated with a monopoly. Similar costs arise in other forms of imperfect competition, though usually to a lesser degree.

Deadweight losses from reduced output. Figure 14.1a shows a monopoly market, with the marginal revenue (MR) curve below the demand curve. For ease of illustration, the marginal cost (MC) of supply is assumed constant. A monopolist maximises profit by selling Q_2 output, where MR = MC, and setting price at P_2, rather than at $MC = P_1$. Consumption falls from the competitive level Q_1 to Q_2. This causes a deadweight loss equal to area ABC.[2] If the demand curve is linear the deadweight loss (DWL) is:

$$DWL = 0.5\,(Q_1 - Q_2)\,(P_2 - P_1) \tag{14.1}$$

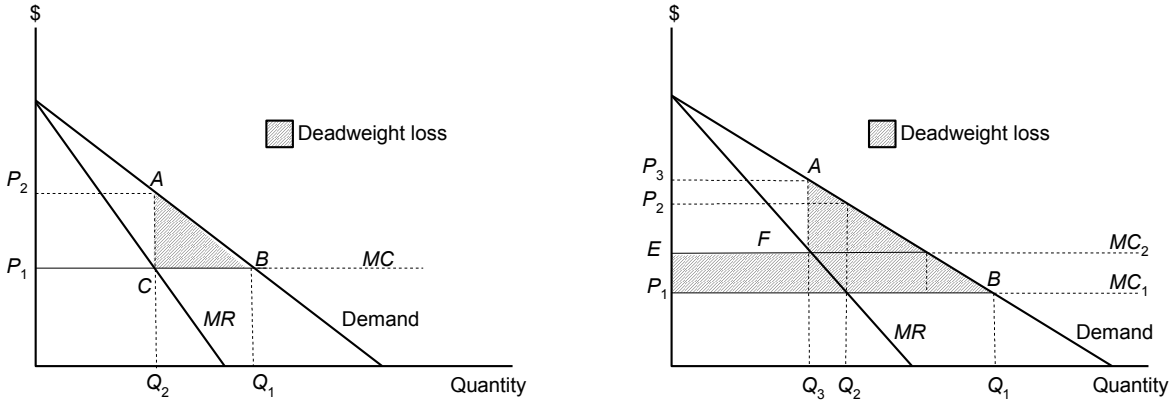
The DWL rises with the size of the price mark-up and the fall in consumption. In this analysis, area P_2ACP_1 is treated as a transfer payment, with losses to consumers offset by gains to the producer.

Equation 14.1 can be rewritten as:

$$DWL = 0.5\eta_d P_1 Q_1 m^2 \tag{14.2}$$

where η_d is the price elasticity of demand and m is the mark-up percentage $(P_2–P_1)/P_1$.[3] Note the role of the price elasticity of demand. The higher the elasticity of demand the greater is the DWL. Equation 14.2 is important not only for estimating the costs of market power but also for developing efficient pricing and tax strategies (see Chapters 17 and 27).

Excessive production costs. In an oft-cited article, Leibenstein (1966) argued that monopolists exploit freedom from competition by being x-inefficient. X-inefficiency occurs when a firm fails to produce its output with the minimum use of inputs.[4]



(a) High price and low output

(b) High price, low output and high costs

Figure 14.1 Costs of monopoly

[2] This triangle is often called a Harberger triangle. Harberger (1954) popularised the idea of estimating the costs or benefits associated with consumer and producer surplus triangles.

[3] The derivation of Equation 14.2 from Equation 14.1 is explained in Chapter 27.

[4] A firm is efficient if it produces output at least cost given the technology available. In perfect competition, all firms are efficient because x-inefficient firms are driven out of business.

In a study of 670 firms in the UK, Nickell (1996) found that growth in total factor productivity was positively related to the competitiveness of the industry. Figure 14.1b incorporates x-inefficiency costs, with marginal cost rising from MC_1 to MC_2. The profit-maximising price would rise from P_1 to P_3 and output fall from Q_1 to Q_3. The DWL is now the shaded area $EFABP_1$. A major part of this loss is the increase in operating costs.

Innovation costs. As we saw in Chapter 5, economists differ on the impacts of industry concentration on the rate of innovation. Schumpeter argued that some monopoly is required to obtain innovation. Other economists contend that investment rises with certainty of profit, which increases in more concentrated industries. Scherer and Ross (1990) reviewed the evidence and found that, on balance, competition is more important for innovation. The recent history of the Australian telecommunications industry after deregulation supports this view. The welfare loss from a lack of innovation is the difference between what consumers would pay for a new good and the cost of producing it.

Rent-seeking costs. These are resources spent to gain economic rent. **Economic rent** is a payment to a factor of production over its opportunity cost (its value in alternative employment). In the discussion above, the price increases borne by consumers (from P_1 to P_2 in Figure 14.1a) were treated as a transfer payment because they were offset by the gain to the firm. However, the area P_2ACP_1 in Figure 14.1a represents economic rent. Firms often spend resources to obtain such rents, especially when they can be gained by government regulation that reduces competition. Expenditure on gaining rents is a deadweight loss. Examples of rent-seeking firms in Australia include major mining companies, firms holding commercial TV licences, owners of taxi licences and pharmacies, just to name a few.

Rent seeking
Trying to persuade government to change the rules in such a way as to create economic rent

Economic rents
Payments to factors of production in excess of

Rent-seeking costs
Resources used in rent seeking

Estimated costs of monopoly and rent seeking. These estimates vary greatly. In a classic study drawing on Equation 14.2, Harberger (1954) estimated that the DWL from *output restrictions* in monopoly industries in the United States in the 1920s was equivalent to only 0.1 per cent of GDP. Cowling and Mueller (1978) recalculated the loss from output restrictions drawing on data for 734 US firms in the 1960s. Making different estimates on demand elasticities and price mark-ups, they calculated that the DWL was about 4 per cent of US GDP. Following a similar approach, Dixon *et al*. (2001) estimated that the welfare loss from monopoly in Australian manufacturing alone was equivalent to 1.4 per cent of GDP.

Other costs of monopoly may also be large. Kamerschen (1966) estimated that excess production costs were equivalent to 6 per cent of US GDP. Also, rent-seeking costs may be substantial. If the probability of achieving an economic rent is high, risk-neutral profit-maximising firms may spend a large amount of the potential rent to achieve or retain it. Drawing on relevant demand elasticities for regulated industries in the United States, Posner (1975) estimated that economic rent represented 10–30 per cent of the value of sales in some industries. This indicates the potentially high cost of rent-seeking expenditures.

Equity. We saw above that the area P_2ACP_1 in Figure 14.1a is treated as a transfer payment, with losses to consumers offset by gains to the producer (usually shareholders). However, government may view transfers from consumers to shareholders as inequitable and undesirable and regard this also as a significant social cost.

Potential competition
Competition from potential rivals not currently in the relevant market

The role of potential competition. Some economists (e.g. Baumol *et al*., 1982) argue that, providing a market is contestable, potential competition constrains monopoly behaviour. A market is contestable if there is free entry and exit. In such a market, to deter entry the incumbent monopolist reduces the price of a good towards its long-run average cost. The critical requirement for free entry and exit is low sunk costs—firms can expect to recoup their

sunk costs or be able to liquidate assets or transfer them to another market without cost. In practice, there is often some unrecoverable sunk cost and completely free entry and exit is rare. A DWL remains when price exceeds marginal cost. But potential competition lowers the price mark-up and the associated loss. This implies that governments should encourage contestability and not place artificial restrictions on entry into a market.

Possible efficiency arguments against competition

Despite the costs of imperfect competition, governments often protect industry from competition. This protection may take the form of import tariff or quota protection, restriction on entry to the industry, direct financial assistance or tax concessions. The Productivity Commission (2011a) estimated that measured gross assistance totalled $17.2 billion in 2008–09, including $9.5 billion in tariff assistance, $3.7 billion in budgetary outlays and $4.0 billion in tax concessions. This was equivalent to about 1.5 per cent of GDP. Industry support programs in Europe total about 2 per cent of GDP.

Clearly, government may assist industry for equity objectives. Can industry assistance also be justified on efficiency grounds? We discuss below four possible efficiency arguments for restricting competition or competitive behaviour.

Predatory pricing
Setting low prices and sacrificing short-term profits in order to reduce competition in the long run

First, competition in the short run may destroy competition in the long run. An example is **predatory pricing**. This occurs when a firm sells goods at low prices and sacrifices short-term profits with the aim of reducing the capacity of competitor firms and increasing own long-run market share and profits. Competition policy generally prohibits predatory pricing.

A second possible case for industry protection arises from the need to encourage innovation. Patents reward innovation by allowing a firm or individual control over new technology for a limited period. Without such protection entrepreneurs may be unable to recoup the costs of research and development (R&D), there would be under-investment in R&D and innovation would fall.

Thirdly, governments may limit entry into an industry, or assist industry self-regulation, to protect quality of service. This applies to providers of professional services, such as doctors, pharmacists, lawyers, accountants and surveyors, as well as to tradespersons like electrician and plumbers. The premise is that consumers cannot be trusted to make well-informed decisions, so there is information failure. We discuss these information issues in Chapter 15.

Fourthly, there has long been concern that start-up firms may not survive against competition and should be protected until they mature. The implicit market failure here is either that capital markets place too high a price on risk. Japan's protection of its steel, coal, machinery and shipbuilding industries is sometimes cited as an example of effective state-coordinated development that would not have occurred without official support. A current version of this view is that firms may need to combine in domestic markets to achieve the economies of scale that will enable them to compete in international markets.

However, it is not clear that there are such significant market failures. Selective assistance policies generally distort the allocation of resources and may reduce rather than increase the aggregate value of output. As Adam Smith observed:

> Every system which endeavours … to draw towards a particular species of industry a greater share of the capital of society than would naturally go to it … diminishes, instead of increasing, the real value of the annual produce of its land and labour.[5]

If assistance is warranted, ideally government would identify the marginal subsidy required to provide the desired increase in output. But this is hard to identify. Often all output is subsidised, not just marginal output. Also, government has limited expertise at deciding when private capital markets are under-funding a private business. Public support for shipbuilding

[5] Smith (1776), *The Wealth of Nations*, p. 651 in E. Cannan (ed.) (1937).

in Germany and Sweden and aerospace industries in Italy and the UK was unsuccessful. In recent years, East Asian experience (stagnation in Japan, the fall of the Korean chaebols and the failure of Indonesian crony capitalism) has also discredited industry assistance policies.

In summary, competition is generally desirable. An important exception occurs when there are natural economies of scale or scope, which reduce the average cost of production. There may also be situations when competitive behaviour aims to reduce competition, innovation requires protecting the gains for the innovator or consumers need to be protected from quack service providers. Other forms of industry protection or assistance may be warranted on distributional grounds, but rarely on efficiency grounds.

Regulating Industry Structure

The aim of industry structure policies (also described as anti-trust policies) is to create and maintain competitive industry structures. This is a structure in which no single firm or small group of firms has a dominant market share that allows it to exercise market power, mark up prices significantly above long-run average costs or be x-inefficient. Dominant market share may arise via organic growth or by acquisitions or mergers. Anti-trust policies generally aim to prevent acquisitions and mergers that could result in excessive concentration in a market (or in Australia to lessen competition in a defined market). It is harder to break up an existing dominant company. The US government succeeded in breaking up the dominant AT&T (telephone) company into smaller companies in 1984. However, the Clinton Administration failed to break up the Microsoft Corporation in the 1990s.

The general approach to dealing with mergers and acquisitions is to determine what constitutes a market and the existing level of competition in it (using a concentration metric) and to estimate the effects of the proposed merger or acquisition on market concentration or competition and hence output and prices.

Defining a market. Estimates of market shares require a definition of a market and estimates of total market sales. However, there is no definitive measure of a market. A market consists of similar products that can be readily substituted one for another. Where buyers or sellers can readily shift between products or areas, these products and areas are part of one market. Formally, substitution is measured by cross-price elasticities. As Stigler (1955, p. 4) wrote: 'all products or enterprises with large long-run cross-elasticities of either supply or demand should be combined into a single industry'. However, deciding whether any cross-elasticity magnitude is large or small is a matter for judgement.

The determination of the market is crucial because the larger the market the less will be the market dominance by one or a few firms. If a market is narrowly defined, it is more likely to be dominated by one or a few firms. On the other hand, the demand for the industry's products may be elastic and the products may be subject to competition from other markets. An important related issue is the appropriate territory. Increasingly, regulators must decide whether markets are defined by a single country, a group of countries or even the world.

As an example of the product dimension problem, in determining whether United Bananas was a dominant company the European court had to decide whether bananas were distinguishable from other fruits and therefore a separate market or part of the overall fruit market in Europe. The Court ruled that this depended on the extent to which bananas were 'interchangeable' with other fruit. However, this is not a precise concept. The concept of interchangeability is operational only if the cross-price elasticity of demand can be estimated. If the demand for bananas falls significantly when the price of apples falls, the two goods are substitutable and hence part of one market.

An Australian illustration of the difficulty of defining markets was the decision of the Federal Court to allow the wholesaler Metcash to take over the loss-making Franklins'

supermarkets. The ACCC opposed this takeover because it would reduce competition in the wholesale/distribution market. But the Federal Court found that the major constraints on grocery competition were the large market shares of Woolworths and Coles in the retail market and that the takeover could increase competition in this market rather than decrease it.

Determining market dominance with concentration ratios. Economists typically measure market power by the excess of price over marginal cost. However, marginal cost is hard to measure. Therefore, other measures of market power are required. Market shares are a simple indicator of the market power of single firms.

Concentration ratio
The proportion of a market shared by a small number of large firms

Concentration ratios measure the proportion of industry sales accounted for by a small number of large firms (typically three to five firms). However, a simple concentration ratio gives no special weight to the largest firm. Suppose that in two industries, *A* and *B*, four firms share 80 per cent of an industry's sales. However, in *A* the largest firm has 60 per cent of the sales and the other three firms share 20 per cent, whereas in *B* the four largest firms each have 20 per cent of sales. Most people would consider that the concentration was greater higher and potentially more damaging in *A* than in *B*.

The Herfindahl–Hirschmann Index (HHI) gives greater weight to firms with large shares. The HHI is defined as:

$$HHI = (100s_1)^2 + (100s_2)^2 + \ldots + (100s_n)^2 \qquad (14.3)$$

where *s* denotes a firm's share of industry sales and there are 1…*n* firms. Regulators in Australia, the European Union and the United States use the HHI as an indicator of whether a merger or acquisition may be allowed or should be examined. Drawing on Yang and Pickford (2011), the following post-merger shares are generally regarded as acceptable:

- Australia: HHI < 2000, or HHI > 2000 and ΔHHI < 50.
- European Union: HHI < 1000, or HHI 1000–2000 and ΔHHI < 250, or HHI > 2000 and ΔHHI < 150, or *M* < 10 per cent (where *M* is market share of the merged entity).
- United States: HHI < 1000, or HHI 1000–1800 and ΔHHI < 100, or HHI > 1800 and ΔHHI < 50 and *M* < 35 per cent.

Conversely, mergers that lead to higher HHI scores would be examined. However, the Australian Competition and Consumer Commission (ACCC, 2011) indicates that these scores are only guides. Clearly, rulings based only on measures of market share would be arbitrary. High market share may not indicate inelastic demand for a firm's product (which is the basis of its power to mark up prices). Concentration measures indicate possible costs but ideally the costs and benefits of concentration should be estimated on their merits in each case.

Estimating the effects of increased concentration. Ideally the regulator would assess the effect of increased concentration on output, prices, x-inefficiency costs and innovation as appropriate. This implies assessing economic performance with and without the acquisition or merger. This is not easy. In the Metcash takeover of Franklins, the ACCC argued that if the takeover were disallowed, another company would take over the loss-making Franklin stores. The Federal Court did not agree that this could be assumed to be likely to occur. Firms involved in mergers and acquisitions invariably argue that there will be significant cost savings from synergies and economies of scale. However, it is important to identify merger-specific efficiencies. Firms may achieve cost savings by contracting out low-volume in-house work or by economies of scale that could be achieved by internal growth. Nevertheless, regulations based on estimates of the costs and benefits of acquisitions and mergers can be expected to produce a more efficient outcome than determinations based on arbitrary concentration ratios.

In practice, applications of competition policy vary. In the United States any lessening of competition is generally viewed as undesirable regardless of possible counter-arguments. In Europe, the main merger test is whether a merger would significantly impede effective competition. However, Vickers (2005) notes that the European Community Courts in Luxembourg are now requiring more economic rigour in merger analysis.

In Australia, the *Competition and Consumer Act 2010* provides the legal basis for determining acquisitions and mergers. Critically, Section 50 prohibits acquisitions and mergers, not on the grounds of concentration ratios, but rather when they would be likely to substantially lessen competition in the market. However, as shown in Box 14.1, the ACCC is required to consider many facets of competition and, in effect, to make a determination on the perceived economic costs and benefits of the proposed acquisition or merger.

Regulating Competitive Conduct

Despite anti-trust policies, concentrated industries occur in most countries. Collusion is an ever-present danger. Even without high concentration, some firms may have significant market power and use this power to restrict competition. Such restrictions include collusion over prices or market shares, price discrimination of various kinds, commodity bundling and vertical restraints on buyers or sellers.

Accordingly, most OECD countries regulate the conduct of firms by prohibiting collusive behaviour and actions that restrict competition. The following are some of the major conduct issues and some policy responses.

Box 14.1 Regulation of acquisitions and mergers in Australia

In Australia, the *Competition and Consumer Act 2010* (the Act) governs acquisitions and mergers. The ACCC is responsible for implementing the Act and determines the outcomes, subject to appeals to the Trade Practices Tribunal. As in other countries, definitions of a market and market dominance are key issues.

Section 4E of the Act defines a market as 'a market in Australia and, when used in relation to any goods and services, includes a market for those goods or services and other goods or services that are substitutable for, or otherwise competitive with, the first mentioned goods or services'. This is a broad definition of a market rather than a precise one as the degree of product substitutability or competitiveness is not defined.

Section 50, which governs acquisitions and mergers, prohibits acquisitions and mergers that would result in 'a substantial lessening of competition' in a 'substantial market'. The ACCC (2011) acknowledges that interpreting 'substantial lessening' requires judgements. A substantial market is interpreted to be at least a state-wide market rather than a regional market.

Section 50 also describes the factors that the ACCC must consider in making its determinations. These include actual or potential import competition, barriers to market entry, the level of concentration in the market, the likelihood that acquisition will remove an effective competitor and the likelihood that acquisition will significantly increase prices or profit margins.

In recent years the ACCC has opposed outright some major acquisitions and mergers. It opposed the Australian Stock Exchange takeover of the Sydney Futures' Exchange, the Caltex proposal to take over the retail assets of Mobil Oil and the National Australia Bank's bid to take over AXA Asia Pacific Holdings. However, it allowed Westpac Bank to take over St George Bank, Mayne Nickless to take over Australian Hospital Care, Fosters to acquire Southcorp and AMP to acquire AXA Asia Pacific Holdings' Australian and New Zealand businesses.

In practice, the ACCC opposes only a few acquisitions and mergers outright. In 2010–11 the ACCC considered 377 cases for compliance with section 50 of the Act and decided that 141 required review. Of these, the Commission cleared 110 unconditionally. Of the remaining 31 cases, the Commission publicly opposed three, expressed private concerns about 4, and agreed that 10 could proceed with variations. Fourteen matters were withdrawn.

This does not mean that the ACCC has little influence over acquisitions and mergers. The nature of the proposals that firms put forward are influenced by both the Act and the power of the ACCC under it.

Price fixing
An agreement between firms to set certain prices

Price fixing agreements by firms generally raise prices and restrain competition. Most countries prohibit price fixing by cartels. In 2007, Australia's largest packaging company (the private Visy corporation) was fined $36 million for fixing prices with its major duopoly competitor, Amcor, which confessed to the scheme. However, it can be difficult to prove price fixing. Firms may adopt similar prices and price changes due to collusion or because they face similar market forces and similar price changes make economic sense. In 2007, the ACCC failed to prove that owners of major chains of petrol stations were colluding on petrol prices although the stations regularly changed their petrol prices by similar amounts and at similar times over the week. Another grey area is fee recommendations by professional organisations to their members. Regulators usually focus on cases of overt, explicit, price fixing because these cases are easier to prove and punish.

Predatory pricing is the practice of setting low prices that sacrifice short-term profits with the aim of driving existing competitors out of a market and deterring new firms from entering it. In Australia, low prices are defined simply as 'below the relevant cost'. In 2010, Cabcharge was fined $3 million for supplying 6178 meters to taxis either free or for $100 (to maintain its monopoly on the taxi EFTPOS system) when the direct costs of acquisition were $250 per unit. Putting excess supply on a market has a similar effect. Qantas was alleged to have done this on some inter-city Australian routes to lock out competition from Virgin Airlines. Predatory pricing is prohibited in most OECD countries. On the other hand, pricing below long-run average cost but above average variable cost is generally not regarded as predatory.

Again, the regulator faces various problems. First, the regulator must define a 'relevant cost'. This could be defined as average variable cost. But there are still awkward questions. Are overheads associated with labour employment variable costs? How should joint costs be allocated? Second, the regulator rarely has independent cost data and must rely on data provided by the regulated firm. Third, the regulator requires estimates of marginal revenue. The ACCC could not show that the building products company, Boral, exercised predatory pricing in that industry. Under Australian law, the regulator must show not only that a firm is underpricing, but also that it is aiming to harm a competitor. Intention is hard to prove.

The principle of unrestricted dealing is critical to a competitive market. This has several implications. One is the need for **third-party access to essential facilities**, such as telecommunication, energy or transport networks. An essential facility has two main characteristics: it is a natural monopoly and it produces a good that is an essential input to the production of another good or service. Vertically integrated firms often have a monopoly on some aspect of transmission and can exclude other firms by setting excessive prices for access. This makes access unprofitable and maintains the down-market share of the vertically integrated firm. In Australia, the (then) major telco (Telstra) was charged frequently with setting excessive prices for access to its copper lines. Again, there is a major question as to what would represent a fair price. Economic efficiency requires that competitors should pay no more for a service than do in-house businesses. However, a vertically integrated firm may argue that it is entitled to recoup research and development and capital expenses and to a return on risk capital. Although the obligation to deal on fair terms applies principally to vertically integrated firms with monopoly control of infrastructure networks, the principle of unrestricted dealing has general application. For example, the European Union fined British Sugar three million euros for charging excessively high prices for bulk sugar to wholesalers to protect its own retail operations.

The principle of unrestricted dealing also implies **avoidance of exclusive dealing with external parties.** Exclusive dealing occurs when one party imposes restrictions on another party's freedom to choose with whom it deals. For example, exclusive purchasing or distribution arrangements between breweries and hotels would be prohibited. In 2010,

Cabcharge was fined $11 million for refusing to deal with a competing supplier (Travel Tab) to allow Cabcharge payments to be processed through a rival EFTPOS terminal. The European Commission has gone further than this and argued successfully in the courts that discounts and rebates designed to tie in dealers or agents to a particular supplier are unlawful. For example, by giving loyalty-inducing quantity rebates, Michelin was found to prevent dealers from being able to select their suppliers freely.

Resale price maintenance is another major restrictive practice. This occurs when firms fix the minimum prices at which their product can be sold by distributors. This restricts competition in both wholesale and retail markets. Most developed countries prohibit resale price maintenance. This regulation usually applies to all firms, not just dominant ones.

Australian practice. As shown in Box 14.2, under the *Competition and Consumer Act 2010* Australia has a comprehensive set of regulations designed to prevent anti-competitive behaviour including by government. Section 2 binds all Commonwealth and state businesses to the conditions of Part IV, which contains the key prohibitions on anti-competitive behaviour.

Box 14.2 Australian regulations prohibiting anti-competitive behaviour

The *Competition and Consumer Act 2010* (the Act) contains many regulations that promote competition and prohibit anti-competitive behaviour.

Part III of the Act deals with provision of third-party access to essential facilities such as electricity grids or natural gas pipelines. The aim is to ensure that third parties can compete in upstream or downstream markets. Under Section 44, the ACCC has the power to require vertically integrated monopolies to provide competitors with access to their facilities at competitive rates. For example, Telstra must provide competitors with access to its local telephone network at rates that allow other firms to compete in the local telephone market. In 2005, following ACCC action, Telstra was forced to lower its wholesale prices for broadband customers.

Part IV deals with anti-competitive practices. The following are some of the most important regulations.

Section 45 prohibits anti-competitive agreements, including agreements that involve market sharing, restrictions on the supply of goods and price fixing. Under this section the ACCC has imposed fines of several million dollars on major transport companies for colluding on freight contracts, on suppliers of concrete for colluding on concrete prices in Queensland, and on large civil engineering companies for colluding on tenders for major public projects in New South Wales.

Section 46 deals with misuse of market power. It prohibits a business with a substantial degree of market power from using its market power to eliminate or substantially damage a competitor or to prevent a firm entering the market or engaging in competitive conduct in a relevant market. This outlaws predatory pricing, which is described as supplying goods for a sustained period at a price that is less than the relevant cost to the corporation of supplying the good. Under this section the ACCC found that BHP breached the Act when it refused to sell important products to a competitor, Queensland Wire Industries. However, the section does not prevent what small businesses often view as oppressive behaviour by a dominant firm, for example the owner of a large shopping mall may set local monopoly rents and oppressive leasing conditions.

Section 47 prohibits anti-competitive exclusive dealing, which has the aim of substantially lessening competition in a relevant market. Thus, it is prohibited to supply goods on the condition that the purchaser will not acquire goods from a competitor of the supplier or that the purchaser will not resupply the goods in a specified place or classes of place.

Section 48 deals with resale price maintenance. Suppliers may recommend a resale price for goods providing that is only a recommendation and the supplier takes no action to influence the reseller not to sell or resupply below that price. However, the Act prohibits suppliers from specifying a minimum price below which goods or services may not be resold or advertised for sale. It also prohibits inducements to resellers not to discount, for example by giving special deals to resellers who agree not to discount. Following action by the ACCC, the skin-care company, Dermatologica, was fined $0.25 million for attempting to force two retailers to stop discounting.

Regulating Dominant Firms

When a firm dominates a market, especially an important one such as telecommunications, energy or transport, general rules of conduct may not be sufficient to protect the public from excessive prices and x-inefficiency. The public interest may require direct regulation of the dominant firm and of the prices it sets. Public ownership is another option. Public ownership enables government to control directly the key output, investment and pricing decisions and provides the government with more information. Here we discuss methods to regulate dominant private firms. But, as discussed in Chapters 16 and 17, similar efficiency and pricing principles apply to a publicly owned firm.

Given that the capacity to charge excessive prices is the main cause of economic inefficiency, price regulation is the logical policy response. However, price regulation is a complex task. The regulator needs to determine the firm's costs, a fair rate of return on capital and the allocation of costs over multiple services. Even an apparently homogeneous product like electricity is characterised by the time of the day or week when it is supplied.

Rate-of-return price regulation
Setting regulated prices to achieve an expected rate of return

Rate-of-return price regulation is the basis of price regulation. Typically, the regulator reviews the total operating and capital costs of the regulated enterprise to establish a revenue requirement. Sometimes the regulator conducts a supplementary exercise to determine whether the expenses are efficient. The aim is to ensure that regulated revenue is not inflated by inefficient expenditures. In some cases, the regulator estimates the costs of separate services by a cost allocation method (see Chapter 16). The regulator then sets the prices for each service. As we will see in Chapter 17, in the short run efficient prices should reflect short-run marginal costs (SRMC). However, SRMC pricing may not produce enough revenue to provide a reasonable return on capital employed. If this occurs, the regulated firm will not be able to attract capital and there will be under-investment in key services. Therefore, most regulators aim to set prices based on long-run marginal cost inclusive of capital costs.

The process can be represented by the following accounting equation:[6]

$$\sum_{i=1}^{n} P_i Q_i = \Sigma E + rK \tag{14.4}$$

where P and Q are the price and quantity of the ith service respectively, the firm supplies $i = 1...n$ services, E is the firm's operating expenses, K is the value of capital employed and r is the rate of return allowed on capital. To determine the appropriate price for each service, the regulator must forecast Q and E and determine appropriate values for K and r.

To forecast operating expenses, the regulator can draw on recent audited expenses. However, the regulator must allow for normal cost inflation. Also, expenses depend on the quantity of services provided. If the quantity of output is considered either too limited or excessive, the regulator will need a cost function that relates costs to quantity supplied. Also, the regulator may consider that expenses include some x-inefficiency costs which should be discounted going forward.

The main methods for determining the capital value employed are: historic cost after depreciation, reproduction cost, replacement cost or market value. Each of these concepts requires some judgement. The first method requires estimates of depreciation. A further major drawback of this method is that, with inflation, the depreciated historic cost may be far below replacement cost or market value. Therefore, revenues based on this measure may be too low to warrant future investment. Reproduction cost is the cost of reproducing the existing plant at current prices. Again, this may understate the needs of the business as it assumes old technology. Replacement cost is the cost of replacing the capacity with plants embodying

[6] The discussion in this section draws on Viscusi *et al.* (2005, Chapter 12).

current technology. This ensures that the revenues will allow the firm to fund new investment. However, like the previous two methods, it assumes that all past investment in capacity was justified. This problem does not arise with the market value of the productive assets, which is a forward-looking valuation method. But there is a circularity problem with market valuations of capital stock: the market value depends on the prices that the regulator sets.

Having established capital value, a regulator attempts to determine a rate of return that the market would require for a business with its risk characteristics. Essentially, the rate of return must reflect the competitive cost of capital. The pre-tax weighted average cost of capital (WACC) is given simply by Equation 14.5, where D is the pre-tax cost of debt, E is the cost of pre-tax equity and w is the proportion of debt finance.

$$WACC = wD + (1–w)E \tag{14.5}$$

However, the literature on how to estimate this seemingly simple equation is vast.[7] At a simple level, following the Modigliani-Miller theorem on capital structure, in a world without taxes, the WACC is not affected by the proportion of debt finance (the value of w). Although debt is lower cost than equity which bears more risk, taking on more debt increases the risk and cost of equity.[8] While an appropriate cost of debt can be inferred from the market, there is debate about whether to use historic or projected debt rates. The cost of equity is typically based on the capital asset pricing model (known as CAPM, see Equation 14.6). This shows the expected return on equity (E) as equal to the risk-free return (R_f) plus a risk premium which allows for undiversifiable market risk. The risk premium is expressed in terms of the expected return on a diversified market portfolio (R_m) and a factor β which is the covariance of the asset returns with the market return. The average value of β across all risk assets is 1.0 but a given asset may have a lower β where returns are less volatile than average and/or returns are regarded as only partly correlated with general market conditions.

$$E = R_f + \beta(R_m – R_f) \tag{14.6}$$

Finally, the regulator sets prices. The regulator set prices for each major service or set a price cap for a group of services (the basket control method). Individual price caps provide greater controls and may be feasible where there are only a few products. However, they are complex to develop and administer. The regulator must determine a fair price for each service that will cover average cost, not just marginal cost, given the required risk-adjusted rate of return. To avoid cross subsidies, information on the allocation of fixed costs and variable overhead costs is required. Alternatively, the regulator may design a two-part tariff system to minimise deadweight losses (see Chapter 17). But this is not practical for a multi-product firm.

Given these complications, regulators often adopt a basket approach. In this case, the regulator specifies the maximum rate at which a weighted price index for a group of products may change. The weights in the index are based on the revenue shares of each product. Thus, a regulator may determine that a firm's weighted price index cannot increase faster than the consumer price index (CPI) or the CPI – X where X is a percentage figure such as 1 or 2 per cent which typically reflects potential productivity gains relative to the rest of the economy. It may also reflect a judgement that a firm's current prices are too high. Basket price caps are simpler to administer than a set of individual prices. They also allow a regulated company to determine its relative prices and to alter prices more flexibly when circumstances change.

---

[7] Useful discussions can be found in reports by the Commonwealth Competitive Neutrality Complaints Office (1998), Australian Energy Regulator (2013) and NSW IPART (2018).

[8] If the interest paid on debt is tax deductible, and the dividends paid to equity are not tax deductible, the pre-tax WACC may fall with the amount of debt used. Dividend imputation partly offsets this effect.

However, the regulator must still judge whether a firm's initial basket of prices is reasonable and whether the basket of prices should decline in real terms at some specified rate because current prices are too high or because of expected productivity gains.

Evidently, rate-of-return price regulation (RoRR) requires a lot of data. Most of this is supplied by the regulated enterprise and the regulator often has incomplete information. RoRR may also weaken incentives to control costs and improve productivity. Under cost-based prices, firms may pass on cost increases in higher prices without breaching the regulated rate of return. Alternatively, RoRR may encourage excess services. Before airlines were deregulated in the United States, they competed with excess services. With the cost increases, the high prices did not breach the regulated rates of return. Firms may also substitute capital for labour when this is efficient and increase their accounting profits (while reducing the rate of return). Generally, firms under RoRR have less incentive to improve productivity if these improvements lead to lower prices rather than increased profits.

Given these problems, there has been much discussion of methods of regulation that would incentivise the regulated firm to be efficient (Weare and Friedmann, 1998). One incentive strategy, sometime called the price cap approach, simply sets maximum prices and allows the regulated firm to retain any profits from productivity gains. However, the effectiveness of the strategy depends on how it is managed. Productivity incentives exist if the regulator holds prices constant in response to the firm's performance but allows prices to change with changes in external factors.

In conclusion. Regulating private firms is complex. The regulator has limited information and must forecast market environment and firm performance. Future events may vary from those forecast, so flexibility is needed. Also, price regulation may simply validate the cost structure of the regulated firm. In a review of Australian regulation, the Productivity Commission (2001a) argued that, due to the complexity of price regulation and the poor information on which it is often based, regulated prices are often more inefficient than those that would occur in an unregulated market. High prices penalise consumers. Low prices discourage investment. Given these challenges, price regulation should focus on major monopolistic markets, especially access to essential infrastructure services, where regulation is most beneficial. And the regulator needs to show that regulation improves outcomes compared with market outcomes. Box 14.3 outlines how firms are regulated in Australia.

Box 14.3 Regulation of dominant firms in Australia

In Australia, Commonwealth and state agencies regulate dominant firms, mainly by price regulation. The regulatory agency is usually an independent agency established by Australian or state legislation.

The ACCC is the most important regulator. The ACCC administers the price regulations clauses in Part III of the *Competition and Consumer Act 2010*, governing essential access to major infrastructure businesses, and the *Prices Surveillance Act 1983*. The ACCC sets price caps for access to telecommunication networks, rail freight and postal services. Other regulatory responsibilities include maintaining competitive safeguards in telecommunications and monitoring major airport, stevedoring and petrol prices.

The Australian Energy Regulator (AER) regulates the wholesale electricity market and is responsible for the economic regulation of the electricity transmission and distribution networks in the national electricity market. The AER is also responsible for regulation of gas transmission and distribution networks and for enforcing the national gas law and national gas rules in all jurisdictions except Western Australia.

Turning to the states and territories, most have regulatory agencies. For example, in NSW the Independent Pricing and Regulatory Tribunal (IPART) regulates the water, gas, electricity and public transport industries. IPART advises the NSW government on maximum prices for monopoly services provided by government agencies. This includes revenues and/or prices of electricity networks and natural gas pricing and third-party access to gas networks. In Victoria, the Essential Services Commission plays a similar role.

Competition Policy in Australia

As we saw in Chapter 2, the Australian economy was traditionally protected and not very competitive. However, since the early 1980s governments from both sides of politics have introduced substantial competition reforms. In 1983–84, the Australian government abolished most controls on currency exchanges, floated the Australian dollar and deregulated domestic interest rates. In the late-1980s the government cut tariffs substantially and introduced competition into the airline and telecommunication markets.

In 1995 the states and territories agreed to participate in the competition reform process by signing with the Australian Government the National Competition Policy (NCP) package of agreements.[9] Under the NCP all governments agreed that there should be no restrictions on competition unless these restrictions could be shown to be in the public interest. The public benefit test requires that non-competitive agreements or behaviour be assessed by benefit–cost analysis. In principle, the reforms would include:

- Structural reforms separating regulatory from commercial functions and natural monopoly elements from contestable elements.
- Competitive neutrality with public agencies competing on the same terms as private firms.
- Independent authorities to set and administer prices for monopoly service providers.
- Introduction of a national regime ensuring that monopoly suppliers of essential infrastructure services provide access to third parties on reasonable terms and conditions.
- A legislative review program to assess whether regulatory restrictions on competition are in the public interest and, if not, the changes required.

To implement the agreements, the *Trade Practices Act* was extended to include government businesses. The ACCC was established to implement the Act, taking over from the Trade Practices Commission and the Prices Surveillance Authority. The National Competition Council (NCC) would monitor the progress of reforms, especially by state governments.

The Commonwealth, states and territories signed the agreement only after protracted arguments over the distribution of benefits and costs. The Industry Commission (1995) estimated that the reforms would increase GDP by 5.5 per cent per annum with four-fifths of these benefits due to reforms by state and local government and only one-fifth to reforms by the Commonwealth. However, whereas the Commonwealth would gain significant tax revenue from the reforms, especially from sales of power, gas and water businesses, the states would lose previously untaxed surpluses from these industries. To obtain agreement to the reforms, the Commonwealth agreed to increase general revenue grants to the states and territories by up to $600 million per annum in 1994–95 dollars, subject to satisfactory implementation of the NCP as assessed by the NCC.

These reforms affected all major sectors of the economy, including agriculture, mining, manufacturing, banking, the professions and public services. In 1998, the OECD (1998, p. 40) declared that 'the regulatory review in Australia is unprecedented in its scope and ambition in OECD countries'. The ACCC (run by economists) enforced the *Trade Practices Act* much more aggressively than did its predecessor, the Trade Practices Commission (run by lawyers). The average effective rate of assistance for manufacturing industry fell from 22 per cent in 1984–85 to 6 per cent in 1996–97 (Productivity Commission, 1991a). Over the same period,

[9] The National Competition Package included the Competition Principles Agreement (CPA), the Conduct Code Agreement and the Agreement to Implement Competition Policy. Clause 5(1) of the CPA provided that 'The Guiding Principle is that legislation (including Acts, enactments, Ordinances or regulations) should not restrict competition unless it can be demonstrated that: (a) the benefits of the restriction to the community as a whole outweigh the costs; or (b) the objectives of the legislation can only be achieved by restricting competition.'

trade intensity (the ratio of imports plus exports to GDP) increased from 30 per cent to 47 per cent. Liberalisation of foreign investment increased competition in local markets. Inward direct foreign investment increased from 1 per cent of GDP to 2 per cent of GDP. The Commonwealth introduced a requirement that a Regulatory Impact Statement, including a cost–benefit analysis, be undertaken for all existing and proposed regulations that may affect business. The state governments reviewed many restrictive laws and regulations.

The Productivity Commission (2005a) argued that these reforms produced important benefits (see below). However, the Commission also noted that reform was incomplete in many areas. Integrated national markets still did not exist in the energy, water and freight transport sectors. There were competitive restrictions in broadcasting, pharmacy and insurance services. State governments had not completed their reviews of restrictive legislation. The Commission also argued that there would be substantial gains from more competition in delivery of human services (community services, health and vocational education and training). Other challenges include the billion dollar assistance for domestic procurement for the defence industry, restricted competition in the provision of public transport, and the structure of industrial relations.

Benefits of Competition Policy

The benefits of competition policies are the reduction in the costs of imperfect competition described previously. They include increased supplies of goods due to more competitive pricing, reductions in x-inefficiency in production and increased innovation. Also, there may be reductions in the rent-seeking expenses associated with protection and industry assistance.

Most estimates of the benefits of competition policy focus on reductions in prices in the industries that are directly affected by competition policies because these are the most readily observable outcome. Lower prices are generally assumed to reflect productivity gains. Improvements in productivity allow resources to be released into other sectors.

However, there are important caveats. First, to estimate the effect of competition policies, the counterfactual (outcomes without the competition policies) must be estimated. Changes due to technical progress should not be counted as a benefit of deregulation. Second, lower costs may reflect lower wages as workers in the now unprotected industries no longer extract part of the economic rent in the industry. If labour is initially paid a wage above its opportunity cost, some savings in costs are transfers from workers to shareholders or consumers and are not real productivity benefits. On the other hand, lower prices may underestimate the benefits if they do not allow for any dynamic effects of competition policy flowing on to other sectors, for example through increased exports.

Ideally, gains from competition policies would be estimated using a computable general equilibrium (CGE) model of the economy. This allows for changes in output and prices in all sectors of the economy and for the release of factors of production into and out of the newly competitive sectors. However, a CGE model has many in-built and not always transparent assumptions. In any case, to estimate the benefits of competition policy it is necessary to estimate the counterfactuals.

Reviews of the effects of competition policies. In an early study Winston (1993) reviewed the effects of deregulation programs in the United States between the late 1970s and early 1990s. In this period, the proportion of GDP produced by fully regulated industries fell from 17 per cent to 7 per cent. Market conditions changed greatly in the transportation, energy, telecommunications and financial industries, which accounted for $600 billion of output. The total estimated benefit of about $40 billion per annum (Table 14.1) is conservative because it excludes benefits for the energy sector (not then analysed) and savings in rent-seeking costs. Ninety per cent of the estimated benefits accrued to consumers via lower prices or improved

services, for example in improved service frequency and reliability and faster travel times. Firms increased profits due to cost savings in the airlines and railroads, but lost profits with lower revenues in the trucking industry. Employment increased in cable television and banking but declined in trucking. In other industries, the employment effect was unclear.

The OECD (1998) reviewed productivity gains in OECD countries, including the United States, due to regulatory reform. Table 14.2 shows the estimated reductions in real prices that followed elimination of economic regulations. The report claims that these price reductions reflect large gains in labour and capital productivity, although it acknowledges that part of the reductions may be attributable to factors other than regulatory reform.

The OECD reports that labour productivity growth rates in the manufacturing sectors most affected by competition-enhancing reforms were double those in other sectors from 1986 to 1991. Capital productivity increased greatly in the road haulage industries. In addition, market liberalisation in telecommunications facilitated a rapid take-up of new services. The effect on employment depends on the output effects of the reforms and the flexibility of the labour market. In many industries, in telecommunications in Japan and Finland, in retailing in Sweden and the Netherlands, and in airlines in the United States, the increase in output more than offset the labour productivity gains, resulting in net job increases.

For Australia, the Productivity Commission (2005a) noted that the rate of increase in real per capita incomes in the second half of the 1990s was as high as at any time in the 20th century. Also, labour force participation was as high and unemployment as low as at any time over the past three decades. It attributed much of this growth to the competition reforms. It also noted that the only CGE modelling of the reforms was the prospective work cited above by the Industry Commission (1995).

In its 2005 review the Productivity Commission estimated the economy-wide gains due to productivity improvements and price changes over the 1990s in the electricity, gas, urban water, telecommunications, urban transport, ports and rail freight sectors, in all of which there

Table 14.1 Welfare effects of deregulation in the United States in the 1980s ($bn per annum at 1990 prices)

Industry	*Consumers*	*Producers*	*Total*
Airlines	8.8–14.8	4.9	13.7–9.7
Railroads	7.2–9.7	3.2	12.4–12.9
Trucking	15.4	-4.8	12.6
Telecommunications	0.73–1.6	-	0.73–1.6
Cable television	0.37–1.3	-	0.37–1.3
Brokerage	0.14	-0.14	0.0
Naturalgas	Na	Na	na
Total	32.6–43.0	3.2	35.8–46.2

Source: Winston (1993).

Table 14.2 Real price reductions after elimination of regulations

Sector	*Countries*	*Price reductions (%)*
Road transport	France, Germany, Mexico, USA	19–30
Airlines	Australia, Spain, UK, USA	20–33
Electricity	Japan, Norway, UK	5–26
Financialservices	UK, USA	30–70
Professionalservices	UK (conveyancing)	33
Telecommunications	Finland, Japan, Korea, Mexico, UK	10–66

Source: OECD (1998), *The OECD Report on Regulatory Reform, Synthesis*.

were significant reforms. In the 1990s, real prices fell Australia-wide by 19 per cent in the electricity sector, by over 20 per cent in telecommunications, by between 8 and 42 per cent for rail freight and by up to 50 per cent for port charges. The Commission concluded that reforms in infrastructure services boosted GDP by 2.5 per cent but this was a partial estimate because it does not include all effects of the policy changes since 2000.

Concluding Observations

In concluding, we note two very different issues. First, we must note that we have not addressed in this chapter possibly the most important industry issue of our time — namely the control of the internet which is dominated by a few extraordinarily large US companies, including Google, Facebook, Amazon, as well as by Chinese giants, Alibaba and Tencent. The issues here are both highly complex and politically important with allegations of international sabotage, fake news, abuse of private information and manipulations of democratic processes in both the United States and the United Kingdom. We will discuss these issues as we review and progress this in-line edition.

Second, and quite differently, in its most recent 5-year review of productivity in Australia, the Productivity Commission (2017) devoted only one of the five main chapters to the role of markets. Here the two main recommendations were reform of the energy markets (a mess largely caused by inconsistent government policies) and creating an effective environment for market innovation. In the other four chapters, the Commission focussed on improving the health of Australians, investing in future skills, more efficient towns and cities and more effective governments. This perspective emphasises the important and integral role of government in increasing productivity. Competitive markets are important, but not the whole productivity story.

Summary

- Competition policy is mainly about creating and maintaining competitive industry structures and behaviour. The aims are to improve the allocative, productive and dynamic efficiency of the economy.
- The relationship between competition policy and regulation is complex. Government often must regulate the conduct of firms to ensure that they act competitively. Also, when there are dominant firms, government may regulate their behaviour to ensure that they act in the public interest. On the other hand, improving competitiveness may require doing away with some regulations.
- The costs of monopolies include high prices due to restrictions on output, excessive production costs and a lack of innovation. Also, firms may incur unproductive rent-seeking costs. Collectively, these costs may total several percentage points of GDP.
- Competition strategies include structure, conduct and performance policies.
- Structural policies aim to create a competitive industry structure, for example by prohibiting acquisitions and mergers that would create a dominant business.
- Competitive policies deal with the general competitive behaviour of firms, for example outlawing collusive and other anti-competitive behaviour.
- Performance policies regulate dominant firms in key sectors, usually by rate-of-return price regulation. However, this may not provide an incentive to reduce costs. Also, there is a risk that such price regulation will reduce investment.
- Over the last 15–20 years, many OECD countries, including Australia, have introduced policies designed to increase competition. These have included opening the public sector to competition.
- These competition policies appear to have significantly reduced the prices of basic services and to have increased productivity and gross domestic product.

Questions

1. Should mergers and acquisitions that substantially reduce competition always be opposed?
2. What are the various ways to measure market power and industry concentration?
3. Why might rate-of-return price regulation of a dominant firm give rise to inefficiencies?
4. Assume that a monopolist produces at constant marginal such that $MC = 40$. The monopolist faces a demand curve given by $Q^d = 110 - 2P$. Determine the following:
   i. The marginal revenue function.
   ii. The profit maximising price and quantity combination of the monopolist.
   iii. The output level that would have been produced if the industry had been perfectly competitive.
   iv. The deadweight loss as a result of monopoly production.
5. Suppose that a firm operating in a perfectly competitive industry can produce at a constant marginal cost of $15 per unit. Suppose also that if the same firm operated as a monopoly it would produce at a constant marginal cost of $20 per unit. If the market demand the firm faces is given by $Q^d = 1260 - 4P$ and the marginal revenue is given by $MR = 315 - 0.5Q$, determine the following:
   i. The perfectly competitive and monopoly outputs and prices.
   ii. The loss of consumer surplus under monopoly production.
6. Suppose that a firm can produce at constant marginal costs of $5 per unit. It has a market demand of $Q^d = 60 - 3P$. Determine the following:
   i. The marginal revenue function.
   ii. The rent-seeking costs that the firm might incur to ensure its market was regulated to enable it to charge monopoly prices.
   iii. The deadweight loss that would arise as a result of the monopoly position.
   iv. The overall loss arising from the regulation.
7. If workers in developing countries work in inferior conditions, is this unfair competition?
8. Virgin Airlines has accused Qantas of engaging in predatory behaviour by increasing flights and reducing prices on inter-city flights. What criteria would you use, and what data would you require, to determine whether a firm is engaging in predatory pricing?
9. Australian competition policy requires that mergers and acquisitions be not allowed when they would result in a 'substantial lessening of competition'. What issues arise in interpreting this principle?
10. Can vertically integrated firms have internally exclusive dealings and thus avoid competitive restrictions on exclusive dealings?
11. What economic factors should a regulator consider in dealing with proposed mergers and acquisitions?
12. In the UK football (soccer) league there are no restrictions on payments to players. In Australia the football leagues impose salary caps on payments. What are the competitive implications?

Further Reading

Australian Competition and Consumer Commission, *Annual Report*, Canberra.

Australian Competition and Consumer Commission (2011) *Merger Guidelines*, ACCC, Melbourne.

Davis, P. and Garces, E. (2010) *Quantitative Techniques for Competition and Antitrust Analysis*, Princeton University Press, Princeton, NJ.

Kaplow, L. and Shapiro, S. (2007) *Antitrust*, Working Paper No. 12867, National Bureau of Economic Research, Cambridge, MA.

King, S.P. (2010) 'Using microeconomics to protect competition', *Australian Economic Review*, 43, 217–224.

NSW Independent Pricing and Regulatory Tribunal (2018), *Review of out WACC Method*, IPART, Sydney

Productivity Commission (2005a) *Review of National Competition Policy Reforms*, Report No. 33, Canberra.

Vickers, J. (2005) 'Abuse of market power', *Economic Journal*, 115, F244–F261.

Viscusi, W.K., Vernon, J.M. and Harrington, Jr., J.E. (2005) *Economics of Regulation and Antitrust*, 4th ed., Chapters 1, 3–4, 6, 9–12, MIT Press, Cambridge, MA.

Yang, Q.G. and Pickford, M. (2011) 'Safe harbours in merger guidelines: what should they be?', *Australian Economic Review*, 44, 13–35.

Chapter

15

Information and Safety

Deception as to the physical nature of a thing offered for sale is practised through false weights and measures, adulteration and misnaming of goods, and dishonest advertising … (Traditionally) the back streets of the manufacturing towns swarmed with small shops, in which the worst of everything was sold with unchecked measures and unproved weights.

Arthur Pigou, *Economics of Welfare*

Information Failures: Nature and Costs ♦ Policy Responses to Information Failures ♦ Product Quality and Safety ♦ Workplace Health and Safety ♦ Regulating Occupations ♦ Private and Social Insurance ♦ Merit Goods: Bounded Willpower and Rationality

When people are misinformed, market outcomes are inefficient and often unfair. In this chapter, we will be concerned mainly with issues relating to misleading information and information asymmetries. The provision of information as a public good with non-excludable or non-rival characteristics is a separate issue and discussed elsewhere, for example in Chapters 4 and 11.

In the first part of this chapter we describe the nature and costs of information failures. We then discuss product quality and safety, workplace health and safety, regulating occupations and insurance issues. Policies relating to information failures usually involve regulations and a theme of the chapter is the need to assess the costs and benefits of regulations. Many of the issues concern safety in the workplace and the community. There is finally a short discussion of merit goods or what today may be described as bounded willpower and rationality.

Information Failures: Nature and Costs

Table 15.1 overleaf provides examples of potential information failures in product, insurance and factor markets. Many would have significant potential costs.

Product markets. Failures in product markets occur when suppliers have information about product defects or risks that consumers do not have. The information asymmetry ranges from major fraud down to minor deceptions. In some cases, firms supply misleading information about product quality. In others, they suppress negative information. Possibly the most infamous case is cigarettes, where tobacco companies were aware of the fatal carcinogenic effects of tobacco consumption long before consumers were (Glaeser and Ujhelyi, 2010). Other examples are defective motor vehicles, ineffective or unsafe patent medicines, unsafe electrical products, unhealthy foods and unsafe or unhealthy buildings. Information problems

Table 15.1 Examples of information failures

Possible information failure	*Examples*
Product markets	
Producer knows more than consumer	Products such as: motor vehicles, electrical products, pharmaceuticals, food products, buildings
	Professional and technical services
Insurance and personal finance	
Consumer knows more than insurer	Insurance products: health, home, vehicles
Uncertainty about future income	Student loans, unemployment
Factor markets	
Workforce uninformed	Occupational health and safety
Asymmetric information in capital markets	Financial products
	Insider trading

may also arise with provision of professional or technical services, where inevitably the supplier has more information than the consumer.

We saw in Chapter 4 that in a market for 'lemons' (for example second-hand cars), inability to discriminate between good and poor products drives out the good products. Here we describe the implications when consumers buy goods in an uninformed market. Figure 15.1 contains demand curves (D_I and D_U) for informed and uninformed consumers respectively. Curve D_U is higher because uninformed buyers are willing to pay higher prices for goods than are informed buyers. Consumer misinformation raises the price from P_I to P_U and consumption from Q_I to Q_U. This raises producers' profit by the area P_UACP_I at the expense of consumers. There is a deadweight loss equal to area ABC. This loss is the difference between the cost of supplying the extra goods ($Q_U - Q_I$) and the price that informed consumers would be willing to pay for these goods.

Insurance markets. Asymmetric information also occurs in insurance markets. In this case. purchasers of insurance products often know more about the risks than do insurers. Purchasers also have an incentive to hide information about risks from the insurer, for example health risks or vehicle accident history, to reduce the price of their insurance. Unless

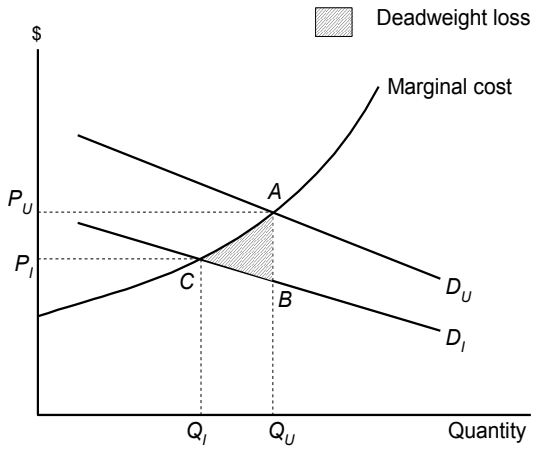


Figure 15.1 Costs of poorly informed consumers

insurers can acquire knowledge about their clients, they are likely to set average actuarial premiums for the population seeking insurance. Consequently, they attract a disproportionate number of poor risks and undersupply insurance services to low-risk firms or individuals (a consequence of adverse selection described in Chapter 4).

A lack of information about future income may also restrict the market for private loans and income insurance. It is hard for students to borrow against future income and for workers to insure against loss of income. Again, there is an adverse selection problem—insurers cannot easily distinguish between high and low risks.

Moral hazard compounds the problem. Once insured, individuals may change their behaviour. For example, they may make less effort to protect their property or to obtain work. To protect against such risks, insurers increase the price of insurance or in some cases decline to offer insurance. Some insurance contracts that would be provided with full information on the risks and no moral hazard are not supplied.

Labour markets. In labour markets, employers may provide minimal or even misleading information about workplace conditions and not provide safe working conditions. When workers are poorly informed about their work environment, they accept inadequate wage compensation for the risks involved. Labour resources are misallocated away from safe or pleasant jobs to unsafe and less pleasant occupations. This is a major area of government regulations.

Financial markets. Information asymmetry also occurs on a large scale in financial markets, in exchanges of assets and in capital raising exercises. We now know of many examples preceding and precipitating the global financial crisis in 2007–08 where leading financial firms sold at premium prices financial products that they knew to be sub-prime and highly risky and, in some cases, then bet that the price of these very same products would fall. Traders in financial markets can make large amounts of money when trading with insider commercial information known to only a few people.[1] Companies may spruik profits and conceal debts to raise their share price and lower the cost of new capital. Individual borrowers may misinform financial institutions about their true financial position. These information failures are inequitable and cause misallocation of capital. Firms and individuals who are poor risks attract more capital than they would from better informed lenders. On the other hand, confidence in the capital market falls and owners of capital require a higher price for providing capital than they otherwise would.

Information deficiencies in financial institutions can have even more critical systemic consequences. Most depositors with financial institutions can neither judge the solvency of these institutions nor insure against losses. Private rating agencies, often paid by the institution that they are rating, have shown that they cannot be relied upon. Indeed, financial institutions themselves may not trust other financial institutions. Consequently, credit markets may dry up, as in 2008–09. A failure of a major financial institution creates not only great personal hardships but may lead to a run of bankruptcies as funds are withdrawn from solvent as well as insolvent organisations. A system-wide run on financial institutions has severe consequences, including a very high cost of capital. Information failure here leads to the collapse of a major public good—the integrity and reliability of the financial system itself.

Market responses. Markets can respond to information deficits in various ways. An active market in expert independent advice or more broadly in information has developed in Australia and most OECD countries to deal with just such problems. Motoring organisations supply information about the quality of second-hand cars. Building surveyors advise on

[1] For insight and an excellent read, see Partnoy (1997) *F.I.A.S.C.O, Blood in the Water on Wall Street*.

building quality. Second or third medical opinions are nearly always available (at a price). Consumer organisations such as Choice in Australia advise on product quality.[2]

Also, most suppliers of goods and services depend on their reputation. The incentive to maintain their reputation is often (though not always) greater than the incentive to mislead. Some firms provide warranties (a form of insurance) for the performance of their goods.

Despite these private responses, asymmetric information between buyers and sellers in any market can raise prices significantly, distort the quantity and quality of goods consumed and create inequities. These distortions and inequities are large enough to warrant government involvement in many markets.

Policy Responses to Information Failures

Table 15.2 lists some policy responses to information failures. It may be observed that nearly all policies dealing with product and factor markets are regulations of one or other kind. The provision of public information to counteract misleading information in the market is a generic alternative policy. But it is rarely a sufficient policy response and is generally complementary to, rather than a substitute for, regulations. Opportunities to use market-based instruments are more limited than for externalities. However, financial penalties are essential to enforcement of regulations. In some cases, for example for serious contravention of consumer product safety provisions and for insider trading, criminal charges are also applied. In general, if penalties are too low regulations are ineffective.

Most regulations are aimed at firms supplying products or services or employment rather than at individuals or consumers. The regulations may be economy-wide measures or industry specific. Most OECD countries lay down general conditions for fair trading and rules for advertising of goods and services. Also, most countries have specific regulations dealing with food products, pharmaceuticals, transport and buildings and so on. In addition, they usually have regulations dealing with occupational health and safety, fair trading in equity markets and prudential safeguards for financial institutions concerning reserves, liquidity and so on.

Table 15.2 Information failure and possible policy responses

Information failures	*Some policy responses*
Product markets	
Product information	Regulation of advertising
	Public provision of information
Product safety	Regulation of quality of products
Professional services	Entry regulation (licensing of services)
Financial services	Regulation of financial institutions
Factor markets	
Health and safety in the workplace	Regulation of workplace standards
Capital markets	Regulations on provision of information
	Bans on insider share trading
Insurance and personal finance	
Risk and insurance	Compulsory third-party motor vehicle insurance
	Public insurance, e.g. public health insurance
Income uncertainty	Public loans to tertiary students
	Social security schemes

[2] There are also thousands of financial advisers in Australia, but many are conflicted because they receive commissions from the companies whose product they recommend. This issue is being examined by government.

These regulations usually have costs as well as benefits. The costs may involve some reduction in sales or increased costs of production. The benefits arise from the avoidance of the kinds of costs described above. For example, regulations on food production typically increase the costs of producing food and reduce foodborne disease. Accordingly, the Australian Office of Best Practice Regulation (2014) recommends that regulations should generally be evaluated using cost–benefit analysis.

The policy response in insurance markets is more complex. Typically, market failure means that various people cannot get actuarially fair insurance. However, governments are often more concerned that poor risks, often disadvantaged persons or families, cannot obtain insurance for say health care or motor vehicles at an affordable price. The principal problem then is one of social justice rather than market failure. In this case, the standard government response is either to mandate community insurance premiums (equal premiums for all persons in certain categories regardless of risk) or to cap premiums or alternatively to make public provision of insurance. All these practices can be found in Australia. For example, the Australian government mandates that private health insurance is based on community insurance principles where high-risk persons are subsidised by lower-risk persons. It also provides general health insurance for everyone. In New South Wales, the Motor Accidents Authority caps premiums for insurance of third-party vehicle accidents.

Consumer protection in Australia

In Australia, the principal legislation dealing with fair and safe trading in the sale of goods and services is the *Australian Consumer Law* (see Box 15.1). This is a schedule to the *Competition and Consumer Act 2010*, which replaced the longstanding *Trade Practices Act 1974*. The new Act applies nationally. It is implemented by the Australian Competition and Consumer Commission (ACCC) and each state and territory's law agency.

Box 15.1 Consumer protection under the *Australian Consumer Law 2010*

The *Australian Consumer Law 2010* (ACL) outlines unfair trade practices and regulates product safety and information requirements.

Chapter 2 of the ACL provides general consumer protections and outlines general standards of business conduct in the market.

Part 2-1 sets out a broad provision that prohibits firms from engaging in any conduct that is likely to mislead or deceive. This covers false representations about attributes of goods such as the standard, quality, price, origin, composition, style, warranty or guarantee, as well as any prior history of the goods.

Part 2-2 is a general ban on unconscionable conduct in trade or commerce plus more specific prohibitions on certain business transactions. The concept of unconscionable conduct has developed in common law.

Part 2-3 makes void unfair contract terms in consumer contracts. A term is 'unfair' when there is any imbalance in the parties' rights or obligations or when it causes any detriment to one party.

Chapter 3 of the ACL provides specific protections to address identified forms of business conduct.

Part 3-1 prohibits certain false or misleading representations, the supply of unsolicited goods or services, participating in pyramid schemes, bait advertising, and practices involving the display of prices, referral selling, harassment and coercion.

Part 3-2 provides guaranteed consumer rights for goods and services, national rules for unsolicited sales transactions, and five basic rules for lay-by agreements.

Part 3-3 creates a new consumer product safety law and regulatory framework that applies nationally.

Part 3-4 creates a single national law that prescribes information standards for consumer goods and services.

Part 3-5 creates national rules governing the liability of manufacturers for safety defects.

Chapter 4 sets out criminal offences relating to certain matters covered in Chapter 3.

Chapter 5 sets out national enforcement powers and remedies relating to consumer law.

Other important product safety agencies are Food Standards Australia New Zealand (FSANZ) and the Therapeutic Goods Administration (TGA).[3] FSANZ is responsible for ensuring safe food in Australia and New Zealand. All food products have to meet detailed high standards across the whole food chain from primary production to manufacturing to distribution and retail. FSANZ also has the power to recall food products from the retail and distribution chain when there is an 'unacceptable safety risk to consumers'.

The TGA regulates therapeutic goods in Australia, under the *Therapeutic Goods Act 1989*, to ensure their quality, safety and performance. Therapeutic goods are defined as any product which is represented in any way to be taken for therapeutic use, including complementary medicines. The TGA assesses therapeutic goods before they can be used and monitors their compliance with standards once the goods are on the market. Like FSANZ, the TGA can recall unsafe products.[4]

Product Quality and Safety

Most products carry some risk. This does not mean that they should not be supplied. Nor does product failure necessarily imply market failure. Suppose that consumers can purchase a product with an expected 1 in 20 probability of malfunction at a price discount of 15 per cent. Providing that the consumers suffer no personal harm, as a group they are better off with the product than without it even though 5 per cent of consumers will experience product malfunction. However, if suppliers are aware of the product risks and consumers are not, there may be misallocation of resources and a need for consumer protection, especially when products may cause personal harm.

As we have seen, governments usually enact general legislation to outlaw unfair trading and misleading or deceptive conduct to ensure that consumers are making an informed choice on product quality and price. Here we focus more specifically on safety and public health issues.

In general, products can be made safer by increased spending on product design, development and manufacture. Establishing an efficient level of product safety can be viewed as a cost–benefit exercise. The safety level is efficient when the marginal benefit of increased product safety equals the marginal cost. Equivalently, as a cost-minimisation exercise, an outcome is efficient when the total cost, including expenditure on product safety and damage costs, is minimised. In Figure 15.2, the efficient level of product safety is Q_E. Panel (a) overleaf illustrates the marginal approach, panel (b) the total cost-minimisation approach. Of course, equity considerations might suggest a higher, or lower, level of product safety.[5]

We consider below four strategies for ensuring an efficient level of product safety: product liability, regulation of information, regulation of products and processes and taxation of unsafe products.

Product liability implies that suppliers of unsafe products bear some or all of the damages from their products. Under a negligent liability standard, a firm pays for the damage costs that would have been avoided if it had adopted an 'efficient level of safety'.[6] Under strict liability, the firm pays all damage costs regardless of whether it has met an appropriate or efficient level of safety. Actually, if consumers are aware of their rights, whichever liability standard is adopted the firm has an incentive to choose an efficient level of product safety (the level at

[3] The US Food and Drug Administration is the equivalent agency in the United States.

[4] For information on FSANZ, see <www.foodstandards.gov.au>. For the TGA, see <www.tga.gov.au>.

[5] The approach to obtaining an efficient level of product safety is of course the same as that employed to obtain an efficient level of environmental quality (see Chapter 13)

[6] This definition comes from Viscusi *et al.* (2005, p. 812). However, Viscusi *et al.* also talk about 'an efficient degree of safety at the medium level of safety'. A non-negligent level of risk is not a precisely defined concept.

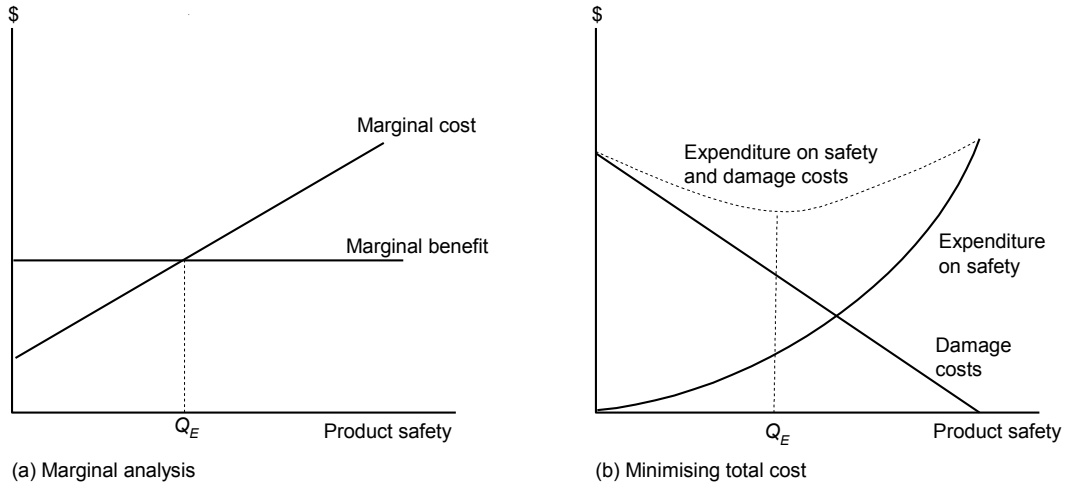


Figure 15.2 Efficient level of product safety

which the marginal benefit and marginal cost of an extra unit of safety are equal). This is an example of the Coasian theorem that markets can achieve efficient outcomes regardless of how property rights are assigned. Of course, the distribution of costs varies with negligent or strict liability.

Despite its potential efficiency advantages, a public liability strategy may not achieve efficient outcomes. Individuals may not want to incur legal costs to obtain redress for minor damages. Even with class actions, consumers may free ride and not engage in the legal process to obtain redress for damages. This would lead to undersupply of safe products. On the other hand, if courts award costs in excess of actual damages, some socially valued products may become unavailable. Viscusi *et al.* (2005) suggest that, in the United States, high court-awarded damage costs have substantially increased the costs of public transport and other products such as ladders and led to the demise of the domestic aircraft industry, the withdrawal of vaccines from the market and the slow development of contraceptive devices.

Regulation of information. Regulations may require firms not only to not mislead their customers but also to advise of any potential safety risks associated with their product. For example, most pharmaceutical products carry warnings against excessive use. Many governments require tobacco companies to carry health warning messages on the packaging of their products. Indeed, the Australian government requires tobacco companies to carry graphic pictures of the risks of tobacco consumption. Before introducing this initially, the government commissioned a cost–benefit analysis that showed the proposed health warnings would provide a substantial net social benefit (Applied Economics, 2003a).

Regulation of products and processes includes regulation of technical standards of products, pre-market approvals and product recalls. A prime example of setting technical standards is the regulation of motor vehicles. Regulations typically include compulsory safety belts with specific design requirements, side door strength, bumper requirements and fuel system integrity standards. Viscusi *et al.* (2005) estimated that the total cost of vehicle safety regulations in the United States exceeded US$1000 per vehicle. In the United States, as in Australia, traffic accidents per vehicle kilometre travelled have fallen a great deal in the last

30 years following the introduction of such safety regulations. However, this is also due to improved roads, better traffic management and controls over driving speeds and drinking and driving. Australian studies (Applied Economics, 2003b) show a high benefit–cost return for such road safety programs. However, it is often hard to estimate the benefits of specific product safety regulations.

A prime example of pre-market testing is licensing of drugs. Here a regulator typically seeks to determine whether a drug is technically effective and safe. Drugs with potentially adverse consequences are not licensed. Usually the determination is based on technical evidence and judgement rather than on economics. However, a regulator can make two kinds of error. The regulator may license drugs that are not safe or deny a licence to a safe drug. Licensing unsafe drugs has clear costs. Failure to license a safe drug also has costs as it denies consumers access to useful and possibly life-saving medicines. Ideally the regulator would review the costs of both kinds of error and determine the licence on the costs and benefits of the decision, rather than simply try to avoid one kind of error (licensing an unsafe drug).

Product regulators may also have the power to withdraw unsafe products from the market. In Australia, FSANZ typically orders about 50 food recalls over a year. In 2004 a major drug manufacturer (Pan Pharmaceuticals) went into liquidation following a major product recall by the Australian Therapeutics Goods Administration.

Regulatory capture. So far, we have assumed that regulations are designed to protect consumers rather than regulators. As discussed further below, this is not always the case. Australia has long prohibited the import of apples from New Zealand on the ground that they could introduce various diseases, but many people suspect that the main reason has been to protect the Australian apple industry. Likewise, several states restrict the number of taxis, ostensibly in part to protect public safety but primarily to protect the value of taxi licences.

Taxation of unsafe products is another approach. An example is the excise tax on cigarettes in Australia. Bardsley and Olekalns (1999) found that this tax had more effect on reducing tobacco consumption than did public anti-smoking campaigns. Where the harmful impact of the product is solely on the consumer, this is essentially a tax on a demerit good that aims to discourage personal consumption rather than a corrective tax on an externality.

However, taxation is not often employed for unsafe products. Governments tend to regard products as either safe or unsafe and do not recognise intermediate cases where taxes may discourage consumption without prohibiting it. Another reason for eschewing injury taxes is that it is often difficult to determine the relative contribution of the product and the consumer to injuries. Accidents with ladders may reflect unsafe use or poor climbing ability rather than unsafe ladders. Taxing ladders would have little impact on accidents with ladders.

Workplace Health and Safety

In informed and competitive labour markets, wage differentials compensate for differences in working conditions. Workers require higher wages to compensate for less healthy or safe working conditions or, equivalently, they are willing to accept lower wages for an increase in non-wage amenity.

The actual wage differential depends on both workers' preferences for health and safety and on the ability of firms to provide a safe environment. In Figure 15.3 overleaf the convex indifference (I) curves show the trade off between wages and workplace safety that an employee is willing to make for a given level of utility. A higher curve represents more utility. On the other hand, the concave iso-profit (P) curves show how firms can trade-off higher wages with lower provision of safety (for any given level of profit, including zero economic profit in a competitive market). Equilibrium is achieved in each market when the marginal rate at which

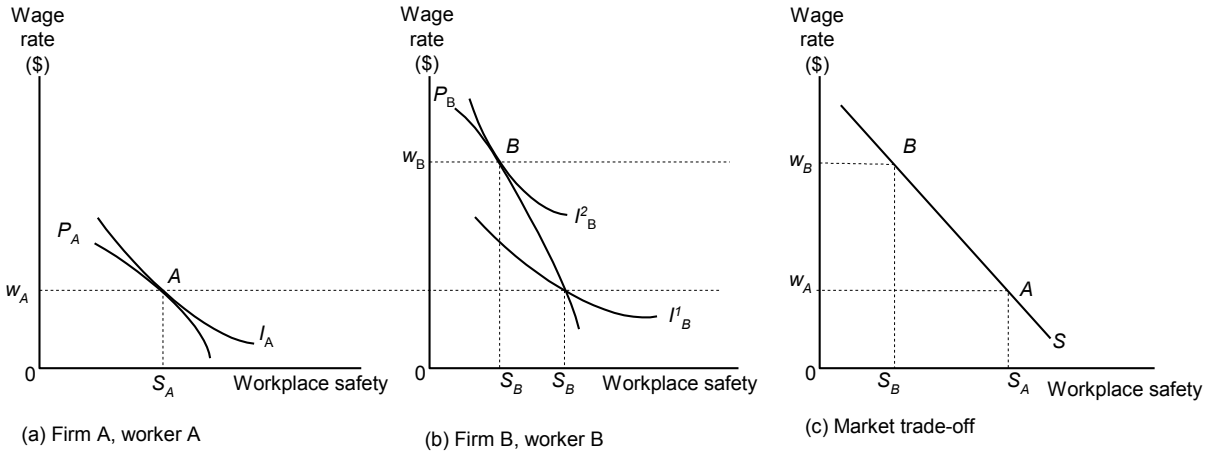


Figure 15.3 Price of safety in the workplace

Source: Based on McConnell, Bruce and Macpherson (1999).

workers are willing to substitute changes in wages for safety equals the marginal rate at which firms can transform wages into safety.[7]

Panel (a) shows a worker with a high marginal value of safety along with a firm that can provide safety at a low cost—the outcome is a low wage. Panel (b) shows a worker with a lower marginal value of safety combined with a firm that can provide safety only at high cost—the outcome is a high wage. Panel (c) shows the relationship between wage rates and job safety in a competitive labour market with heterogeneous workers and jobs. Other things such as productivity being equal, when workplace safety varies there is a trade off between wages and safety, with a higher wage for working in a less safe environment.

With the compensating wage differential model, a profit-maximising firm invests in safety so long as the marginal savings from greater workplace safety due to lower wages, less workplace disruption, and a more stable workforce exceed the marginal costs of providing the extra safety. Figure 15.4 shows the marginal cost and benefit of an extra unit of job safety. The efficient level of workplace safety is Q_E.

In a perfectly competitive and informed labour market, workers are compensated for any risk that they bear. If a worker is not compensated adequately, he or she moves to another firm. If there are no externalities, this level of job safety is efficient. The amount of job safety provided by the employer matches the level required by the employee.

Policy responses to market failures

However, markets provide an efficient amount of safety only when workers are aware of the risks and have alternative employment options. Viscusi *et al.* (*ibid.*) suggest that workers are generally aware of explicit technical risks, such as the risks of working at heights, but much less aware of chronic health risks, such as exposure to asbestos. In practice, information about occupational health and safety is incomplete and costly to obtain. If employees underestimate workplace risks, wages under-compensate workers for risks and employers under-provide workplace safety.[8] Job safety is also under-supplied when some costs are externalised and the

[7] This concept underlies the wage equations used to derive the value of life described in Box 11.2.

[8] On the other hand, if workers overestimate job risks, wage premiums for safety may be high and employers may oversupply job safety.

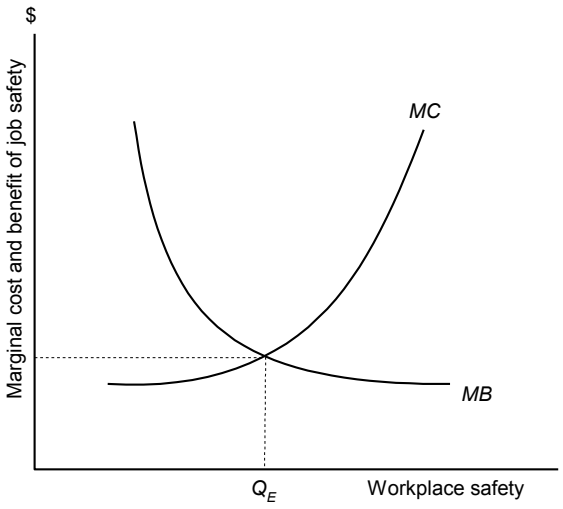


Figure 15.4 Optimal workplace safety

public health care system bears some of the injury costs. In this case, the marginal social benefit of job safety is greater than the marginal private benefit to the firm and the socially efficient level of job safety is again greater than the market-provided amount.

What is the appropriate policy in these cases? Public provision of information is the most direct response to misinformation. It eliminates directly the informational market failure. Viscusi *et al.* (*ibid*) cite an example relating to chemical labelling that suggests that hazard warnings may be effective. However, the information approach relies on market forces to change workplace conditions. It does not require employers to improve conditions. Therefore, regulators rarely rely on information alone to improve workplace safety.

Worker compensation policies imposed on firms are a market-based strategy that can influence employer incentives. Firms would bear the costs of injuries of workers either directly or via the intermediation of insurance premiums. This internalises the injury costs. With an efficient insurance market in which firms bear all the costs of employee injuries and there are no externalities, businesses would have an incentive to provide an efficient level of workplace safety. Additional regulations would be unnecessary.

However, a combination of regulatory and market failures may create an inefficient outcome. In Australia regulatory failure occurs because state governments prescribe insurance premiums for each type of business. Firms within each insurance category are not rewarded for a good safety record. Market failures include moral hazard (insurance makes workers less careful and more likely to claim for an adverse incident), transaction costs especially for court cases, and external costs borne by other parties, usually by government through the public health care system.

In practice, most governments employ a range of regulations to ensure safe practices in the workplace. In Australia, state agencies such as Workcover NSW prescribe safety practices for numerous activities and occupations. For example, there are detailed controls over the type of plant and equipment allowed in the construction industry, on farms, for long-distance truck driving and in brothels (see www.workcover.nsw.gov.au). These conditions deal with both the technology of plant and equipment and behaviour in the workplace. Commonwealth and state industrial courts also prescribe safety conditions in industrial awards.

As discussed in Viscusi *et al.* (*ibid*), regulators typically set safety standards on the basis of their 'affordability'. However, affordability is not a precise concept and there is rarely explicit consideration of an efficient level of safety based on the marginal costs and benefits of a

safety regulation. Regulators often know more about safety risks than employees do but less about both the safety risks and the costs of producing safety than the businesses themselves do. To achieve an efficient level of safety a regulator needs to know not only the technical risks in each case but also the marginal costs and benefits of producing safety.

In addition, the regulator must be able to monitor and enforce the regulation(s) without undue cost and the regulated party must expect to be caught and punished for any significant infraction of the regulation. And, of course, a regulator must not be captured by the parties that are being regulated.

Regulating Occupations

A major information concern is the quality of service supplied by professional and technical services. How do we know that our doctor, our builder or our electrician is competent? To ensure the quality of such services governments often regulate either the structure of the occupation or the conduct of practitioners in the occupation.

As shown in Table 15.3, there are several potential regulatory instruments. These include controls on entry, ownership of businesses, professional conduct, advertising and fee scales. These regulations are intended to provide social benefits by assuring quality services. However, most regulations have costs. For example, entry restrictions ensure that the provider

Table 15.3 Regulating occupations

Regulation	*Effect of regulation*	*Benefits*	*Costs*
Controls over entry	Excludes people lacking skills from providing a service	Ensures a competent level of service; reduces risk of harm; minimises flow-on effects of malpractice	Increases prices due to restricted supply; limits choice of provider; adds administration costs of registration
Controls on business ownership	Precludes some business structures and sources of capital	Minimises conflicts of interest between professional service providers and business interests; maintains trust between client and professional	Increases practice costs; inhibits innovation by restricting entry of capital and entrepreneurial expertise
Controls on conduct of professionals	Provides opportunities for removal of incompetent practitioners and redress of consumer complaints	Maintains integrity of profession and standards of service	Transaction costs of dealing with complaints; inefficient if controls are not related to professional services
Controls on advertising	Reduces false and misleading claims	Protects consumers; reduces wasteful expenditures; discourages competition by quality sacrifice	Discourages new entry into market; reduces consumer information on quality and price of services; reduces price competition
Controls on fee scales	Provides price stability and financial certainty	Removes risk that price cutting would reduce service quality; reduces transaction costs	Reduces ability of professionals to compete; may increase prices; restricts innovation

meets a competent standard and minimises the flow-on costs of malpractice to the community. On the other hand, they reduce service supply, increase the price of services and have administration costs. Controls on advertising prevent misleading information, but reduce service competition, increase the costs of entry into an occupation and may reduce the capacity of consumers to make an informed choice of a provider.

Figure 15.5 illustrates the possible deadweight loss from due to restricting entry into an occupation. With a free-entry occupation the market would supply Q_M labour hours at w_M wage per hour. If government limits the supply of services to S_2, the quantity of services would fall to Q_R labour hours and the wage would rise to w_R. Would-be suppliers of labour who are prevented from supplying services suffer a loss of surplus shown by the area *BCE*. If consumers are well-informed, they would lose a surplus equal to area *ABE* from the non-availability of services. Thus, the total deadweight loss is given by area *ACE*. Consumers also bear the higher prices, but the rectangle w_RABw_M is a transfer from consumers to producers rather than a deadweight loss.

In Australia, under the National Competition Policy governments are required to justify all such regulations as being in the public interest or to repeal the regulation. In effect, a benefit–cost test is required. The benefits and costs of each regulation should be quantified where possible in monetary terms and the benefits shown to justify the costs. Box 15.2 summarises a cost–benefit analysis of a NSW state regulation that requires that all boundary surveys be done by registered surveyors.

In this case the costs of regulation are low because registered surveyors circumvent the regulations by employing other technicians, usually unregistered surveyors or engineering draftsmen, to do most of the work. On the other hand, the regulation probably reduces land disputes and saves some demolition and rebuilding costs. Consequently, the estimated benefits of the regulation exceeded the costs. However, the conclusion is not the main issue here. The purpose of the example is to show that regulations, like other policies, have benefits and costs that are amenable to cost–benefit evaluation.

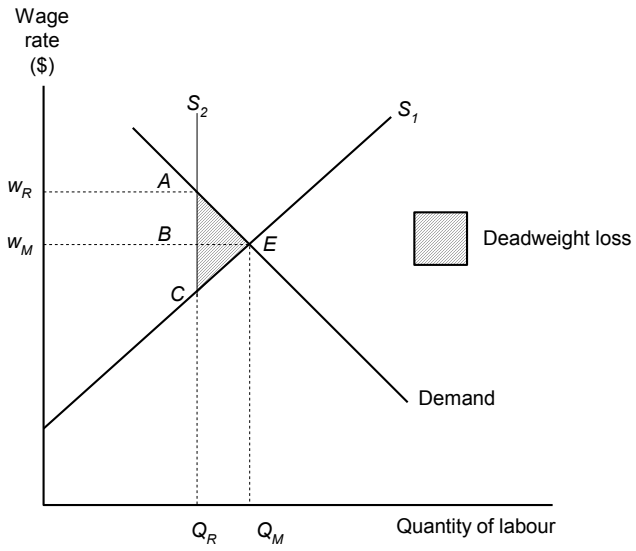


Figure 15.5 Costs of occupational licensing

Box 15.2 Regulation of land surveyors: cost-benefit analysis

Most Australian states have legislation that requires property boundaries to be certified by persons qualified and registered by the local Board of Surveyors. This regulation may be justified by the public's lack of knowledge about the qualifications and experience required for land surveying, the costs of incorrect surveys to property owners and adjacent landholders, and the importance of accurate mapping for property exchange and contracts. Thus there are information, externality and public goods arguments for some regulation of the industry.

However, the regulations prevent unregistered surveyors, engineering technicians and geographical systems experts from certifying property surveys, which increases the costs of surveys. In practice, the cost increase is not high because registered surveyors often employ engineering technicians to do much of the work and the surveyor simply certifies the product!

The then NSW Department of Information Technology and Management (2000) estimated that the regulation of land surveying in New South Wales increased surveying costs by only $2.5–$5.0 million per annum.

The study estimated that the benefits of regulation (savings from dispute reduction and lower demolition and rebuilding costs) probably exceeded the cost of regulation. It also found that the costs of using market mechanisms to deal with poor information, for example by insurance of property titles against error, would probably exceed the costs of the regulation. The study concluded that regulation is justified, but that some relaxation of the qualifications required for registration of boundary surveyors should be considered.

Regulatory capture and rent-seeking costs

Regulatory capture
Situations in which a regulator advances the interests of the industry or occupation that they are set up to regulate

Although regulations may be intended to protect consumers, not infrequently when designed they give as much or more assistance to producers. An industry may seek and obtain regulation to reduce competition and raise prices. As Friedman and Friedman (1980, p. 282) observed, 'there is no occupation so remote that an attempt has not been made to restrict its practice by licensure'. **Regulatory capture** occurs when a regulator identifies with the interests of an occupation or industry that they are supposed to regulate.

In a classic study of regulatory capture, Stigler (1975) explained how regulations develop because of the advantages that they confer on the regulated industry. Stigler argued that regulations are made because industry demands them. Stigler then showed that regulation of occupations in the United States is more likely when the occupation is large (carries many votes), per capita income is high (individual gains are large), urban concentration is high (organising costs are low) and opposition is dispersed. Although consumers generally have more votes that the regulated party, producers often succeed in extracting and maintaining regulations that restrict competition. Producers have more to gain individually, are more concentrated geographically, and are better organised and resourced.

Regulatory capture is not a zero sum game in which producers simply gain from higher prices what consumers lose. Deadweight losses occur not only as a result of supply restrictions (as we have observed), but also in the form of rent-seeking costs. The latter costs are resources that firms expend in order to gain economic rents (area $w_R A B w_M$ in Figure 15.5). Firms and industry associations of all kinds often expend considerable resources to create and maintain this economic rent.

Private and Social Insurance

Can private insurance markets protect or compensate consumers and workers against uninformed decisions? If the answer is 'no' for whatever reason, there may be a case for publicly provided social insurance. This could replace, or be additional to, other regulations designed to protect consumers or workers.

Insurance (private or public) generally provides two main benefits. First, it compensates individuals for injuries or misfortunes. Insurance against damage to properties, motor vehicles or health are prime examples. Second, it evens out consumption over time. By purchasing

insurance against future adverse effects, whether it is a motor vehicle accident or loss of a job, an individual can ensure that their consumption level is smoothed out regardless of what happens. Gruber (2016, Chapter 12) shows that, if individuals experience decreasing marginal utility from consumption, in a perfectly functioning insurance market, individuals will purchase insurance to smooth out their consumption fully against various possible states of the world.

However, as we have seen, markets may fail to produce an efficient amount of insurance because of asymmetric information and related adverse selection problems. Consequently, insurance companies charge average or higher than average insurance premiums and low-risk individuals cannot buy actuarially fair premiums.[9] In addition, individuals may under-insure because of a lack of information about risks or because they do not understand the benefits of insurance. In the latter case insurance would be a merit good.

Policy responses to insurance market failures include doing nothing, mandating that individuals hold a certain level of private insurance, subsidising private insurance, providing full or partial public insurance, and providing social assistance to persons in need. For example, the Australian government mandates that at least 9 per cent of all wages must be paid into private superannuation funds (a form of compulsory consumption smoothing) and all states require that motorists purchase third-party motor vehicle insurance. The Australian government is also considering mandating that insurance companies must offer households compensation for flood damages. On the other hand, the Australian government subsidises private health insurance and provides a form of universal health insurance (from consolidated revenue). Under a social assistance strategy, government picks up the costs of risks such as unemployment or disability, but government generally does not cover damages to property except in extreme cases.[10]

Evaluating policy responses is complicated by various factors, notably externalities and social justice issues. Often there are third-party effects, for example with motor vehicle accidents or fire damage to properties. And often the party suffering the injury will be disadvantaged in some way.

Focusing here on the efficiency issues arising from adverse selection and moral hazard, government faces similar problems to the private sector and its intervention may not improve matters. For example, mandating that everyone must be fully insured against health risks is likely to require some low-risk consumers to pay average premiums when they would prefer to be self-insured. They are required to take out more insurance than is efficient and are implicitly taxed to support the market. Similar problems arise with public provision of insurance. If it is funded by an equal levy on everyone, low-risk individuals are required to over-insure. Also, public insurance may crowd out some private insurance.

In addition, insurance of any kind (private or public) raises problems of moral hazard. Insured individuals may change their behaviour and take more risks than they otherwise would. Also, they may claim more compensation or use more services, such as medical services, than they would otherwise. Therefore, although the provision of insurance, whether mandated or publicly financed, is designed to reduce market failure, it introduces another set of costs.

In conclusion, there are limited effective policy responses to insurance market failures. Government is unlikely to have superior selection knowledge to private firms or more capacity to handle moral hazard behaviour. However, when there are third-party effects, there may be a case for mandating insurance for individuals or for some social insurance, but this

[9] Gruber (2011) shows that risk-averse individuals may be willing to pay a higher premium and still purchase insurance.

[10] The Australian government does provide drought relief for farmers and occasional compensation to households who experience extreme fire or flood damages.

raises the potential of moral hazard. Also, when adequate private insurance is not available, for reasons of social justice government may provide assistance to individuals suffering substantial adverse events, so long as this does not itself create significant moral hazard and unduly reduce the precautionary actions that individuals would otherwise take.

Merit goods
Goods that are believed to provide more benefits to consumers than they themselves believe

Demerit goods
Goods that are believed to provide more harm to consumers than they themselves believe

Bounded willpower
Occurs when individuals cannot behave in their best long-term interest

Bounded rationality
Reflects an inability to process information sensibly

Merit Goods: Bounded Willpower and Rationality

Merit goods are goods that are believed to provide greater benefits to consumers than they themselves believe. **Demerit goods** are goods whose consumption is believed to provide more harm to consumers than they themselves believe. Such goods arise principally because of inadequate information, bounded willpower or bounded rationality.[11] **Bounded willpower** occurs when people have difficulty implementing strategies that are in their best interests in the long run. **Bounded rationality** reflects incomplete information processing and an inability to weigh up the costs and benefits of alternative actions. In each case, an individual would be better off if she consumed more, or less, of a good than she would ordinarily do.

Inadequate information is not inconsistent with the concept of consumer sovereignty. Consumer sovereignty assumes that (1) personal preferences matter and (2) individuals are the best judges of their own welfare if they have adequate information. Thus, when information is lacking, provision of information is often an appropriate policy response. On the other hand, bounded willpower and rationality are inconsistent with consumer sovereignty. In these cases, the individual can no longer act consistently in their own best interest. Here, provision of information is not a sufficient policy response.

Examples of bounded willpower, or incomplete self-control, include excessive current consumption (too little saving) and addiction spending. Pigou (1920) contended that most people undervalue future consumption even when it is 'perfectly certain to occur'. However, the evidence for this is not clear. Becker and Murphy (1988) famously argued that some people can rationally choose to consume addictive goods such as drugs, tobacco and alcohol. Some drug and alcohol addicts and possibly even gamblers are aware of the consequences but claim that on balance the pleasure from these activities increases their well-being. The view that drug consumption and gambling are demerit goods implies that these claims are misguided.

Evidently, the line between bounded willpower and bounded rationality is often grey. But bounded rationality, or incomplete information processing, also includes situations where individuals rely on rules of thumb rather than data to make decisions, make biased probability judgements or are over-confident or anchor on irrelevant information (Diamond and Vartianen, 2007). A classic merit good and example of bounded rationality is the value of education, where it was traditionally contended that poorly educated households underestimate the benefits of education.[12] Another example would be the failure of gamblers to understand that expected statistical losses often far exceed expected gains, or indeed to understand the real odds of gambling at any point in time (see Box 15.3).

Merit (or demerit) goods should be distinguished from moral goods. For example, some people contend that homosexual acts between consenting adults, abortion and euthanasia should be banned because they are morally wrong. This is a quite different contention to the merit good one that an individual's personal utility would be enhanced if they consumed more of some goods and less of others.

[11] Individuals may also under-value the benefits of some goods, such as cultural goods or heritage, because of a lack of experience or education in them. Such under-valuation does not fall readily into any of these three categories.
[12] Musgrave (1959, p. 14): 'the advantages of education are more evident to the informed than to the uninformed'.

Policies for merit and demerit goods

As we have noted, the policy response to inadequate information is improved information. Other issues arise with failures of consumer sovereignty when individuals may not be the best judges of their own welfare.

Economists have generally considered that most people are the best judges of their own welfare most of the time and been wary of policies justified by reference to merit goods. As John Stuart Mill observed, individuals can learn from their mistakes and there are advantages in treating individuals as if they are the best judges of their own interests. Political judgements about the best interests of citizens may themselves be poorly informed. They may also be moral judgements rather than merit good judgements.[13]

Nevertheless, most people would agree that government should protect the young, the mentally disadvantaged or disturbed and the elderly infirm who lack adequate family support. Many governments have offices to assist with the financial affairs for people deemed incapable of handling these matters themselves.[14] And compulsory education up to a certain age is generally justified as a merit good as well as an equity issue. Taxes on addictive and harmful goods, such as drugs and alcohol as well as on gambling, are widely justified on demerit good grounds as well as externalities. Recently the literature on behavioural economics has greatly expanded the potential areas of consumer areas of consumer failure and correspondingly increased the scope for possible government intervention (Bernheim and Rangel, 2008).

Box 15.3 Gambling and public policy

The Productivity Commission's (2010) report on gambling in Australia identified the main impacts of the industry and implications for government regulation.

The report found that estimated losses exceeded $19 billion in 2008–09. It also found that there were 115 000 'problem gamblers' and that another 165 000 people were at 'moderate risk'.

It estimated that the social cost of gambling is at least $4.7 billion a year. The costs include depression, relationship breakdown and family neglect, domestic violence, crime, lowered work productivity and job loss and the drain on public resources.

Gambling takes many forms. It includes electronic gaming machines (EGMs, called 'pokies'), casino gambling, wagering, lotteries, pools, keno and online gaming. EGMs account for 62 per cent of turnover and for most of the problem gamblers.

Significantly, many players of EGMs have no idea of the real odds or price of playing. Over half the players believe that they can recover losses by continuing to play ('chasing losses') and that machines run 'hot' or 'cold', neither which is true and which, in any case, are inconsistent beliefs. Receipts are not issued, so losers do not know how much they have lost and tend to underestimate losses.

EGMs in particular are associated with significant social costs (negative externalities), information failure and bounded rationality. Notwithstanding the pleasure that many people obtain from gambling, a significant number are causing self-harm. To deal with the variety of problems, the Productivity Commission recommended that:

- The cash that can be fed into EGMs at any one time should be limited to $20 (now $1000)
- The bet limit per button push should be $1 instead of $5-$10
- There should be longer and more time-effective shutdown periods for gaming in hotels and clubs
- Players should be allowed to set binding pre-commitment levels on their losses
- There should be improved information and warnings in EGM venues
- On-line gambling, including offshore sites, requires careful regulation
- There should be more gambling counselling services
- There needs to be ongoing research into gambling issues.

[13] In Australia most state governments prohibit euthanasia. This ruling seems to reflect the moral beliefs of most politicians rather than a merit good view that those who would choose euthanasia to end suffering would be better off if they could not choose this course of action.

[14] For example, in New South Wales, the Office of the Protective Commissioner has this function.

Box 15.4 Obesity and public policy

Crowle and Turner (2010) discuss the costs of obesity, especially in the young, and various possible policy responses.

The weight of Australian children has increased significantly and over 8 per cent are now defined as obese (based on Body Mass Index) and 17 per cent as overweight.

Access Economics (2008) estimated that the total (all age) cost of obesity in Australian in 2008 was $58 billion, with $50 billion being lost health and well-being and $8 billion being financial costs (productivity losses, public health care costs, carers and transfers). Thus a high proportion of costs in this case are personal. There is a correlation between obesity when young and later in life.

However these estimates make no allowance for the pleasures of eating or the pain of dieting. Thus they may overestimate the welfare costs of obesity.

While the basic cause of obesity is an imbalance between energy consumed and expended, the contributions of physical, home, environmental and social factors are hard to determine.

International evidence to date indicates that consumers have limited responsiveness to food taxes, that the link between television viewing and childhood obesity is small in magnitude and that mandatory posting of calorie content in restaurant menus has led to only a small reduction in energy intake, Australian experience suggests that the most encouraging interventions appear to be community-based programs to encourage healthy eating and physical activity, But these interventions could be resource intensive and costly.

Crowle and Turner conclude that there is a lack of clear evidence on both causal factors and cost-effective policy responses and that more research will be needed to improve policies to reduce obesity.

In these circumstances, what advice can an economist contribute? In Boxes 15.3 and 15.4 we describe analyses of the problems associated with gambling and obesity and policies to deal with them. Here we outline a general approach to policy making.

The first task is to identify the nature of the problem(s). Are the problems externalities, information failures or problems of bounded willpower or rationality? Often, as with gambling and obesity, there is a combination of all these factors.

The policy response(s) depend on the problem(s). If the problem arises from a lack of information more information should be provided. The provision of information maximises the exercise of personal preferences and minimises coercion. If the problem is one of negative externalities, fiscal incentives (taxes or subsidies) may be appropriate (see Chapter 13). This approach internalises the impacts and allows individuals to continue to exercise personal preferences. Fiscal incentives may also be an efficient way to encourage a marginal increase or decrease in consumption of a merit or demerit good respectively, although they may have limited impacts on addictive behaviour (including over-eating).

Regulations provide another strategy especially when people face discrete choices about whether to consume a good or not. For example, government may require all persons up to a certain age (say 16 years) to attend school. On the other hand, it may prohibit the consumption of specified drugs, ban young persons from purchasing alcohol or tobacco and restrict places of gambling. Indeed, regulations can be found for most problems. An advantage of regulations is that they can tackle problems directly.

However, regulations usually restrict choices and impose various costs on people. Regulations also have transaction costs associated with administering, monitoring and enforcing the regulations and compliance costs for firms and households. To ensure that the expected benefits exceed the costs, regulations should be subject to cost–benefit tests. However, as can be seen from the studies of gambling and obesity, policy makers are still dealing with many uncertainties.

Summary

- Information failures may occur because of the public good nature of information, misleading information, information asymmetries or bounded willpower or rationality.
- Asymmetric information (when some people are better informed than others) occurs often in product and factor (labour and capital) markets. It causes resources to be misallocated and the less informed party bears unexpected and sometimes high costs.
- Markets provide incentives for product and workplace quality and safety. Firms can charge more for known quality products and pay lower wages for providing safe working conditions. Nevertheless, these incentives are not always sufficient to ensure efficient and fair outcomes.
- Policies for achieving efficient product quality and safety include prohibition of misleading or deceptive conduct, product liability policies, regulation of information, product or process regulation and taxation of unsafe products.
- The main public policies for achieving an efficient level of workplace safety are regulation of information, workers' compensation policies and workplace regulation.
- Government may also regulate occupations to ensure quality of service, but this may reduce levels of service and increase prices.
- Insurance markets may be inefficient because of adverse selection and moral hazard. However, government faces similar problems. Governments may regulate or provide insurance to further social justice objectives but there is a limited efficiency case for substituting social insurance for private insurance.
- Government may promote consumption of merit goods and discourage demerit goods because individuals cannot make informed and rational choices. However, there is a risk that government may impose its own moral values on citizens.
- Fiscal incentives can address some information failures, but regulations address them more directly and are used more often than fiscal instruments.
- Most regulations have costs as well as benefits. They can restrict services and raise prices. Cost--benefit analysis aims to ensure that regulations meet a public benefit test.
- As a practical matter, regulations need to be well-informed, enforceable and not subject to regulatory capture.

Questions

1. Is insider trading simply an equity issue? Does insider trading have a deadweight loss?
2. Can legally imposed damage costs result in inefficient outcomes?
3. Government policy towards insurance:
   i. Should house insurance be made compulsory?
   ii. If a household is uninsured and their house is flooded or burned down in a bush fire, should government provide any compensation?
4. Government regulation of health products:
   i. What kinds of factors should a drug regulator like the Therapeutic Goods Administration in Australia take into account when deciding what medicines should be allowed for public sale?
   ii. Should complementary health products that make therapeutic claims also be regulated? If so, how?
5. What problems arise with determining workplace safety standards by an affordability criterion?
6. What are the costs and benefits of food safety regulation? Can these costs and benefits be quantified?
7. What is the appropriate policy for goods such as cigarettes? Is it to ban false advertising, to tax the product, to regulate where smoking can occur or to produce government advertisements with an alternative point of view? How would a balance between these policies be struck?
8. If workers are fully informed on the risks associated with their workplace, should government also impose conditions on workplace activities and methods?
9. What would constitute an efficient system of workers' compensation insurance?
10. Is it necessary to regulate entry into the hairdressing trade?
11. What is regulatory capture? Is it a general problem? When is it most likely to occur?
12. When and how should government intervene to reduce private self-harm behaviour as distinct from reducing social costs? Discuss in relation to (i) gambling and (ii) obesity. Should an individual's enjoyment of gambling or eating affect public policy?

Further Reading

Bernheim, B.D., and Rangel, A. (2007) 'Behavioural Public Economics: Welfare and Policy Analysis with Non-standard Decision Makers', pp. 7–4 in *Behavioural Economics and its Applications*, eds: Diamond, P. and Vartiainen, H., Princeton University Press, Princeton, New Jersey.

Competition and Consumer Act 2010, Attorney-General's Department, Canberra.

Glaeser, E. and Ujhelyi, G. (2010) 'Regulating misinformation', *Journal of Public Economics*, 94, 247–257.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapter 12, Worth Publishers, New York.

Productivity Commission (2008) *Review of Australia's Consumer Policy Framework*, Final Report, Canberra.

Pyndick, R.S. and Rubinfield, D.L. (2005) *Microeconomics*, 6th ed., Chapter 17, Pearson Education, New York.

Stigler, G.J. (1961) 'The economics of information', *Journal of Political Economy*, LXIX, 213–225.

Stigler, G.J. (1971) 'The theory of economic regulation', *Bell Journal Economics and Management Science*, 2, 3–21.

Stiglitz, J.E. (2000) 'The contributions of the economics of information to Twentieth Century Economics', *Quarterly Journal of Economics*, 115.

Viscusi, W.K., Vernon, J.M. and Harrington, Jr., J.E. (2005) *Economics of Regulation and Antitrust*, Chapters 22 and 23, MIT Press, Cambridge, MA.

Viscusi, K. (2018) Pricing Lives: Guideposts for a Safer Society, Princeton University Press.

Part 6

Public Supply of Goods and Services

Chapter

16

Delivering Efficient Public Services

I see it is impossible for the King to have things done as cheap as other men.

Samuel Pepys, Diary

Issues in the Supply of Public Services ♦ Principles of Public Sector Management ♦ Public Expenditure Budgeting and Management ♦ Estimating the Costs of Services ♦ Organising Service Delivery

Publicly funded services account for about a fifth of the total output of goods and services in most OECD countries. In previous chapters we have examined why and when governments should provide these services. In this chapter we discuss issues in the delivery of efficient public services.[1]

The first part of the chapter outlines some issues in the public supply of services and general principles of efficient management in the public sector. We then turn to issues associated more specifically with the efficient management of public expenditure and with budgets that lie at the heart of the government process. In the second half of the chapter we discuss two core issues for delivery of efficient services: how to estimate the costs of services and how to organise efficient delivery of these services, including whether to deliver publicly funded services in-house or by outsourcing.

Issues in the Supply of Public Services

At one level, efficient provision of public services is simple: identify the benefits and costs of services, prioritise, deliver and regularly evaluate. This is essentially the message of the Productivity Commission's major five-year review (2017). However, public services have two important characteristics. First, many services are intermediate rather than final goods. Examples are school classes, hospital operations or police on the street. In the public administration literature, these services are usually described as 'outputs'. These outputs are, of course, intended to provide benefits (outcomes) to individuals or firms, such as educational skills, improved health and increased security. But, whereas outputs are usually visible, outcomes are less certain. A related second feature of many public services is that, in the absence of market prices, the outcomes have no immediately observable value. This means that there is no clear connection between the value of the outcomes and their cost.

[1] In this chapter, 'services' should be understood to include goods and services. Also, 'public' services should be understood to mean all publicly funded services.

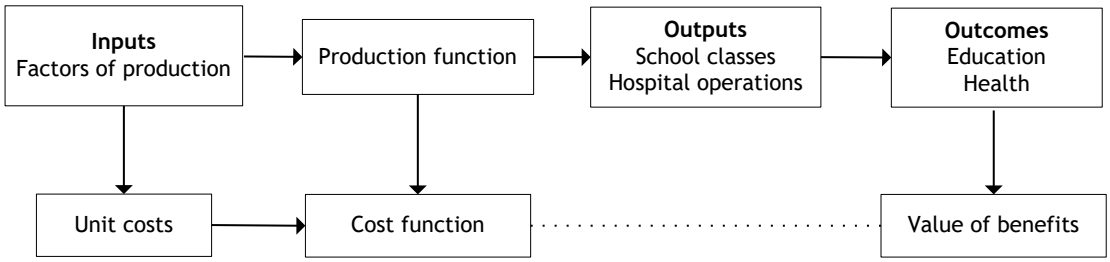


Figure 16.1 Framework for analysing the supply of publicly funded services

Production function
A function that relates the level of output in physical units to quantities of inputs

Cost function
Shows total cost as a function of output

The process of supplying public services, from inputs to outputs to outcomes, is illustrated in Figure 16.1. Integral to this process are the production function and the cost function. A **production function** relates levels of output measured in physical units to quantities of inputs of land, labour and capital also measured in physical units.[2] A **cost function** shows total cost in dollars as a function of the level of output. A cost function is generally derived from a production function (which shows the inputs required for given amounts of output) and the unit costs of the inputs. Thus, a change in the production function will generate a change in the cost function.

To evaluate the supply of public services we need estimates of both the costs of producing the outputs and the value of the outcomes generated from these outputs. However, the identifiable and measurable service deliverables are most often the outputs. Thus, many public expenditure management decisions are based on the costs of delivering outputs in different ways.[3]

The supply process is depicted in Figure 16.2. Panel (a) shows a partial production function, with output as a function of labour inputs. In this panel, inputs of capital and land are assumed to be fixed and output increases at a declining rate as labour inputs are added to these inputs. Panel (b) shows a cost function inclusive of fixed and variable costs. The variable cost here rises throughout at an increasing rate as marginal productivity falls in panel (a). Panel (c) shows a relationship between outcomes and outputs, for example between educational courses and educational attainment. The concave curve implies declining marginal effectiveness of outputs, in this case of additional courses. Panel (d) shows the value of outcomes in relation to physical measures of outcome, for example the value of increasing educational

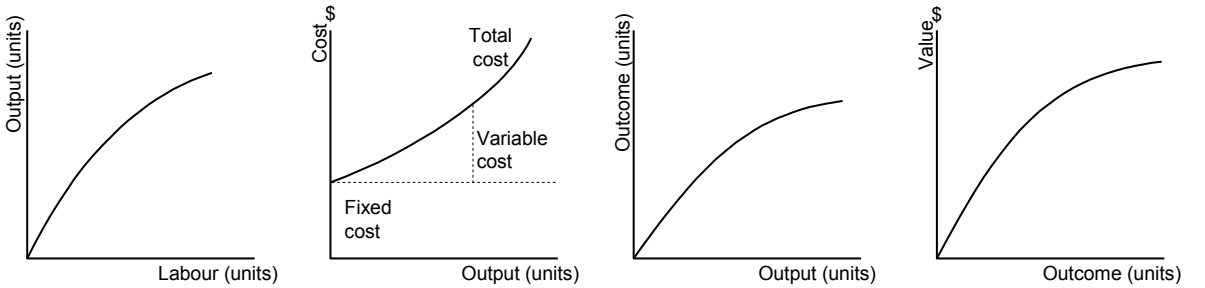


Figure 16.2 Graphical depiction of the supply process

[2] In Chapter 5 we described an aggregate production function for a whole economy.
[3] Methods for valuing outcomes were discussed in earlier chapters, notably in Chapters 6, 8 and 11.

achievement. The concave curve implies declining marginal value of outcomes. Thus, increasing educational achievement would have a positive, but declining, marginal value. Of course, these curves are hypothetical and could take many other shapes. Note also that the slope of the curve at each point in each panel shows the **marginal** change in the dependent variable (on the *y*-axis) as a function of a **marginal** change in the determining variable (on the *x*-axis). Because efficiency in supply requires that the marginal benefit from a change in output should equal the marginal cost, we need to estimate the shape of the curves or the relevant parts of each curve in each panel.

Principles of Public Sector Management

As we saw in Chapter 10, there are significant obstacles to efficient supply of public services. These include monopoly supply, a lack of measurable outcomes and the decoupling of supply from demand. In the 1980s and 1990s, concern about the efficiency of public service provision generated a new management literature for the public sector, described as the New Public Management (NPM) (see Box 16.1). Although this literature was generated mainly by political scientists and public management experts, much of it draws on economics and responds to the government failures discussed in Chapter 10. The following are the main principles of NPM.

Objectives of government agencies should be stated as precisely as possible. Many government departments observed by the writer have unclear, non-quantified, objectives. Without clear and quantified objectives, efficiency is difficult to achieve and impossible to monitor. To achieve this, the New Zealand government broke up large conglomerate ministries into performance-based agencies or business units, headed by chief executives on fixed-term, output-based contracts with authority to hire and fire personnel. Singapore also has performance-based statutory boards.

Government should serve consumers rather than producer interests. A corollary of this principle is that, where feasible, citizens should have choice of service. For example, parents should be able to choose schools for their children (and whether their children should attend scripture classes). Low income households receiving government support should be able to choose their own housing.

Government activities should be separated according to functions. Public service was divided traditionally into policy making and administration. NPM proposes that the major administrative functions (revenue raising, purchasing of services, delivery of services and regulation) should also be separated. In particular, funding and delivery of programs should

Box 16.1 New Public Management (NPM)

In 1987 the New Zealand Treasury produced a systematic exposition of the new ideas in public management (*Government Management: Brief to the Incoming Government*). Over the next 10 years the NZ government implemented most of the proposals (see Boston *et al.*, 1996). Osborne and Gaebler (1993) espoused similar ideas *in Reinventing Government*, which influenced the Clinton Administration in the United States. Lane (2000) provides a useful summary of the models of public sector management and their implementation in many OECD countries.

In Australia the National Competition Agreement, signed in 1995 by the Commonwealth and state governments, embodied some of the principles of NPM. The report of the National Commission of Audit (1996) also recommended sweeping changes in government practice. As described in Chapter 14, National Competition Policy has had a significant effect on the Australian economy. However, arguably, Australian governments have tended to adopt the recommended efficiency practices piecemeal rather than with total enthusiasm.

be separated—this is known as the purchaser–provider split. This facilitates monitoring of work and competition among suppliers of services and helps to ensure that suppliers do not determine policy or choice of supplier. Also, regulation should be separated from service provision. Referees should not be players. Institutional arrangements should minimise the scope for regulatory capture. For example, an agency supplying water and sewerage services should not also be the agency for environmental protection.

Activities should be separated according to responsibilities. The responsibilities of ministers and departmental chief executives should be separate and distinct. Ministers should determine the outcomes to be pursued and provide the funds to achieve these outcomes. Departmental heads should be responsible for producing these outcomes with minimum interference from ministers and central agencies. A core mantra of NPM is 'let the managers manage'. But this must be subject to checks and balances (review and accountability). Without regular review and accountability, managers run an unaccountable monopoly service.

Government should adopt best practice management. Government should supply services as efficiently as possible. The public service should not itself be a welfare service. Employment relations in the public sector should draw on the more efficient flexible relations in the private sector. This may involve the use of contract work, fixed terms and remuneration scales to attract appropriate staff. This would be combined with use of individual incentives.

Where feasible, government should employ market pricing rather than taxes. User charges for services and cost recovery encourage financial and management disciplines. Prices improve measurability and provide information about consumer preferences. However, we should note a caveat here about pricing of welfare services (see Chapter 17).

Public supply should be provided competitively where feasible. Service deliverers, whether public or private, should compete by tender or other methods for the right to provide publicly funded services. Subject to appropriate specifications of work required (see further discussion below), this helps to ensure service efficiency.

Provision of services should be decentralised. Decision making should be located as close as possible to the place of implementation, subject to an efficient scale of service being achieved. Information cannot travel quickly up and down a large centralised hierarchy—there is commonly administration overload. A local agency is better informed about local preferences and more likely to provide customers the services that they want. Local provision provides more variety of services. Also, costs are more transparent in local administrations.[4]

Improve transparency: where subsidising, do so directly and transparently. In *Yes Minister*, when Jim Hacker (the newly appointed Minister for the Department of Administrative Services) meets Sir Humphrey Appleby (Permanent Secretary of the Department), Hacker observes that 'Opposition's about asking awkward questions.' 'Yes', Sir Humphrey replies 'and government is about not answering them'. By contrast, transparency is another mantra of NPM. For example, government subsidies should be provided by explicit grants to consumers or by 'community service obligation' payments to organisations that provide deemed community services, such as public transport or telephones. Subsidies should not be supplied by tax expenditures or by cross-price subsidies. These hidden payments cannot be readily assessed.

[4] As someone who has over 12 years' experience in working with both state and local jurisdictions in Australia, I would strongly endorse these generic observations on decentralised decision making and services.

Suppliers of public services should be accountable. There are three main principles for effective accountability. First, responsibilities and outputs must be specified. Agencies should publish regular plans that explain proposed outputs (and where possible outcomes), strategies to improve output, and progress against these plans. Output measures are an essential basis for performance appraisal. This facilitates benchmarking as an aid towards best practice productivity. Second, there should be freedom of information. The general principle should be that information paid for by taxpayers belongs, and is available, to them other than for clearly defined exceptions, for example for public security and occasionally details in commercial-in-confidence contracts. Watchdogs such as the Auditor General and parliamentary committees should have full access to government information. Third, provision and monitoring of services should be separated. There should be independent auditors of powerful departments, such as police, health and education departments.

Public Expenditure Budgeting and Management

Efficient management of public expenditure has three major components:

1. Aggregate fiscal discipline—total government expenditure should be set before decisions are made on the separate parts of the budget, and this limit should be maintained unless circumstances change. Expenditure should not simply accommodate the demands of spending departments.
2. Allocative efficiency—public expenditure should be allocated to the highest priorities.
3. Operational efficiency—public services should be delivered at least cost.

The central instrument of expenditure management is the **government budget**. The principles and practice of budgeting are critical to efficient provision of public services.

The basic principles of budgeting are that budgets should be comprehensive (including all revenue and expenditure), accurate (recording actual transactions and flows), cover a defined period (usually a year), be authorised appropriately and transparent (government should publish timely information about estimated and actual flows). These principles in conjunction with the formal procedures for developing the budget and for auditing expenditures comprise due budgetary process. Typically, due process is achieved by centralising the management and control of public expenditure.[5]

In practice, central Treasury departments have traditionally emphasised expenditure control and funded each government agency on an annual basis for their expected use of inputs rather than for supply of outputs. Budgets itemised the amounts that departments could spend on each category of inputs. Line-by-line allocations identified the allowable expenditure for regular employees, casual workers, overheads such as workers' compensation, equipment and supplies, travel expenses, materials and so on, often in considerable detail. Agencies were expected to conform to these line-item budgets. Expenses could be audited before or after expenditure. With a pre-expenditure audit, each expense is vetted in advance, which is burdensome. Post-expenditure audit is based on a sampling of expenditures after they have been incurred to ensure that the process works properly.

This traditional budgeting process establishes the basis for effective financial controls and management of public expenditure. However, it has significant limitations. This process encourages spending by the rules, conforming to budget estimates, getting approval before taking actions and, above all, spending all allocated funds. Relatively little attention may be paid to prioritising expenditures in relation to outcomes. Typically, agencies get more funds by requesting incremental funds rather than by shifting funds from lower to higher priorities.

[5] The discussion in this section draws especially on Schick (1998).

Once established, public programs often escape systematic and rigorous evaluation. Such budgets, based on input controls, are not related closely to outcomes. Nor do they provide spenders with incentives to economise. Expenditure is audited on the basis that spending is duly authorised rather than in terms of what it has achieved.

Public expenditure management

Governments have tried to deal with the limitations of traditional budgeting in various ways. In the 1980s and 1990s, governments introduced various budgeting instruments, described as **program budgeting, planning–programming budgeting systems** (which included evaluation of program budgets), and **zero-based budgets** (which attempted to avoid incremental budgeting by requiring that budgets be developed from scratch each year). The common aim was to relate budgets to outputs and thereby to integrate program evaluation and financial management. However, limited success was achieved. The aims of the reforms were too ambitious for the instruments available. Programs often ran across departments. The information needs of program evaluation and management were considerable and hard to integrate with line-item, agency-based accounting systems. Departments threatened with loss of control over expenditure decisions to central agencies or loss of resources resisted changes. Accordingly, many governments continued to use traditional financial control methods and to generate program data on an 'as needs' basis.

In recent years, many governments have rebadged the reforms as **performance or output budgeting**. With output budgeting, agency funding depends on performance of an explicit stipulation of the outputs that the agency should deliver. For example, payments for hospital services in many parts of Australia are now based on provision of services based on "defined diagnostic related" groups. Output budgeting is consistent with the purchaser–provider principle, in which the function of purchasing (financing) services is separated from that of supplying the services.

Under output budgeting, central agencies (the purchasers) determine strategic objectives and priorities and inter-sectoral strategic expenditure claims. However, delivery agencies are responsible for intra-agency expenditure decisions. Under the key principle of **managerial accountability**, managers are accountable for delivering pre-specified outputs. Managers receive global budgets based on the expected costs of the outputs to be delivered. They decide on the mix of inputs to deliver the outputs. They are responsible for delivering the outputs and for controlling costs. Managers need budget flexibility and control over resources. They are accountable for program results. Spending must be justified by what it achieves.

Another important management principle is **contractualism**. The principal who funds the services should have an explicit service agreement with the agent who supplies them. Contractualism decouples policy from operations. Performance budgeting aims to overcome principal–agent problems, reward good performance and encourage effort. However, to be effective, performance measures must be clearly defined and easy to observe. In particular, performance budgeting must deal with three important issues: output heterogeneity, contingent capacity and cost containment incentives (Robinson, 2002). Outputs should be defined clearly in terms of quality as well as quantity. For example, university outputs may be gauged by the number of effective full-time students enrolled or by the number of research publications. But if there is no allowance for the quality of the educational process or the publications, these performance measures may be misleading. Second, excess capacity may be required for some services, especially for emergency services. Third, if budgets are determined by outputs the agency may have more incentive to expand services than to contain costs. Separate mechanisms are required to provide incentives to control costs.

Accrual accounting is also important for efficiency. As described in Chapter 2, accrual accounting records expenses and receipts when they are incurred, not when cash is exchanged. Accrual accounts also include the costs of using capital equipment. Annual

balance sheets include actual and contingent liabilities. Accrual accounting shows how much an agency is really spending. It also improves transparency and efficiency in use and maintenance of assets. However, accrual accounting can be misused. Operating expenses and asset valuations depend on depreciation estimates which require technical judgements and which may be distorted to achieve a preferred paper outcome.

Schick (1998) summarises the instruments for achieving operating efficiency as follows. Managers are given global operating budgets—a single allocation for all operating expenses. Line officers should also have devolved global operating budgets. Accountability of outputs is maintained by a series of performance agreements and contract-like documents. There is an efficiency dividend. Agencies should receive part of the savings they achieve (providing that they are due to lower costs, not lower outputs). Outputs are specified in the budget or related documents. Purchasers and providers are separated. The costs of supplying services should be market tested by comparisons with the costs of outside suppliers. Agencies should regularly report quantified outputs (generally called key performance indicators, KPIs) and outcomes. KPIs would typically include measures of volume of services, timeliness, quality (service errors), recipient satisfaction and unit costs. And there should be independent audits.

Conclusions. Traditional budgets emphasise input controls and financial process rather than policy analysis. Current public expenditure management emphasises outputs and outcomes. The move to performance management is designed to increase allocative and operational efficiency. The reform bargain is managerial freedom over inputs in return for accountability over outputs. Implementation requires performance contracts.

However, problems remain. First, a true performance budget is a variable budget. But variable budgets are hard to design. Budget appropriations are generally fixed limits on expenditure but not tightly related to performance. Second, in so far as budgets are related to performance, they are generally based on outputs, which can be measured and verified, rather than outcomes. The link between outcomes and services is often hard to demonstrate. For example, it is hard to show whether public health programs change individual behaviour and improve health outcomes. Also, outcomes are often hard to value. Third, performance budgets require accurate information on costs and outputs. However, government agencies control the information flow to the central agencies and have little incentive to minimise costs if savings accrue to Treasury. Also, the contribution of individual public servants to outputs is hard to identify. This may explain the slow development of effective in-house performance contracts. As Box 16.2 indicates, Australian practice falls short of the aims of performance budgeting.

Estimating the Costs of Services

Estimates of the costs of services are fundamental to efficient delivery of services. However, before we can estimate costs, we need to recognise that there are various kinds of costs and to identify those that matter for the relevant policy decision.

The following are the key categories of costs.

- Total cost: fixed costs plus variable costs.
- Fixed costs: costs that do not vary with output.
- Variable costs: costs that vary with the level of output.
- Average or unit cost: total cost divided by the number of units supplied.
- Average variable cost: total variable cost divided by the number of units supplied.
- Short-run marginal cost: the cost of producing an additional unit of output with fixed capital stock.
- Long-run marginal cost: the cost of producing an additional unit in the long run inclusive of the cost of additional or new capital stock.

Box 16.2 Performance budgeting in Australia

In 1999–2000 the Australian government introduced accrual accounting and an explicit purchaser–provider model involving an output and outcomes framework. For this budget, 45 major agencies accounting for nearly all Commonwealth expenditure were provided with an accrual-based budget, inclusive of depreciation allowances, capital injections, estimates of long-term liabilities and capital charging for assets at 6 per cent. The government defined outcomes as the key results that the government sought to achieve. Outputs were the goods and services that government purchased from the agencies to achieve these outcomes. Once quantities of outputs were determined, the agency and the Department of Finance and Administration (DoFA) agreed a set of prices or costs to achieving these outputs. In the words of DoFA (2000), the:

> *outcomes and outputs are at the centre of how agencies plan, budget, manage and report … Under accruals, agencies are resourced for the price of their outputs—what they produce to contribute to outcomes. The agency output price will include full costs, such as depreciation and employee leave entitlements.*

Although the reforms were intended to improve expenditure efficiency, it is not clear whether real improvements in resource allocation or expenditure control have been achieved. A major problem with the new process is the high-level nature of the proposed outcomes.

For example, in the 2002–03 budget, Outcome 1 under Population Health and Safety is 'promotion and protection of the health of all Australians and minimising the incidence of preventable mortality, illness, injury and disability'. For the Department of Education, Science and Training, Outcome 1 was defined as 'school systems that provide their students with high quality foundation skills and learning outcomes'. Other proposed outcomes in the health and education portfolios are equally vague.

By contrast the UK public service agreements set much more precise outcome targets. For example, there are numerical targets for reducing mortality rates from heart disease and cancer and for improving the literacy and numeracy outcomes for schoolchildren (UK Treasury, 2002).

Hawke (2007) described how the Commonwealth agencies regularly report on a range of key indicators and how the Department of Finance has developed a set of questions to diagnose the quality of performance and improvement over time. Portfolio budget statements include details of uses of funds by outcomes. However, he noted that a key challenge remains 'to ensure that the links between programmes, outputs and outcomes are clear and measured effectively'. Also, limited use was made of performance information for decision making for the budget. But it should be noted that this view is dated and practice may have improved.

The cost of interest depends on the issue. Fixed cost, total cost and average cost include past or sunk costs. Variable and marginal costs are forward looking. Past costs affect balance sheets, debt and financing costs. However, current policy decisions should be based on present and related future costs. Sunk costs are not relevant. Of course, an estimated total cost function provides information implicitly on most kinds of costs.

We should also recall the differences between financial and real costs. Financial costs are the monetary costs to government in cash or accrual terms. The real cost of employing resources (factors of production) is their opportunity cost (the value of output forgone). This may be greater or less than the financial cost. In budgeting, government agencies are generally most concerned with financial costs. However, the real (opportunity) cost of production is the relevant cost for determining service delivery. For example, the use of public land has an opportunity cost but often little or no financial cost.

Service costs can be estimated in two main ways. The most common method is to estimate the quantities of inputs required for supply of a service (essentially to estimate a production function) and then to apply relevant unit costs to estimate the costs. The inputs required can be estimated in various ways, notably by experts in the relevant production activity, statistical estimates of a production function or activity-based costing. Alternatively, it may be possible to estimate a cost function directly from financial and output data.

Estimating production functions. Estimates of the inputs required for a given service or product are typically provided by experts in the relevant area drawing on a combination of experience and data. For example, a quantity surveyor may estimate the material and other

inputs required to construct a building. An educational expert may estimate the labour and other inputs required to run schools of various sizes. This approach requires detailed knowledge and specification of the relevant production activity, often gained from data in similar situations. Once the inputs have been estimated, the cost is the product of these inputs and their unit costs.

Generally, a production function may be estimated from relevant data as follows:

$$Q = f(K, L, X_i) \tag{16.1}$$

where Q is a physical measure of output, K is capital inputs, L is labour and X_i is a vector of other $i = 1 \ldots n$ inputs. For example, Q could be bus passenger kilometres, K the number of buses and L the number of bus company employees. The capital units may be disaggregated into buses of various sizes and the labour units into various kinds of labour, for example bus drivers and other employees. Equation 16.1 could be estimated from either time series or cross-sectional data. Hensher (2007) reviewed production functions for urban buses. Box 16.3 overleaf provides some further details on production functions.

More sophisticated techniques for comparing production efficiency across agencies have been developed. For example, stochastic frontier analysis and data envelope analysis aim to identify the production frontier, that is, the maximum feasible output given various input quantities. The methods then aim to provide measures of shortfalls between observed outputs and the maximum feasible output. These are measures of relative efficiency (Coelli *et al*., 1998; NSW Treasury, 2001).

Activity-based costing
The allocation of an agency's inputs to its outputs and estimating the costs of the inputs

Activity-based costing. Using this method, an analyst allocates an agency's inputs to its outputs, usually based on a survey of an agency's work over a sample period, such as two to four weeks. The analyst then uses the agency's financial data to estimate the cost of services. Typically, this process involves three main steps:

1. For each line section of the agency that produces final services, determine the inputs (labour, materials and so on) that vary with the level of each service and attribute these to that service.
2. Determine the fixed inputs of each line section (such as line section administration and equipment) and allocate these inputs to the services.
3. Allocate the overhead costs of the agency (such as property rent) and the costs of internal service agencies (such as IT services) to final line service sections by use of some formula. For example, rent may be allocated according to the amount of space occupied by the line section. The line agency then distributes these costs to its services.

When all costs of an agency are accounted for, costs are described as 'fully distributed'. The process can involve several steps. Take, for example, a hospital. It may be quite straightforward to estimate the variable and fixed costs associated with a ward; however, to allocate central hospital expenditures, such as accounting, laundry and catering, to final service areas such as wards, theatres and so on, it may be necessary first to allocate part of these expenditures to service departments such as radiology, pathology and cleaning. The costs of these service departments must be allocated in turn to hospital wards and theatres, along with the direct nursing and material costs. After the total costs of wards or theatres are estimated, unit costs can be estimated by dividing total cost by activity level.

There are three main difficulties with activity-based costing. First, standard budget line accounting shows costs incurred but not how inputs are allocated. Accordingly, sample period surveys are required to determine how inputs are allocated to services. The results may reflect the period chosen. Second, book costs may not reflect real (opportunity) costs, especially for use of capital assets. Third, financial accounts do not specify which costs are fixed and which

Box 16.3 Estimating production functions

The Cobb–Douglas production function, which we introduced in Chapter 5, is a commonly used production function. The specific form of this function is: $Q = KL^aK^b$, where Q is output, K and L are units of capital and labour respectively, and A, a and b are constants ($0 < a$ and $b < 1$).

The exponents a and b represent the elasticity of output with respect to L and K respectively. When $(a + b)$ is equal to one, this production function exhibits constant returns to scale, which means that output changes proportionately with any change in inputs. When $(a + b)$ is greater (less) than one, there are increasing (decreasing) returns to scale.

The logarithmic form of the Cobb–Douglas production function for firm i can be written as:

$$\log Q_i = A^1 + a \log L_i + b \log K_i + u_i$$

where u_i is the stochastic disturbance term, which accounts for variations in the productive capabilities of the i^{th} firm. Subject to the availability of the data on Q_i, L_i, K_i, the parameters A^1, a and b can be estimated directly from cross-sectional or time series data, assuming the parameters and prices are the same for all firms.

Estimates of the production function may be handicapped by poor data on capital inputs. Another problem is the aggregation of the capital used in production due to its diversity of components. The imputed depreciation value of capital may be problematic.

Direct estimation of the production function may also be problematic if the exogenous variables $\log L_i$ and $\log K_i$ are jointly determined with $\log Q_i$, which leads to a problem of simultaneous equation estimation. If L_i and K_i are dependent on each other, there may be a multicollinearity problem. If the variance of u_i is not constant, a heteroskedasticity problem may arise.

An alternative approach is to assume constant returns to scale (i.e., $a + b = 1$). The production function can then be written as:

$$\log Q_i = A^1 + a \log L_i + (1-a) \log K_i + u_i$$

With some manipulation this can be represented as output per worker as a function of the capital to labour ratio:

$$\log \frac{Q_i}{L_i} = A' + (1-a) \log \left(\frac{K_i}{L_i}\right) + u_i$$

This form of equation reduces the problems associated with multicollinearity and heteroskedasticity. However, because of the assumption of constant returns to scale, we cannot test for increasing or decreasing returns to scale.

For empirical work, the constant elasticity of substitution (CES) production function is also commonly used. The general form of the CES production function is $Q = A[bK^c + (1-b)L^c]^{1/c}$, where A, b, c are constants. The ratio of the proportional changes in quantities of inputs to the proportional changes in relative prices is the elasticity of substitution. When $c = 1$, the elasticity of substitution is 1 and the CES production function becomes a Cobb–Douglas function. For more details see Intriligator *et al*. (1996, Chapter 8).

ones vary with output. More analysis is needed to obtain a breakdown between fixed and variable costs. The analyst can then estimate a cost function by extrapolation based on estimated variable costs as a function of output.

Estimating a cost function directly. It is sometimes possible to estimate directly the relationship between cost and output from a sample of operating experiences, using statistical techniques such as multiple regression analysis. The analyst typically draws on observations of costs and services from time series data for a single agency or cross-sectional data for several agencies to estimate a cost function. In both cases the cost data are generally based on accounting measures of cost. Time series analysis for a single agency is usually interpreted as providing a short-run cost function as capital assets do not change significantly between periods. On the other hand, cross-sectional data from several agencies yield long-run cost functions if it is assumed that producers have adjusted their fixed assets to optimal levels to meet demand.

However, interpretations must be made carefully. An underlying problem with both time series and cross-sectional analyses is that the quality of services may vary over the sample period or firms. In addition, the usual problems of multiple regression analysis occur. Various functional forms of the relationship may have to be tested (see Box 16.4 overleaf). Also, the appropriate variables need to be included in the analysis. If this is not done, cost differences may be ascribed wrongly to output levels instead of to other relevant factors.

Box 16.4 Estimation of cost functions

We have seen in Box 16.3 that, for estimation of production function, output is assumed to be endogenous and the inputs are exogenous. For estimation of cost functions, cost and input quantities are assumed to be endogenous and output exogenous. The Cobb–Douglas cost function, which is related to the Cobb–Douglas production function, can be written as:

$$C = f(w, r, Q)$$

where C is total cost, Q is the output, w is the wage rate and r is the return to capital. The elasticity of the cost function with respect to output is the reciprocal of the degree of homogeneity of the production function.

A function is said to be homogeneous to the degree n if multiplying all the independent variables by a constant λ multiplies the value of the dependent variable by λ^n. The elasticities a and b can be estimated from the linear model (considering firm i):

$$\log C_i = B' + \frac{a}{a+b}\log w + \frac{b}{a+b}\log r + \frac{1}{a+b}\log Q_i + u_i$$

Similarly, a cost function related to a CES production function can be defined and estimated.

Benchmarking
Comparing an agency's costs with the costs of similar providers

Efficiency comparisons. The cost-efficiency of service delivery may be assessed in three main ways: by detailed micro assessment, benchmarking and market testing.

A detailed micro assessment of an agency's operations may be made by personnel employed in the service, who are often aware of ways to make operations more efficient, or by an independent external expert who would review the operations.

Benchmarking involves external comparisons with other agencies. For example, the Productivity Commission provides an annual report on the costs of a wide range of government services across Australian jurisdictions (see Productivity Commission, 2011b). Benchmark comparisons range from the informed to the not very precise. The former may be based on detailed costing studies across a range of similar agencies and subject to detailed statistical analysis. Given adequate sample size, statistical analysis may provide a cost model, which shows total or unit costs as function of several variables, along with estimated variances. For cost comparisons to be meaningful, outputs and production environments should be similar. Canada is a more relevant comparison for Australia for, say, transport or telecommunications costs than is Hong Kong. Also, the output and cost data must be reliable, often in the absence of markets for the factors of production services provided. Reliable information may be hard to obtain if public agencies have an incentive to conceal the true costs.

Market testing, via contracting for services, is a major response to these problems. It introduces competition, benchmarks and increased cost transparency.

Organising Service Delivery

In this section we discuss the role of contracts in service delivery, the choice between in-house delivery and outsourcing of services, and user cost models.

The role of contracts in service delivery

Government employees who supply in-house services have employment contracts. Under such a contract the employer usually has exclusive power over the work and output of employees in specified working hours (and sometimes even outside these hours) and can usually require the employee to work as directed. Government agencies may complement employment contracts with specific service duty statements. However, the services are often specified in broad terms.

On the other hand, under a service contract, typically an outsourced contract, firms or individuals are engaged to provide specific services. The contracted agent can usually provide

services to other contractors where there is no conflict of interest. A service contract is usually, but not always, for a shorter period than an employment contract.

Following Schick (1998), there are five conditions in efficient contracts.

1. The two sides must have an arm's length relationship.
2. The purchaser must be able to purchase the services from alternative suppliers.
3. The supplier must have freedom to produce the contracted services.
4. The contract must specify the cost of the services.
5. The contract must specify the performance required of the supplier.

These contract conditions can be achieved with in-house or with outsourced contracts, but are usually easier to achieve with the latter.

Outsourced service contracts may be fixed-price or cost-plus contracts. A fixed-price contract stipulates the work to be done for a fixed fee. It may allow for cost escalation for inputs based on a recognised price index and may stipulate rates payable if variations in output are required. A fixed-price contract provides incentives to contractors to economise on inputs. On the other hand, the contractor bears all or most of the risk, depending on the details of the contract specification. Typically, the contractor builds the cost of risk into the offer price for the fixed-price contract.

With a cost-plus contract the purchasing agency pays all costs incurred by a firm in completing a project. In this case the contractor does not charge for bearing risk. The government can control costs by negotiating low work payment rates, by progressively specifying the work to be done and by carefully monitoring the work. However, the contracted firm has an incentive to increase inputs and costs rather than to reduce them.

If a purchasing agency wishes to minimise the total cost of service delivery, it would select a cost-plus contract (competitively delivered) when it can clearly specify and monitor the work and thus control the risks. Alternatively, it could select a hybrid contract in which some costs are fixed and other costs that can be specified and monitored are paid for on a cost-plus basis.

In practice, most public agencies prefer to pay for risk via a fixed-price contract rather than run the political risk of a cost-plus contract. The main exception arises when a contracting agency cannot specify all the work in advance due to various uncertainties, as in technical design work. Cost-plus contracts are often used for professional advice, such as legal work, and development of military hardware. However, this work needs to be carefully monitored to avoid cost blow-outs.

Choosing between in-house and outsourced services

The two main economic arguments for in-house supply of public services are achieving economies of scale and minimising transaction costs. Where economies of scale exist, one supplier can provide services more cheaply than can two or more suppliers. There is limited scope for ongoing competition and if a private supplier is established there is a risk of establishing a private monopoly. Transaction costs are the costs of negotiating, monitoring and enforcing contracts. Outsourcing raises these costs. Surveillance of private firms is harder than surveillance of in-house employees. In contrast, the continuous and vertically integrated nature of government organisation reduces the need for regular formal contracts, transaction and monitoring costs (Khursheed and Borcherding, 1998).

There are two related sets of costs associated with outsourcing services. First, when services are difficult to describe, and contracts cannot be specified clearly, transaction costs escalate. When contracts are difficult to monitor, costs cannot be controlled and quality of service may fall. Firms may take advantage of this lack of surveillance by not fulfilling their contracts or by exploiting loopholes in contracts to cut services or seek extra payments.

Second, potential costs also increase when contracting out is feasible only for long periods, in which case a private contractor may exploit an effective monopoly supplier position. Problems also arise when contractors default and supply is disrupted.[6]

On the other hand, outsourcing as some significant advantages. First, most public agencies suffer from structural problems, such as rigid appointment procedures and wage setting. Private agencies tend to be more flexible and innovative (Rainey and Chun, 2005). Also, links between behaviour, performance and rewards are greater in private firms than in the public sector. Thus, it may be possible to define operational outcomes and to provide financial incentives related to achieving these outcomes.

Lundsgaard (2002) noted the following related advantages. Contracting

- focuses the attention of service suppliers on core objectives and user needs,
- provides suppliers with clear incentives to operate efficiently,
- reveals best practices so that slack cannot be hidden, and
- makes budget constraints harder.

In addition, competition for contracts allows the purchasing agency to obtain best value for money. Even if there is to be a sole supplier, a competitive auction for the right to provide this supply may produce significant cost savings.

Savings in supply costs with outsourcing. Turning to the evidence, many studies have shown that private firms generally supply services more cost effectively than do public agencies. Borcherding *et al*. (1982) reviewed 52 studies of productive efficiency in 19 economic sectors in five OECD countries, mainly in Germany and the United States, and concluded that in 40 studies private supply was 'unequivocally more efficient'. Only in three cases was public supply less costly than private. Khursheed and Borcherding (1998) and Mueller (2003) cite many studies of the cost advantages of the private sector, mainly from the 1970s and 1980s. Lundsgaard (2002) provided more recent case studies. Table 16.1 gives results from Bailey (2002). These results show that competitive contracting of services traditionally supplied by the public sector, such as waste collection, office and street cleaning, water supply, printing and road and park maintenance, produced large savings often of 20 per cent or more.

In Australia, Domberger (1994) and Domberger *et al*. (1995) reported that contracts let by NSW Departments of Transport, Education, Health and Housing reduced costs by an estimated average of 20 per cent compared with the cost of in-house services. The Productivity Commission has also reported significant cost savings from contracting out services in various reports. For example, the Productivity Commission (2002) found that the 200 private members of the Job Network provided public employment services, such as job matching and job searching, at significantly less cost than the Commonwealth Employment Service that they replaced, although the effectiveness of the service was similar. Abelson (2005) found significant cost savings and service quality improvements in seven case studies of outsourcing by local governments.

On the other hand, contracting for services is not always efficient. Box 16.5 (p. 291) describes a costly failure by the Australian government to achieve a major outsourcing initiative for information technology.

---

[6] Private firms use various methods for dealing with long-term contracts. These include price adjustment mechanisms, minimum purchase requirements, methods for monitoring quality and dispute resolution procedures. Public agencies can use similar methods.

Table 16.1 Cost savings from competitive contracting: international evidence

Country	*Activity*	*Reported savings*
Australia	Water supply	Estimated potential cost savings of 15 per cent
Canada	Refuse collection	Public collection up to 50 per cent more costly
Denmark	Fire services	Public provision almost three times more expensive than private contractors
Germany (West)	Office cleaning	Public sector provision 42–66 per cent more expensive
Japan	Refuse collection	Municipal collection 124 per cent more costly
Sweden	Road and park maintenance, water supply, sewerage	Cost reductions of 10–19 per cent in several municipalities
	Waste collection	Average cost reductions of 25 per cent
	Leisure activities	Cost reductions of 13–15 per cent
	Child care	Cost reductions of 9–15 per cent in nurseries
Switzerland	Refuse collection	Costs of private contracts 20 per cent cheaper
UK	Domestic services in local government and the NHS	Saving of 20–26 per cent, costs being reduced by more than a third in some cases
	Refuse collection	Cost reductions of about 20 per cent
	Central government (e.g. IT and printing)	Average saving of 25 per cent
USA	Refuse collection	Savings of 29–37 per cent
	Street cleaning	Savings of up to 43 per cent
	Office cleaning	Savings of up to 73 per cent
	Federal government	Cost savings of up to 35 per cent with an average saving of around 20 per cent
	Mass transit	Potential savings of 20–50 per cent

Source: Bailey (2002) based on Parker (1990) and OECD (1993).

Most studies find that the key driver of the cost savings is competition. Competitive tendering ensures that the lowest cost producer supplies the service and provides data on costs and efficiency benchmarks. Some economists argue that, given both competition and information on costs, public services can significantly increase their productivity and lower costs and can be as efficient as private sector. For example, Hoxby (1994) found that when public schools in the United States compete strongly with private ones, both educational attainment and graduation rates improve.

Welfare gains. It is also important to note that financial savings to government from competitive tendering may not equate to welfare gains. Financial savings are welfare gains in so far as they reflect productivity improvements, which release resources to other sectors. When competitive tendering produces financial savings only by reducing wage rates, there is no efficiency gain. In this case, outsourcing represents a loss of rent to public employees. The savings here represent a transfer from the workers providing the service to the purchaser of the service (in effect taxpayers).

Box 16.5 Outsourcing can go wrong

In the 1997–98 budget the Australian Government announced that all information technology (IT) infrastructure and telecommunication requirements of its agencies would be consolidated in common groups to be offered to the market. The initiative was based on a centralised model, in which the needs of organisations of varying sizes and differing functions were clustered together for the purposes of tender. The government forecast that consolidation and outsourcing would improve service delivery, achieve economies of scale, improve management of IT services and deliver cost savings of $1 billion over a seven-year period.

The initiative met with strong bureaucratic resistance ranging from opposition to change to concerns about the security of sensitive information relating to personal privacy, scientific research and intelligence.

With the outsourcing initiative three times over budget and two years behind schedule, the government commissioned an independent review. The review (led by Richard Humphrey) found that IT outsourcing had been exercised without adequate regard to the sensitive risks and complex processes of transition and without adequate regard to the business needs of each government agency.

Despite these criticisms, the Commonwealth government remained committed to outsourcing IT services. However, responsibility for outsourcing has reverted to the individual agencies.

Conclusions. Competitive tendering encourages clear specification of the services required and ensures that services can be delivered most cost effectively. Occasionally, in-house suppliers may win contracts; more often, competitive tendering results in outsourcing.

Services can be outsourced most efficiently when the key elements of supply (quantity, quality, delivery times and so on) can be specified precisely, transaction and monitoring costs are low, and the service can be competed for on a regular basis. Outsourcing works well for routine services (such as street or office cleaning, vehicle maintenance, printing and waste collection) for which complete contracts can be written. It can also work efficiently for more complex services such as property maintenance, construction of roads, hospitals and schools, management of recreational and sporting facilities, and for specialised services such as architectural design, traffic modelling, or some policy advice, which can be contracted on an as needed basis. Indeed, the scope of contracting for services is large.

Outsourcing is not efficient when contracts are vague or incomplete or involve numerous contingencies. Thus, the administration of foreign policy could not be contracted out, although running small consular offices can be. Also, functions essential to the continued running of government would rarely be outsourced. Governments needs core management and technical skills, the capacity to manage contracts, continuity and corporate memory.

Finally, the politics of outsourcing should be recognised. Outsourcing may upset groups, notably politicians and public sector workers, who have a vested interest in public production. Thus, to achieve outsourcing, the efficiency benefits must be clearly demonstrable. Some elements in government may also be inclined to closed non-competitive arrangements. Indeed, sometimes government may be captured by private business interests. Such arrangements are generally contrary to the efficiency objectives of outsourcing.

User choice models

User choice models are another way to open service provision to competition. Service users can choose between service suppliers who are reimbursed according to the quantity and kind of services that they provide. Service providers compete for clients. Prices are fixed, but service providers compete in quality and relevance of service.

Alternatively, government can provide service vouchers to users. For example, government may provide vouchers for housing rents to low-income households or vouchers for public transport services to senior citizens. The service provider receives from government the income equivalent of the vouchers it receives. In this case service providers can compete on price as well as quality of service.

Consumers gain directly from user choice because they can choose the service that maximises their welfare. They also gain indirectly from the service and cost competition among suppliers. On the other hand, when government provides funds direct to service suppliers, it may encourage them to absorb higher costs rather than improve service.

An example of the user choice approach in Australia is the funding of tertiary advanced and further education colleges (TAFEs). Increasingly state governments are funding public and private-owned TAFEs based on the numbers and types of students they enrol. Victoria has moved to a user choice model with funds going first to users rather than to training institutions. Another example of user choice is government subsidies for child care. These payments may be made either to a child care centre or to families directly. Either way, the families can choose which (accredited) child care centre they send their children to and the centres are remunerated for the volume of services provided.

Although the arguments for user choice models are strong, there appears to be less clear evidence of the effects of user choice on efficiency and other outcomes than there is for competitive tendering and contracting. Rosen and Gayer (2014) report that allowing parents to choose between schools increases competition and improves learning outcomes of students as measured by test scores. However, these outcomes may also reflect parental motivation and support and peer interaction. Lundsgaard (2002) concluded that for services like long-term care for the elderly and disabled, child care and employment services, there are indications that incentives from funding systems matter but that there are too few systematic studies to draw firm conclusions.

Notwithstanding these reservations, mainstream economic thinking generally favours user choice as a starting point unless there are cogent reasons for not providing user choices. And this is certainly the mindset of the Productivity Commission (2017).

Summary

- Production of services involves turning inputs into outputs, which in turn generate outcomes.
- Efficient supply requires (1) that expenditure is allocated to ensure that the marginal value of an output exceeds its marginal cost and (2) that outputs are produced at least cost.
- Under the principles of New Public Management (NPM), government should set clear objectives to meet consumer rather than producer interests and managers should deliver services. NPM also involves competitive supply, empowerment of consumers, decentralisation of service delivery, transparency and accountability.
- Traditionally public expenditure management focused on budgetary processes and provision and control of inputs. Modern management focuses on outputs and performance budgeting that tie agency funding to outputs rather than inputs.
- To estimate costs, it is generally necessary first to estimate a production function that shows the inputs required to produce a given output. This can be done via expert estimates of inputs, formally estimated production functions or activity-based costing methods. Benchmarking compares an agency's costs with the costs of other providers.
- In-house supply of services usually has lower transaction costs than outsourced supply. On the other hand, contracting for services tends to improve service specification; focuses attention on core objectives and user needs; increases incentives to operate efficiently; reveals best practices so that slack cannot be hidden; and makes budget constraints harder. International and Australian evidence indicates that contracting for services often produces cost savings of 20 per cent or more.
- User choice models provide similar benefits by empowering consumers and encouraging competition among suppliers.
- Competitive tendering for delivery of services is desirable when work is specialised and contracts for services can be clearly specified, monitored and enforced. However, government should maintain core management skills, corporate memory and the capacity to manage contracts efficiently.

Questions

1. What are the main differences between outputs and outcomes? Give some examples.
2. What are the various ways of estimating production functions for public services? What are their strengths and weaknesses?
3. What is the relationship between cost and production functions and how are cost functions estimated?
4. What are the main difficulties that arise in estimating average and marginal costs for a public agency?
5. Public and private sector costs are often compared using either 'before and after' or cross-section analysis. How are these methods applied? What issues might arise in assessing the relative costs of private and public production from these methods?
6. Suppose that you were asked to determine whether public or private buses operated at lower unit cost. What kind of data would you need? What kind of statistical model would you use?
7. In Australia, job search assistance has been contracted out to private organisations. What are likely to be the costs and benefits of doing this?
8. If government wishes to increase the use of child care services, what is the best way to do so?
9. How can users of publicly financed services be given increased choices? What are the advantages and disadvantages of giving users more choice over consumption of public services? Give an example.
10. How would you measure the inputs, outputs and outcomes of police services? Is it possible to estimate the average and marginal costs of police outputs? Does this provide benchmarks for comparing the efficiency of police services?

Further Reading

Friedman, L.S. (2002) *The Microeconomics of Public Policy Analysis,* Chapter 9, Princeton University Press, Princeton, New Jersey.

Hawke, L. (2007) 'Performance budgeting in Australia', *OECD Journal on Budgeting*, 7, 1–15.

Khursheed, A.F. and Borcherding, T.E. (1998) 'Organising government supply', Chapter 2 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Lundsgaard, J. (2002) *Competition and Efficiency in Public-Funded Services*, Working Paper No. 17, Economics Department, OECD, Paris.

Productivity Commission (2017) Shifting the Dial, Five Year Productivity Review, Report No, 84, Canberra.

Pyndick, R.S. and Rubinfield, D.L. (2005) *Microeconomics*, 6th ed., Chapters 6 and 7, Pearson Education, New York.

Rainey, H.G. and Chun, Y.H. (2005) 'Public and private management compared', pp. 72–102 in E. Ferlie, L.E. Lynn Jr. and C. Pollitt (eds*), Oxford Handbook of Public Management*, Oxford University Press, Oxford.

Schick, A. (1998) *A Contemporary Approach to Public Expenditure Management*, World Bank Institute, Washington DC.

Vining, A.I. and Weimer, D.L. (2005) 'Economic perspectives on public organisations', pp. 234–256 in E. Ferlie, L.E. Lynn Jr. and C. Pollitt (eds), *Oxford Handbook of Public Management*, Oxford University Press, Oxford.

Chapter

17

Pricing Public Services

They (the Nabobs) raised the price of everything in their neighbourhood, from fresh eggs to rotten boroughs.

Thomas Macaulay, *Critical and Historical Essays*

Pricing Objectives ♦ Pricing Options ♦ Efficient Pricing ♦ Full Cost Recovery ♦ Equity Issues ♦ Conclusions

Most governments charge for publicly produced utility services, such as water, electricity and gas, post or telephone services, bus and rail services. Indeed, governments often establish business enterprises to supply and sell these services. Increasingly, governments also charge for other services, such as tertiary education, road use, property information and waste disposal.[1]

There are three main issues in public pricing. First, whether to charge for a service. It is not possible or desirable to charge for non-excludable public goods such as national or domestic security. More relevant to the discussion in this chapter are equity issues. It would be inappropriate to charge for a service which is regarded as provision of social welfare. Second, once charging is determined, efficient charging is usually regarded as marginal cost pricing. However, as we will see, determining marginal cost is not always easy and marginal cost pricing is not always efficient. This leads to the third issue: full cost recovery. Again, there are issues in determining what this means and when it should be applied.

In this chapter we start by outlining objectives for pricing public services and various pricing options. We then discuss efficient pricing policies when government is the sole supplier of a service and when there are competing suppliers. The fourth main section discusses full cost recovery including setting financial targets and how to fund trading deficits efficiently. The last major section addresses equity issues.

Pricing Objectives

In an ideal world, prices would be set to maximise some measure of social welfare that allows for both efficiency and equity objectives. Indeed, in the discussion of optimal taxes in Chapter 28, we will see that tax structures can be designed to maximise welfare according to some social welfare function.

[1] Economists distinguish between user charges (prices) and fees. A user charge is a payment for a service. A fee is typically a payment for a right. Examples are payments for motor vehicle registration, to sell alcohol or to emit pollutants. This chapter discusses user charges.

However, this would be impractical for setting prices for large numbers of goods and services. Therefore, in setting prices, economists usually seek first to determine whether pricing for the relevant goods meets equity criteria. If this is established, we then seek to set out efficient pricing principles. However, as we discuss below, there may remain equity considerations in setting prices.

Efficiency objectives. Efficiency objectives include producing the desired quantities of each good at least cost and providing goods to those who value them most highly.[2] Prices (user charges) play a critical role in achieving these efficiencies. They show (1) the values that consumers attach to goods and (2) the relative costs of production of alternative suppliers. Therefore, prices guide agencies to produce the goods that consumers want and to supply the goods to consumers who value them most highly. They provide benchmarks for productive efficiency and so encourage efficient production. Also, they promote efficient investment.

User charges (along with financial targets) have other related benefits. First, they facilitate efficient management. A financial target (generally a rate of return on capital) establishes performance goals and benchmarks for performance assessment; provides a basis for making investment decisions; forces agencies to consider the costs of capital tied up in their assets; and enables government managers to monitor performance. Clear financial objectives usually improve agency efficiency.

Second, user charges help to fund services. This reduces the tax revenue required to fund services and therefore reduces the distortions and deadweight losses associated with taxation. On the other hand, prices above marginal cost may also create a deadweight loss (DWL) by discouraging consumption when users are willing to pay for the marginal cost of a service. In designing an efficient price/revenue structure, the benefits of lower taxation need to be weighed against the DWL of user charges above marginal cost.

Also, user charges involve transaction costs. In any application of pricing, the benefits need to outweigh the transaction costs. To minimise transaction costs, user charges should be economic to administer, easy to understand and reasonably constant over time.

Equity objectives. Charging for services meets the equity benefit principle. This principle asserts that the cost of services should be born by beneficiaries of the services, not by others. This implies (1) that individuals should pay for the services they receive and (2) that consumers collectively should pay for the total cost of the service.

However, following an alternative principle of fairness, individuals should be charged for services according to their ability to pay for them. It follows that higher-income households would be expected to pay higher prices. This does not occur when prices are the same for everyone. A given price generally reduces the welfare of a low-income person more than the welfare of a high-income person.[3] Evidently, the ability-to-pay principle of fairness may conflict with the beneficiary principle.

Conclusions. The multiplicity of objectives creates conflicts that require resolution. Prices that satisfy one objective may not satisfy another. Efficient use of resources may require marginal cost pricing; efficient management may require full-cost recovery pricing; equity may require price discrimination between consumers. Economists should be able to show the benefits and costs of alternative pricing policies. Policy making is difficult when government objectives are unclear and/or inconsistent.

---

[2] This is equivalent to the three major efficiency criteria in Chapter 3: product mix efficiency, technical efficiency and exchange efficiency respectively.

[3] This presumes diminishing marginal utility of income and interpersonal comparability of utility.

Pricing Options

In this section we identify five main pricing options. Four of these are based on the cost of producing services, including in some cases the costs of capital investment as well as operating costs. The fifth option, setting market prices, may involve setting prices above production cost for various reasons.

Short-run marginal cost
The cost of supplying an extra unit of a good or service in the short run

Short-run marginal cost (SRMC) pricing is based on the cost of supplying an extra unit of a good or service in the short run.[4] The short run is defined as the period over which capital assets are fixed. In principle, the cost of supplying one extra unit includes any depreciation of capital with use, but this may be negligible. Fixed costs, such as interest payments, overhead administration and operating expenses that are invariant with output in the short run are not included in SRMC pricing.

Average variable cost
The average cost of supplying goods over some specified period, excluding fixed costs

Average variable cost (AVC) pricing is based on the average cost of supplying extra goods over a given period such as one, three or twelve months. AVC pricing excludes all costs that are fixed over the relevant period. However, the longer the period, the more costs may vary with output. Therefore, AVC pricing usually includes more inputs than SRMC pricing.

Long-run marginal cost
The cost of supplying additional units of output in the long run

Long-run marginal cost (LRMC) pricing is based on the cost of supplying additional units of output in the long run, typically over a few years. It includes the cost of replacing fixed assets, as well as the cost of labour and other inputs required for supply of the good. Like SRMC, LRMC is forward looking. Sunk costs are ignored. However, future capital costs must be allocated over time periods and users.

Long-run average cost
The total cost of output divided by number of units supplied

Full cost recovery. This is, in effect, **long-run average cost** (LRAC) pricing. LRAC equals the total cost of supplying a good divided by the number of units supplied. LRAC usually includes any sunk costs that are relevant to the supply. It may also include an allocation of corporate overheads that are fixed and independent of long-run marginal cost. LRAC pricing is sometimes described "break-even" pricing.

Market prices are the prices that consumers are willing to pay for goods. When a good is in fixed supply, notably in the short run, the market price is the market-clearing price. This is the price that equates market demand with supply over a defined period. The price reflects the opportunity cost of the good or service to another consumer. In some cases, market prices may equate to cost-based prices. However, this is not necessarily when capacity is constrained and there is excess demand at SRMC or even at LRMC prices.

Conclusions. These definitions of costs provide useful guides to pricing strategies. However, the definitions are based on somewhat arbitrary distinctions between short- and long-run periods. Estimates of both SRMC and LRMC depend on judgements about which costs are fixed and would be incurred in any case, and which costs are variable. These judgements are often judgements about whether government is committed to provide a certain service rather than simply technical judgements.

It should also be noted that these pricing concepts are usually based on the assumption of efficient costs. Where production is inefficient, government could choose to offer a price discount on costs incurred. However, few governments would admit to this.

[4] Public officials often use the term 'avoidable cost' instead of 'marginal cost'. The terms are interchangeable. Avoidable cost is the cost avoided by not supplying an extra unit of a good.

Finally, government agencies may select price regimes that do not correspond to any of these categories. Indeed, they may choose prices that reflect what the public may bear politically that have no obvious economic explanation in relation to costs.

Efficient Pricing

Government usually supplies services either as sole supplier or with only a few competitors. It rarely supplies services in competitive markets. In competitive markets, private firms generally supply the market efficiently and government provision of services would achieve little, if any, added value. Accordingly, we consider here the principles of efficient pricing when (1) government is sole supplier of a service and (2) there are few competitors.

Efficient pricing as sole supplier

To start with, we assume no short-run supply constraint and that government can supply services at constant marginal cost.[5] This is plausible when a utility, such as water supply, has excess capacity or for labour-based human services. Therefore, in Figure 17.1a the horizontal line (P_1D) represents an SRMC schedule with perfectly elastic supply. Given the demand for services shown, Q_1 is the efficient level of output, where marginal benefit equals SRMC, and P_1 is the efficient price. At a higher price such as P_2, the marginal benefit exceeds marginal cost and there is a deadweight loss equal to area ABC. At a lower price such as P_3, marginal cost exceeds marginal benefit and there is a deadweight loss of area CDE. When price equals SMRC, welfare (defined as the sum of consumer and producer surpluses) is maximised and there is no deadweight loss.

Now consider the implications of a supply constraint, such as plant capacity, on output. Panel (b) illustrates the issues. The maximum supply of water per period is Q_2. Up to this point, water can be supplied for a constant SRMC per kilolitre (kL) of P_1. Once the supply constraint is reached, the SRMC of water is based on the concept of opportunity cost. The marginal opportunity cost of supplying water to one consumer is the marginal benefit forgone

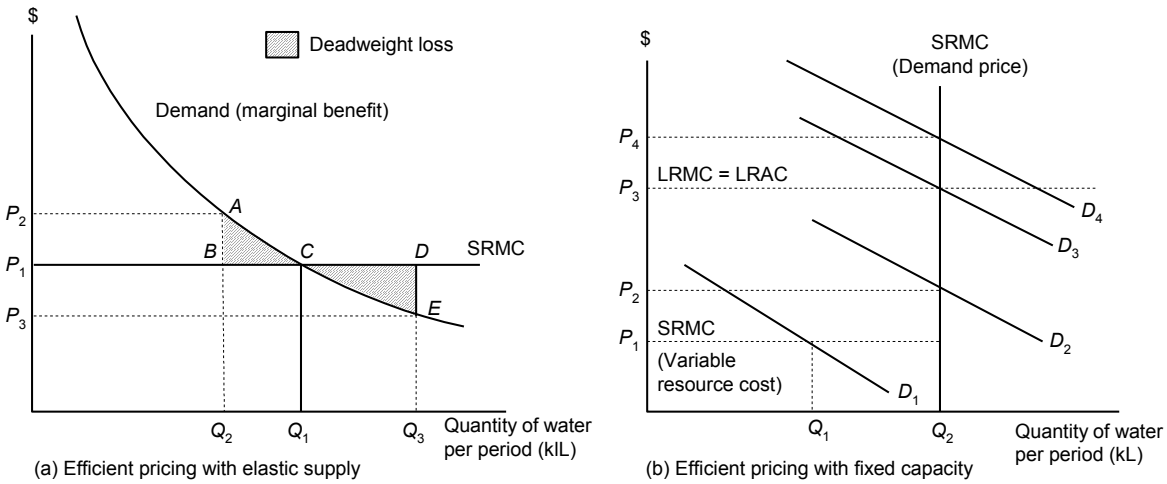


Figure 17.1 Efficient pricing when government is sole supplier

[5] To simplify the discussion in this chapter externalities are ignored. As discussed in Chapter 13, pricing should allow for externalities.

by the marginal user, which is the highest amount that he or she is willing to pay for a marginal kL of water forgone. This depends on the level of demand. Figure 17.1b shows four demand curves, D_1 to D_4. Each point on a demand curve represents the marginal benefit of an additional kL of water. The marginal benefit forgone is shown by the intersection of the relevant demand curve and the vertical supply curve at capacity. Thus, with D_2 demand, SRMC equals P_2. Given D_3, SRMC equals P_3 and so on. If the marginal resource cost of supplying water is \$1 per kL but someone is willing to pay \$2 per kL, the SRMC of supplying a kL to another consumer is \$2.

Social welfare is maximised when price is set equal to SRMC as just defined. Once capacity is reached, this is the market clearing price, which is the efficient strategy with excess demand. This price ensures that water is available to the users with the highest demand for it. Other methods of rationing scarce water would allocate water to some users with lower marginal benefit. Suppose, for example, that demand is given by the D_2 curve. If price is set equal to P_3, capacity would be under-utilised and there would be a large loss of consumer surplus. If price is set equal to P_1, water will be supplied to many consumers with relatively low demand for it. This happens in many countries where an excessive amount of water is supplied to agriculture and too little to industry and urban households (see Box 17.1).

In practice, demand may vary between or within periods. Demand for water is usually higher in summer than in winter. On the other hand, demand for electricity varies over the day, with peak-hour demand much higher than off-peak demand. Differential pricing according to season or time of day encourages efficient use of resources. Suppose that, in Figure 17.1b, demand curve D_3 is summer demand for water and D_1 is winter demand. The efficient charges would be P_3 and P_1 per kL of water in summer and winter respectively. This ensures the water is allocated efficiently in both summer and winter.

Efficient pricing is also important for investment decisions. High prices combined with high consumption indicate a need for more capacity. On the other hand, low prices combined with high consumption may suggest a need for more capacity when there would be little demand for it at prices that would justify the investment. Importantly, the design capacity of many systems is determined mainly by peak demand. If there is excess demand for only a few weeks of the year, it may not be economic to augment capacity. Seasonal pricing may significantly reduce excess demand for water in summer and thus defer the need for expenditure on new

Box 17.1 Water pricing in Australia

In Australian cities water is typically sold at a little over \$2 per kilolitre (kL). It has risen from less than \$1 per kL over a few years. However, the price of water is often much lower in regional and rural areas.

These prices have significant implications. Despite the price increases, at current prices under normal rainfall conditions there is significant excess demand for water in many cities and towns. Also, the marginal value of water use is often much lower in rural areas than in urban areas and there is an inefficient misallocation of water.

Low prices for water (relative to market clearing prices) tend to result in overuse of water and under-investment in replacement of leaking pipes, recycling and new capacity (dams, reservoirs and pumping systems). Consequently, water use is often restricted, rationed and misallocated.

Urban businesses and households who would be willing to pay more for water are unable to obtain it, while rural users have less incentive to economise on their use of water. In the cities, households are forced to adopt inefficient plant and labour-intensive methods to conserve water.

On the other hand, rather than raise water prices further, governments across Australia have invested heavily in expensive, energy-intensive, desalination plants. According to Hoang *et al*. (2009) the production costs of the desalination plants usually work out at between \$1 and \$2 per kL. However, this excludes the resource cost of water. To determine whether desalinisation is an efficient policy we would have to compare the costs of producing potable water via desalination plants with other methods of supplying water, including by recycling.

infrastructure. The discussion on pricing and investment in transport infrastructure in Chapter 19 illustrates these points.

Finally, it should be noted that in finding that SRMC pricing is efficient we made three important assumptions. First, that SRMC a precise concept. This may not be the case. Consider, for example, supply of an extra bus service. The marginal cost of this service depends on the services that would be provided, and the plant and equipment available without this extra service. The marginal cost of an additional service is much lower if a spare bus is available than if an extra bus is required. Also, a judgement is required on the services that would otherwise be provided. Costs, including labour costs, which are variable in some circumstances, are fixed in others. The shorter the decision period, and the higher the proportion of costs regarded as fixed, the lower is the marginal cost of production. This complicates the case for SRMC pricing. If SRMC depends on the circumstances, it is a less exact, and more discretionary, concept than economists commonly suggest.

Second, SRMC pricing is efficient only if there are no substitute products (the cross-price elasticity of demand is zero) or if the markets for substitute and complementary products are perfectly competitive (and adopting marginal cost pricing). Third, we have assumed that, if there is a trading deficit, there is no efficiency cost associated with funding it. We address these issues below.

Efficient pricing with competitors

Many public services are supplied in markets with a small number of other suppliers. For example, a public bus company may compete with some private bus operators. A publicly owned electricity retailer may compete with private electricity or gas suppliers. Here, competitive neutrality is required to achieve an efficient outcome.

When private firms have high fixed costs and low marginal costs, LRAC may exceed SRMC and the firms must set prices above SRMC to stay in business. When a private firm supplies a differentiated product, the demand curve for its product is downward sloping and the firm is again likely to set price above SRMC. In these cases, if a competing public company sets price equal to SRMC, consumption choices are distorted and there is generally excessive public output. This is an example of a second-best environment.

Suppose that a public company and a private firm produce competing, similar quality bus services and that the long- and short-run costs per passenger trip are as shown in Table 17.1. The private firm must charge at least $6 per passenger trip to pay for its operating costs and achieve a normal return on capital. Now, if the public company charges the SRMC of $4 per passenger trip, most bus passengers will use the public service. There is a resource cost of $1 per passenger trip in the short run and $2 per trip in the long run.

Inefficiencies may also occur if the two operators produce different quality services. Suppose that the public company can match the private firm's marginal and average costs per passenger trip of $3 and $6 respectively by using older buses and that passengers are willing to pay $6 for a comfortable private bus trip but only $5 for a less comfortable public bus trip. Now, if the public company charges only $3 for a bus trip, everyone will travel in the less comfortable public buses because this gives them a consumer surplus of $2 per trip. Given that service costs are equal, the deadweight welfare loss is $1 per passenger trip—the difference between what users are willing to pay for private and public bus trips.

Table 17.1 Public and private bus trip costs (example)

Cost per passenger trip	*Public supply*	*Private supply*
LRAC ($)	8	6
SRMC ($)	4	3

In these cases, public sector SRMC pricing causes a loss of welfare. This conclusion is consistent with the general finding of second-best theory that, when there are one or more departures from perfect competition, welfare is not necessarily maximised by adopting perfectly competitive conditions in all other markets (Lipsey and Lancaster, 1957). On the contrary, welfare may be increased by departures from these conditions in other markets.

What pricing policy should a public agency adopt in such cases? This depends on the degree of substitutability between the public and private good. If they are fully substitutable, the absolute mark-up on the public good should be the same as the mark-up on the private good. Let us call the public agency (X) and the private firm (Y). Further, we assume that Y sets its price above marginal cost and that its output cannot be controlled directly. Agency X wishes to choose the level of output that maximises welfare defined as the sum of consumer and producer surpluses.

Now, if X raises its price above marginal cost and demand falls, there is a loss of consumer surplus. However, if consumers switch to Y, the private firm will gain producer surplus equal to $(P_Y - SRMC_Y)$. When there is one-for-one substitution, if the price mark-up over marginal cost is larger in Y than in X, there is a net gain from the switch to Y.

As shown by Rees (1984), the efficient mark-up for P_X is:

$$P_X - MC_X = -\beta\,(P_Y - MC_Y) \tag{17.1}$$

where $\beta = \Delta Q_Y / \Delta Q_X$. The parameter β is the ratio of the change in consumption of Y to the change in consumption of X following an increase in P_X over MC_X.[6]

When the two goods are substitutes, $\Delta Q_Y / \Delta Q_X$ is negative (because ΔQ_X is negative) and $-\beta$ is positive. Equivalently the efficient price for the public service (X) is:

$$P_X = MC_X - \frac{\Delta Q_Y}{\Delta Q_X}(P_Y - MC_Y) \tag{17.2}$$

If the services provided by X and Y are perfect substitutes, $\beta = -1$ and P_X should be set equal to MC_X plus $(P_Y - MC_Y)$. This price adjustment preserves the equilibrium quantities that would prevail if both services were charged at marginal cost. If there is no substitution between services from X and Y, $\beta = 0$ and P_X should equal MC_X.

Suppose that $\beta = 0.5$. Then P_X would equal $MC_X + 0.5\,(P_Y - MC_Y)$. Adopting the numbers in Table 17.1, P_X would rise from \$3.0 to \$4.50. Suppose that this mark-up caused 1000 passengers to stop using public buses and that 500 transferred to private buses. Firm Y would gain profit of \$1500 = (500 × \$3). On the other hand, 1000 passengers would lose an average consumer surplus of \$1.50 × 0.5 = \$0.75 (assuming a linear demand curve) and experience total loss of \$750. There would be a net social gain.

If agency X has two competitors (Y and Z) the efficient price for X becomes:

$$P_X = MC_X - [\beta_Y(P_Y - MC_Y) + \beta_Z\,(P_Z - MC_Z)] \tag{17.3}$$

where β_Y and β_Z are similar to β in Equation 17.1 but here relate to firms Y and Z respectively. If agency X has several competitors, the efficient mark-up of P_X over MC_X would be the weighted average of the mark-up of all other substitute goods where the weights would reflect the extent to which the other goods are substitutes for X (see Box 19.1 on p. 325).

Equations 17.2 and 17.3 show the nature of second-best pricing policy and are straightforward in principle. But the practical requirements are demanding. To apply these equations, public agencies need estimates of the relevant compensated demand functions and the marginal costs of supply for both their own service and their competitors' services.

---

[6] This solution is based formally on compensated demand curves as defined in Chapter 6.

In the absence of such information, when there is competitive supply, resources may be allocated most efficiently when public agencies set prices for their services based on estimated LRMC. LRMC pricing encourages competition and reduces relative price distortions compared with SRMC pricing. It also provides financial discipline for the agency.

In some cases, a public agency is a sole supplier but there is potential competition. In such situations there is again a case for LRMC pricing instead of SRMC pricing. LRMC pricing ensures fair competition where competition exists and provides the appropriate incentive for competition where it does not currently exist but might do so.[7]

Efficient pricing with externalities

Where there are externalities, efficient pricing requires that these should also be incorporated into the pricing formula. Thus, in the case of negative externalities (with say road congestion) the SRMC or LRMC should represent full marginal social costs and include the cost of the negative externality. This is discussed in some detail in our discussion of pricing for transport in Chapter 19.

Full Cost Recovery

So far, we have set prices to achieve efficient use of resources on the assumption that if trading deficits occur, they would not matter.[8] Of course, deficits may matter politically if they are perceived to reflect inefficient management and mean higher taxation.

The main economic reason for full cost recovery is that a deficit is generally funded from consolidated revenue by increased taxes, which usually involve a deadweight loss. For example, taxation of labour income affects work hours. As we will see in Chapter 27, the estimated marginal deadweight loss associated with taxation is usually at least 20 per cent of revenue raised. Therefore, we need to consider whether pricing strategies to reduce trading deficits have a lower deadweight loss.

There are other reasons for full cost recovery. As we noted above, a financial target encourages an agency to manage its assets efficiently. Funding from consolidated revenue reduces the incentive to manage assets efficiently. Also, such funding of deficits is often not a transparent process. In addition, when a public agency is competing with one or more private firms, resources are generally allocated more efficiently with competitive neutrality. This usually implies full-cost recovery. Indeed full-cost recovery is a common objective of Australian state jurisdictions (see Victorian Department of Treasury and Finance, 2013).

However, a caveat is in order. When there is little competition, full cost recovery may not indicate efficiency. An agency may achieve a target rate of return by raising prices rather than by improving productivity. In this situation, measures such as total factor productivity or cost per unit of service may be a better indication of efficiency than the rate of return on assets (Waters and Street, 1998).

Defining financial targets

The concept of full cost recovery implies a financial target. As we saw in Chapter 14, this target is a function of the value of capital employed and a target rate of return on this capital.

---

[7] The case for LRMC pricing is looking forward and (rationally) ignores sunk costs. Because private firms cannot ignore financial liabilities, there is arguably a case for full-recovery cost (LRAC) pricing. This is the recommendation of the Victorian Department of Treasury and Finance (2013).

[8] Under some circumstances SRMC prices broadly defined do not result in a financial deficit. As shown in Figure 17.1b, demand can exceed capacity output at LRMC prices. If it does, SRMC pricing is equivalent to market pricing and potentially produces super profits.

The total revenue target is then set to cover operating expenses and achieve an acceptable return on capital.

A break-even target. A break-even target is a common accounting concept. This is the revenue required to meet all financial obligations, including debt payments and interest. Equation 17.4 represents a break-even requirement that revenue (R) equal or exceed costs:

$$R \geq VE + IE + DR \tag{17.4}$$

where VE is variable expenses (not including depreciation), IE is interest expenses and DR is debt repayments during the relevant period. However, this does not account for equity capital or any return on equity capital.

An accounting rate of return. Private firms also often adopt a revenue target based on a rate of return on the book value of capital employed. This target can be represented by:

$$R \geq VE + IE + BD + (K_B \times r) \tag{17.5}$$

where BD is book depreciation, KB is the depreciated value of capital owned by the business at the start of the year based on historic book values and r is a required rate of return on K_B.

This approach is based on the standard form of account keeping and is simple to administer. Also, it provides for the main financial needs of a business agency, that is, the revenue to meet obligations to creditors and a prudent level of reserves.

However, in this formulation, asset values and depreciation are based on historic costs. These costs do not reflect the true cost of using capital tied up in the assets. Nor does income include real changes in asset values. Thus, the revenue requirement may be understated.

An economic rate of return. A revenue target based on the economic rate of return (ERR) is designed to overcome these problems. The ERR is the return on the market value of assets owned by the business at the start of the period:

Economic rate of return
The return on the market value of assets owned by a business at the start of the period

$$R \geq VE + IE + \Delta K_M + (K_M \times r) \tag{17.6}$$

where K_M is the market value of assets owned at the start of the period and ΔK_M is the change in market value of assets over the period under consideration. This revenue target is based on current asset values and measures expenses correctly.

However, as we saw in Chapter 14, there are significant problems in establishing appropriate values for K_M and r. Market values for public assets are often unavailable or unreliable. In many cases, the assets could not be sold and employed elsewhere. And their market value in their current location depends on their market power and countervailing government regulation. Thus, there is often no ready measure of their market value.

Alternatively, assets can be valued at replacement cost. This is the cost of purchasing or reproducing an asset that can provide similar services using current technology. This cost is written down to reflect any economic depreciation in the value of the asset. Replacement cost is backward looking in that the revenue target is based on paying for existing assets rather than new ones. However, it generally provides a useful basis (in conjunction with a required rate of return) for estimating the revenue required from a set of assets.

The appropriate value of r is the opportunity cost of capital. Where this is equity capital, the main issue is allowance for systemic (of market) risk. The general principle is that the rate of return should reflect the return that can be achieved by private investment allowing for systemic risk that cannot be diversified away. However, an allowance may be made for public trading enterprises that have a below average systemic risk.[9]

[9] The points about the required rate of return in this paragraph pick up on points made in Chapters 8 and 14.

Efficient prices to fund deficits

The two main pricing strategies for dealing with trading deficits are mark-up pricing and non-linear pricing.

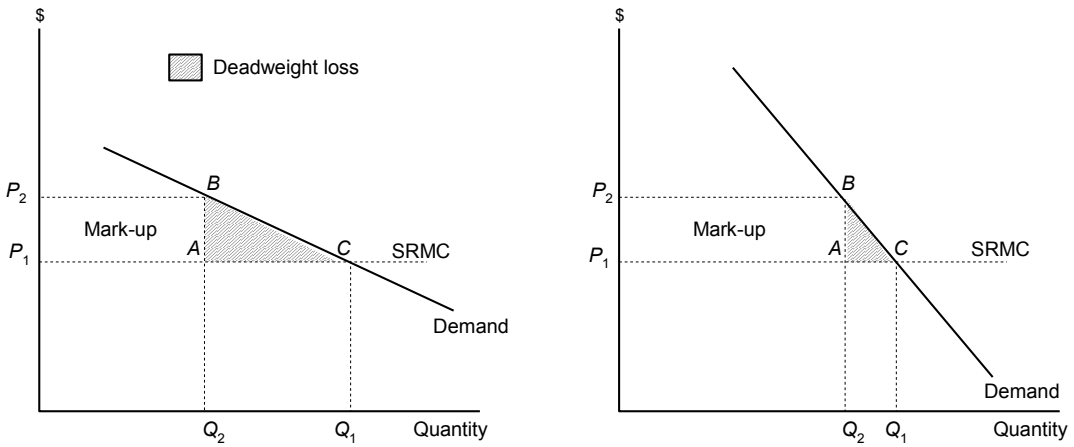
Mark-up pricing. This involves increasing prices above marginal cost until the revenue target is reached. However, when a sole-supplier agency sets prices above SRMC, there is a DWL of consumer surplus. As we saw in Chapter 14, DWL rises with the size of the price mark-up and the elasticity of demand. Figure 17.2 shows DWL with elastic demand in panel (a) and less elastic demand in panel (b). DWL rises with more elastic demand.

The aim therefore is to find the mark-up prices that minimise the DWL due to lower consumption. This is achieved by raising prices for goods in inelastic demand or by charging higher prices to consumers with an inelastic demand for the goods. If demand for a good is perfectly price inelastic, even a very large mark-up does not affect consumption. There is a transfer payment from consumers to the producer, but no reallocation of resources.

Ramsey (1927) showed that to meet a revenue target with least DWL the percentage reductions in the quantities of each good contributing to this revenue must be equal. Strictly these are reductions in quantities along compensated demand curves. To achieve this, providing that there are no cross-price effects (i.e. the price of one good does not affect the demand for another), the percentage price mark-ups should be inversely proportional to the compensated price elasticities of demand for each good.[10] This can be expressed as:

$$\frac{(P_i - MC_i)}{MC_i} = \frac{k}{\eta_i} \tag{17.7}$$

where k is a constant, η is the compensated price elasticity of demand and i is the ith good. The constant k is set so that the mark-ups yield the revenue to meet the revenue target.[11] This form of mark-up pricing is known as Ramsey pricing.[12]



(a) Deadweight loss with elastic demand

(b) Deadweight loss with inelastic demand

Figure 17.2 Deadweight loss and demand elasticity

[10] A more detailed proof of these propositions is given in our discussion of optimal tax in Chapter 27.

[11] Allowing for cross-price effects, Weare and Friedman (1998) show that if two goods are strong substitutes (complements), the optimal Ramsey mark-up is increased (reduced).

[12] Ramsey (1927) is usually credited with the principle of least-cost mark-up pricing. The same principle applies to efficient taxation (Chapter 27).

If an agency supplies one or more goods with inelastic demand, an efficient mark-up strategy can raise revenue with low deadweight cost. However, if mark-ups are not based on the Ramsey principle, mark-up pricing can result in distortions of consumption and deadweight loss.

Non-linear pricing. With a non-linear price structure, prices per unit of consumption vary with the level of consumption. A volume discount is an example of this.

A two-part tariff, sometimes described as access pricing, is a common non-linear tariff structure. With a two-part tariff, users are charged for access to a system (such as a telephone network or water utility) and for use of it. Access charges can pay for the fixed costs, user charges for variable costs. Figure 17.3 shows how this can work. The efficient level of output is Q_1, which reflects the intersection of the market demand curve and the SRMC line (i.e. the equality of marginal benefit and cost). If the agency charges P_2 to cover LRAC, consumption would fall to Q_2. With a two-part tariff, consumers pay P_1 per unit of consumption plus a fixed charge that equals the fixed cost (area $ABCP_1$) divided by the number of consumers. This system is generally efficient. Because the demand for access to utilities is usually highly inelastic, an access charge has little effect on consumption. On the other hand, if user charges reflect SRMC, the key relationship between user charge and marginal cost is maintained.

However, a two-part tariff can present problems. First, access charges can create a DWL when consumers are deterred from access and drop out of the market altogether. This effect may be minimised by offering customers options with various fixed and user cost components so that they can choose their preferred combination. Second, access charges have limited application when services have no obvious access component, for example data on property titles. The marginal cost of supplying these data is low. However, because all property owners benefit from the security of title provided by the work of the agency, the agency's fixed cost could be recovered by a fee when properties are transferred or an annual fee on all property owners. This would have a minimal adverse resource allocation impact. A third potential weakness of access charges is that they are generally flat charges across the population and independent of use. Therefore, they may be more regressive than charges that vary with use.

In conclusion. Ramsey mark-up pricing and non-linear tariff structures are the main ways to raise revenues to meet a budget target with least deadweight loss. When prices are marked up on goods in inelastic demand or when tariffs are levied on access to a service for which demand is inelastic, mark-up prices and access charges have low DWL. In other cases, mark-up pricing and non-linear tariffs can have high DWL, especially if applied inefficiently.

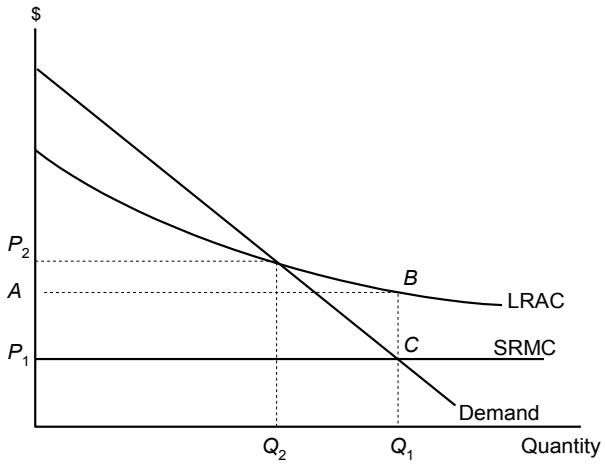


Figure 17.3 Two-part tariff or access pricing

Equity Issues

As we have seen, equity in pricing can be viewed in terms of two, not always consistent, principles: the benefit principle and the ability-to-pay principle. Pricing services at full cost is consistent with the benefit principle. But the ability-to-pay principle suggests that government should subsidise essential services, such as power or transport, for less well-off households.

In practice, governments often use public trading enterprises to achieve social goals, such as cheap urban transport and subsidised communications for regional and rural households. The subsidies may be financed from consolidated revenue, as typically occurs with rail subsidies in Australia. Other subsidies may be financed by cross-price subsidies, with some users paying prices above long-run average cost for their services. This often occurs with communications and electricity in Australia.

Two main questions arise. First, are subsidised services the best instrument for achieving equity goals? Second, if a subsidy is appropriate or politically unavoidable, what is the best way to provide it?

As discussed in Chapter 22, untied income support generally raises the welfare of households more than the equivalent income tied to specified goods. Most people prefer $1000 in cash or untied tax relief to $1000 in subsidies for public transport or other services that they might not purchase with untied income. Second, fiscal transfers target less well-off households better than do subsidised goods. For example, subsidising goods sold to regional households by taxing metropolitan consumers is a crude form of redistribution as there are affluent households in regional areas and poor ones in urban areas. Third, fiscal transfers usually have lower costs of administration than service subsidies. Economists usually conclude that equity goals are achieved best by fiscal measures and that departures from efficient pricing policies are an inferior instrument for achieving equity goals.

Nevertheless, governments often subsidise services because they believe that fiscal transfers are not sufficient. A public choice explanation would be that administration of subsidised services enhances the role of public servants more than do simple income transfers. Accordingly, we need to consider how best to provide these subsidies.

Service subsidies can be provided in four main ways:

1. Undercharging all users of selected services and financing the deficit from consolidated revenue.
2. Undercharging targeted services or users, such as pensioners, and financing the deficit from consolidated revenue.
3. Undercharging for some services and overcharging for others (i.e. by cross-price subsidies).
4. Providing subsidies, in vouchers or cash, to targeted users.

Governments generally employ a mixture of the first three strategies, although the fourth strategy is gradually being used more.

Strategy 1 (undercharging all users) is a blunt instrument. Subsidies would be provided to rich and poor alike. The Industry Commission (1994) showed that a high proportion of the annual billion dollar subsidy for public transport in Australian cities accrues to medium- and high-income households. Moreover, there is no benchmark for how much subsidy to provide and no financial or management discipline for the trading agency. The deficit is often treated as an open-ended bucket of public money.

Strategy 2 (undercharging selected users) is a more direct and generally lower cost subsidy approach. It is often implemented with a **community service obligation** (CSO) payment. A CSO requires that government and the trading agency agree on (1) the subsidies to be provided, (2) the cost of the subsidies to the trading agency and (3) the method of

reimbursement by government. This approach has several advantages over strategy 1. First, government, not the trading agency, decides the intent and level of subsidy. Second, a CSO is a transparent payment. Third, the trading agency is subject to commercial accountability—the CSO is incorporated into the revenue stream of the trading agency. However, the strategy requires that the subsidies and their costs be well defined. This means determining the extent to which service costs exceed charges. This depends in turn on whether the benchmark cost is SRMC, LRMC or LRAC. Government and a trading agency may disagree on the size of a CSO. Sometimes, strategy 2 is hard to distinguish from strategy 1.

Cross-price subsidies (strategy 3) are rarely equitable or efficient. Their real incidence is often obscure. For example, transport or telecommunication subsidies to households who live in designated areas generally benefit property owners (who can charge higher rents) rather than renters. On the other hand, the real burden of high input prices for business often falls on consumers via higher prices, rather than on shareholders. Even if the subsidies are well directed, it is not clear why other service users should bear the cost rather than the whole community via consolidated revenue, as in strategy 2. In terms of efficiency, charging some consumers above LRAC to pay for the subsidies discourages efficient consumption and creates a deadweight loss.

Providing subsidies to users directly rather than to suppliers (strategy 4) can work in two main ways. Government can provide vouchers to targeted individuals that entitle them to a specific service (e.g. electricity) at a reduced price. On receipt of the vouchers from a service provider government reimburses the provider for the agreed amount of the voucher. Alternatively, government can reimburse targeted users on receipt of specified expenses. Whichever means is adopted, this strategy has significant advantages. The consumer can choose how to use the subsidy to maximise his or her service benefits. Also, the strategy maximises the opportunity for competition between service providers. The main disadvantage is the administrative cost of dealing with large numbers of consumers rather than with a small number of service suppliers, sometimes only one.

Conclusions

The principles for efficient pricing of public services can be summarised as follows. When government is sole supplier of a service, price should generally be set equal to SRMC of production. When there is a supply constraint, the SRMC is the highest price that someone would pay for a marginal unit of output as this reflects the opportunity cost of supplying the good to one consumer rather than to another.

When actual or potential private firms can supply substitute services and private sector prices are set to cover LRAC, the principle of competitive neutrality (or second-best arguments) requires that public prices be marked up in similar fashion to the private sector. This may require that a public agency price its services at least at LRMC.

However, SRMC or LRMC prices may not achieve full cost recovery, with consequent demands on consolidated revenue and related DWL. This may lead to a preference for full cost recovery. To achieve cost-recovery efficiently, agencies should mark up prices according to the inverse price elasticity rule or use a non-linear price structure like a two-part tariff.

There are several constraints to implementing these principles. One is the amount of information required on cost functions and demand elasticities. Competitor firms will not willingly provide these data. Calculation of efficient prices often requires information on demand price elasticities for specific services and specific consumer groups, including cross-price elasticities. Consequently, precise applications of Ramsey pricing are rare.

Another issue is the notion of the marginal increment in output. The marginal costs of picking up one extra bus passenger, of making a regular additional bus stop and of running an extra bus service are each different. Similar examples abound in other sectors. Determining

the relevant marginal cost depends on the decisions to be made. In the case of the bus service, the main decision may be whether, or not, to run the service. A secondary decision will be where to pick up passengers. Prices cannot be adjusted to the marginal cost of every decision, such as whether to pick up an extra passenger.

Third, when agencies are required to recover full costs, joint costs must be allocated. For instance, rail or telephone infrastructure may have to be allocated to users of the networks. Regulators often allocate these overhead costs among users according to their proportionate use of the facilities. These proportions may be estimated as the percentages of use (e.g. rail passenger kilometres), of revenues generated or of variable costs attributed to the service. However, these regulatory practices may be inefficient. Often joint costs reflect the need for capacity, which may be administrative or infrastructure capacity. The costs should be allocated to the services that require this capacity. For example, the costs of new rail infrastructure should be allocated to peak-hour services rather than to off-peak services.

Transaction costs are another problem. Transaction costs include metering use, administering a pricing system and enforcing prices. If transaction costs were zero and the demand for services such as road use or electricity were continuously variable, the solution would be real-time (instantaneous) pricing. This is rarely practical. Transaction costs must be factored in to development of an effective pricing strategy.

In addition to efficiency issues, there are frequently concerns about equity especially where the services are perceived as welfare services. Here, decisions are required as to whether services should be subsidised and, if so, what is the best method of subsidy. If a service is to be subsidised, a strong case can be made for setting an efficient price and for providing a transparent subsidy to users rather than reducing prices to all users or adopting cross-price subsidy strategies.

Given these constraints, the pricing of public services rarely resembles the detailed prescriptions for efficient pricing described in this chapter. However, the prices of public services are probably closer to the economist's prescriptions for efficient pricing today than they were 10 or 20 years ago.

Summary

- Prices for public sector services have several functions: to promote efficient use of resources, to facilitate efficient management, to distribute services fairly and to raise revenue with minimum deadweight loss. The multiple objectives create potential conflicts that require resolution.
- When government is sole supplier of a service, efficiency requires that prices be set equal to short-run marginal cost. If there is a supply constraint, the short-run marginal cost is the opportunity cost of supplying a service to one consumer rather than to another.
- When there is competition in supply of a service, setting public prices equal to the long-run marginal cost of supply encourages a level playing field and efficient allocation of resources.
- Also, short-run marginal cost prices may create trading deficits. Deficits are inefficient if they weaken management incentives to be efficient or are financed by taxes that create deadweight losses.
- Cost recovery revenue targets depend on the valuation of capital assets and required rates of return on capital as well as on operating expenses.
- The most efficient ways to pay for fixed costs, or to reach revenue targets, are by mark-up pricing on goods in inelastic demand or by a non-linear pricing system, such as a two-part tariff scheme.
- If government wishes to subsidise a service, it should provide a transparent community service obligation payment either to targeted users or to service provider(s). Open-ended public subsidies to fund deficits and cross-price user subsidies are often inefficient ways to achieve social goals.

Questions

1. The Productivity Commission (2001b) points out that many regulatory agencies, such as the Therapeutic Goods Administration, recover all or a high proportion of their costs, whereas most information agencies, such as the ABS, recover only a small part of their costs.
   i. What principles should determine whether government agencies should recover costs for services provided?
   ii. Should regulatory agencies aim to recover 100 per cent of their costs?
   iii. Should information agencies aim to recover some or all of their costs?
2. What are the main strengths and weaknesses of short-run marginal cost pricing?
3. When does short-run marginal cost pricing lead to full cost recovery?
4. Suppose government is the sole supplier of a good to the public. Demand for the good is given by $Q^d = 50 - P$. Short-run marginal cost is constant at \$10 per unit of output. Suppose the government is producing 30 units of the good. Is this an efficient level of production? If not, what is the efficient level and price to be charged?
5. Suppose government adopts a strategy of mark-up pricing for a good. Demand for the good is given by $P = 50 - 2Q$. The government can produce at a constant marginal cost of \$10 per unit. Government marks up the price by 20 per cent over marginal cost.

   Determine the following;
   i. The price and quantity under the mark-up pricing strategy.
   ii. The efficient price and quantity combination.
   iii. The deadweight loss that emerges as a result of the pricing strategy.
   iv. Suppose that demand for the good becomes more inelastic. What will happen to the size of the deadweight loss as a result of the mark-up?
6. Suppose that a bridge can be built for \$30 million and that there are no operating costs. The expected demand is $Q^d = 800 - 100P$ where Q is the number of crossings in thousands per annum and P is the price per crossing.
   i. If a private company builds the bridge, what will be the profit-maximising price?
   ii. Will that price lead to the efficient number of crossings? Why or why not?
   iii. What will be the company's profit or loss? Will the company build the bridge? Note this may involve a discussion of discount rates.
   iv. If the government were to build the bridge, what price should it charge?
   v. What is the social benefit from the bridge and what is the social rate of return? Should the government build the bridge?
7. What is the main difference between an accounting and an economic rate of return? If a government agency makes an economic rate of return, does that mean that it is making an economic profit?
8. If a government agency is operating at a loss, what are the most efficient ways to reduce the deficit?
9. A government agency produces two goods, *X* and *Y*. Selling both goods at prices equal to their respective short-run marginal cost, the agency sells 20 units of *X* at \$10 and 30 units of *Y* at \$20. The price elasticities are –1.0 and –2.0 for *X* and *Y* respectively. To cover its fixed costs the agency needs to raise \$100 over its short-run costs. At what prices should the agency sell *X* and *Y*? What is the deadweight loss associated with the mark-up in each case? Show why these mark-ups are efficient.
10. If government wishes a publicly owned electricity supplier to provide subsidised electricity to low-income households, what method would you recommend and why?
11. When a public agency supplies water, how should it determine charges? When is it sufficient for charges simply to cover (i) operating costs and (ii) operating and capital costs?

Further Reading

Bos, D. (1987) 'Public sector pricing', Chapter 2 in *Handbook of Public Economics*, A.J. Auerbach and M. Feldstein (eds), Elsevier, Amsterdam.

Hillman, A.L. (2003) *Public Finance and Public Policy*, Chapter 8, Cambridge University Press, Cambridge.

Organisation for Economic Cooperation and Development (2010) *Pricing Water Resources and Water and Sanitation Services*, OECD, Paris.

Productivity Commission (2001b) *Cost Recovery by Government Agencies*, Report No. 15, Ausinfo, Canberra.

Victorian Department of Treasury and Finance, 2013, *Cost Recovery Guidelines*, Melbourne.

Weare, C. and Friedman, L.S. (1998) 'Public sector pricing: an institutional approach', pp. 597–658 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Chapter

18

Public or Private Production and Ownership

When choosing between two evils, I always like to try the one I've never tried before.

Mae West, actress

Ownership: Introduction ♦ Ownership and Efficiency ♦ Distributional Effects ♦ Evaluating Changes in Ownership ♦ Public-Private Partnerships

The economic theory of public goods tells us that government has a major role in the financial provision of these goods, but it does not tell us **how** these goods should be supplied. This applies both to user-pay services such as the major infrastructure networks for energy, water, telecommunications and transport and to major social services like health and education that are generally provided free. All these services are of fundamental importance to the economy and to most households. Many of them, especially the infrastructure services, contain elements of natural monopoly, so duplication would be inefficient, and many are integrated or networked services. These features led many governments historically to believe that public ownership of the network or service was necessary for efficient and socially just production (see Box 18.1 overleaf). Subsequently, poor performance in many cases led many governments first to develop quasi-independent public corporations (public trading enterprises) and then later in some cases to privatise these services (by sale to the private sector). However, the issue of public or private ownership and production remains an important one where the arguments are not always well understood.

We discussed some related issues when we considered the relative merits of in-house or outsourced production in Chapter 16. There we were principally discussing efficient short-term delivery of recurrent services. Here we discuss the long-term and structural foundations for the supply of fundamental economic and social services, including the ownership and control of assets. While some similar issues arise, we will also need to draw here on theories of incentives and behaviour, the nature of the firm and the influence of market structure.

The debate has also become complicated by the development of public–private partnerships (PPPs). Under the traditional model, either government or the private sector was responsible for all elements of the business including financing, owning and operating it. PPPs create the options where government is responsible for some elements of production and the private sector for others. Many permutations of public and private arrangements are possible.[1]

[1] Public–private partnerships are sometimes described as private finance initiatives (PFIs) especially in the United Kingdom. The terms PFI and PPP are essentially interchangeable.

Box 18.1 Nationalisation and privatisation: A brief history

Economic fashions change. Following the Great Depression in the 1930s and the Second World War, governments around the world took over (nationalised) many large private businesses providing power, transport, telecommunications and other basic economic services in the belief that public ownership would be more efficient and equitable.

In the 1980s and 1990s, many governments sold these same businesses, thinking that private owners would manage them more efficiently. Between 1979 and 2000, the share of state-owned enterprises in global GDP fell from 10 per cent to 6 per cent (Megginson and Netter, 2001).

However, debate over ownership of key economic services continues. Government has re-nationalised various rail services and the London Underground in the United Kingdom, the print office in Germany, airport security services in the United States and the railways in New Zealand. And recently such private icons as General Motors, the Bank of America and the Bank of Scotland and numerous other banks have been temporarily at least nationalised.

In Australia, governments traditionally from the colonial days owned most of the infrastructure networks in the energy, water, telecommunications and transport sectors. They also owned several banks and insurance companies.

Since 1990 government revenue from asset sales in Australia has been well over $100 billion. In the first half of the 1990s, the Commonwealth (Labor) government sold Qantas and the Commonwealth Bank. Following this the Liberal-National coalition government sold the national telecommunications company (Telstra) and several airports. In the 1990s, state governments sold most of their publicly-owned banks and insurance companies. The Victorian government also sold its electricity generators and distributors and its transmission grid. In the last 15 years the NSW government has sold numerous assets, including electricity transmission and retail businesses, ports and the property registration service, and the Queensland government has sold its major railway lines.

On the other hand, in 2008 the Commonwealth (Labor) government set up the wholly state-owned National Broadband Network Corporation with a monopoly charter to provide broadband infrastructure to every home in the land at an greatly under-estimated capital cost of $45 billion.

This chapter starts with a discussion of the nature of ownership, reasons for privatisation, methods of privatisation and PPPs. We then discuss efficiency and equity issues associated with public and private ownership and production and how to evaluate changes in ownership. Finally, the chapter discusses some key issues for PPPs.

Ownership: Introduction

Ownership implies rights and risks. Ownership of assets entitles the owners to determine how the assets are used and how the business is managed. Ownership also entitles the owners to the residual earnings from the assets of the business after all prior claims have been met. These claims include payments to workers, suppliers of equipment and materials, debt and interest payments and tax liabilities. On the other hand, the owners bear the prime risk of business failure. Accordingly, a key issue is: 'do the financial incentives associated with ownership result in efficient management of the business and efficient allocation of the risk?'

Turning to business models, two important issues should be noted. The first relates to the role of regulation. Publicly owned firms are in effect highly regulated. But private firms can also be subject to strong regulations as well as to general commercial regulations. When government transfers a business to the private sector, it must decide the kind of regulation that will apply. If it transfers a natural monopoly, such as an electricity transmission grid, there is likely to be substantial ongoing regulation. In this case government must choose between running a publicly owned business and regulating a privately-owned monopoly. In other cases, the choice lies between a publicly owned business and a competitive private business.

Second, most large businesses can be broken into separate vertical and horizontal components. For example, a vertically integrated electricity industry can be broken down into power generation, transmission and distribution businesses and into separate geographical areas. In major construction projects, the works can be broken down into design, construction, finance, management of the asset, provision of operating services and so on. Ideally

government would find the public–private mix for each of these component services that produced the best overall result. The capacity to break businesses or projects into component parts has been an important factor leading to the development of PPPs.

Objectives of privatisation

The fundamental objectives of privatisation are the same as the objectives for any use of resources, namely efficiency and equity. We will see how these objectives can be applied to evaluate privatisation proposals below.

It is sometimes suggested (for example Price Waterhouse, 1989) that there are other economic objectives such as subjecting state-owned enterprises to market discipline or providing the opportunity to introduce competition. These are important considerations, but they are essentially means to produce efficient outcomes rather than objectives in themselves.

Governments often sell assets for financial reasons, especially to avoid or reduce a financial deficit. We also discuss this below and conclude that, in the absence of economic (efficiency) reasons to sell assets, financial reasons rarely justify selling assets.

Finally, governments sometimes have avowedly political motives for selling public assets. The UK (Thatcher) government in the 1980s saw wider ownership of equities (shares) as a vehicle to reduce the power of the labour unions and to promote capitalism and hence voter support for its ideology and the governing (Conservative) party.

Methods of privatisation

The most common way to transfer a business is by sale of shares to the public. Most often, government issues a number of shares at a set price. And generally, in return for an underwriting fee, private financial institutions guarantee that the shares will be taken up at the offer price. If there is excess demand for the shares at the offer price, the shares requested are scaled down, usually according to a predetermined formula, to the number available. Where this occurs, after the shares are listed on the stock market the price rises above the government's offer price. Alternatively, government may auction shares to the highest bidders. This may increase the price received by government, but it increases the risk borne by government as the price may be lower than in an underwritten tender.

Government may also sell a business to a single firm or to a consortium via a trade sale, usually by competitive tender. The Australian government used trade sales, for example, for airports. A trade sale is easier and quicker to organise than a sale to the general public. However, when there are few tenderers it may raise less money. Also, it may result in an increase in private market power. Businesses may also be placed with groups of investors or sold via a management and employee buyout, as occurred with the sale of the National Freight Consortium in the UK.

Government can also transfer assets simply by distributing shares free to all or selected citizens. This method has been used in Eastern Europe (World Bank, 1996). The method is fast and fair. But it raises no revenue for government and, initially at least, it may create a governance vacuum.

Importantly, the rights, obligations and constraints that attach to the transfer of assets or related business are critical to the sale price, the competitive environment and the subsequent operation of the assets. For example, when the UK government floated British Telecom (BT), it legislated that only limited competition for BT would be allowed. On the other hand, in Australia it was always understood that the privatised Telstra would compete with other privately-owned telecom companies.[2]

[2] But few investors anticipated that Telstra would be forcibly divested of its wholesaling business to make way for a new publicly owned telecom company (the NBN Corporation).

Government may also require a private firm to provide specified services that are not commercial. Regulations are critical to performance of the business especially when there is little competition, as, for example, with airports and power transmission facilities.

PPPs: mixed forms of ownership and finance

Table 18.1 shows a spectrum of ownership possibilities from full private to full public ownership and various PPP alternatives. In all cases government has the final regulatory control although this control may be constrained by contractual arrangements.

A common PPP arrangement is a BOOT-type arrangement, under which a private firm builds, owns and operates an asset, but ownership transfers to the state at end of a specified period. Usually the private sector arranges the capital funding. Sometimes, government contributes finance and guarantees some revenue streams, especially for social facilities.

However, there are many variations. Build, own and operate (BOO) arrangements are close to complete private provision and ownership, except that government usually has a major role in the planning stage. On the other hand, with build, transfer and operate (BTO) schemes the private sector may finance, build and operate the facility but ownership is transferred after construction back to government.

Leasing arrangements provide another approach. For example, government may lease a state-owned asset to a private firm for an agreed use for a period of, say, 10 or 20 years. The NSW government has leased its interstate rail lines to the (Commonwealth-owned) Australian Rail Track Corporation for 60 years. Alternatively, government may sell an asset to the private sector and lease it back for an agreed period and price. This is sometimes done with office buildings. The NSW government also did this for a large power station (Eraring).

Ownership and Efficiency

In this section we examine how ownership may affect efficiency. Specifically, we consider whether ownership affects allocative efficiency (the quantity and type of goods produced) or productive efficiency (the unit cost of goods supplied). Key issues are the impacts of ownership on (1) business objectives and (2) individual motivations and hence on business performance. We will also observe how markets may affect efficiency.

Table 18.1 Some forms of public-private partnership arrangements

Areas of involvement	*Full private ownership*	*BOOT-type arrangements*			*Public ownership with contracting out of supply*	*Traditional public ownership*
		BOO	*BOOT*	*BTO*		
Plan	P	P/G	P/G	P/G	G	G
Design	P	P	P	P	P/G	G
Finance	P	P	P/G	P/G	G	G
Construct	P	P	P	P	P/G	P/G
Operate/maintain	P	P	P	P	P	G
Ownership	P	P	P→G	G	G	G
Payment for services	C/G	C/G	C/G	C/G	C/G	C/G
Regulate	G	G	G	G	G	G

Key: P = private; G = government; C = consumers; BOO = build, own and operate; BOOT = build, own, operate and transfer; BTO = build, transfer and operate.

Source: Developed from EPAC (1995).

Under public ownership, government ministers determine objectives. The objectives may include political aims that are inconsistent with economic efficiency. On the other hand, most private firms aim (broadly) to maximise profits. In competitive markets, this objective is consistent with economic efficiency. Also, management incentives are generally clearer and more conducive to efficiency in the private sector than in the public. However, in less competitive markets Adam Smith's invisible hand becomes weaker and outcomes are less efficient. Maximising profits may mean increasing prices or reducing service quality.

Allocative efficiency

Supply of an efficient quantity of goods depends on efficient pricing and investment. As we have seen, efficient pricing is complex. While, in the short run, with a single supplier efficient supply usually requires short-run marginal cost prices. In the long run, businesses must have funds, or be able to access them, to invest to provide new services.

Historically, market failures, especially a lack of competition in the supply of basic infrastructure services, such as water and energy, provided a strong case for public ownership. Dominant private firms may set prices well above marginal cost and restrict output. Also, they are often weak on innovation. Restricted supply of key infrastructure services may impede the growth of other industries that depend on infrastructure. Such views supported the nationalisations of businesses after the Second World War.

However, publicly owned businesses may also set prices and invest inefficiently. The objectives of public enterprises are usually vague. For example, the statutory objective of Australia Post is 'to best meet the social, industrial and commercial needs of the Australian people for postal services'. There are no clear performance criteria and Australia Post is protected from direct competition. And the recent General Manager pocketed a salary of over half a million dollars.

There is also a risk of short-term political direction. Politicians may direct state-owned enterprises to provide subsidised services to favoured groups. Such directions depend more on political considerations than on efficiency objectives.

Without precise objectives and competitive constraints, managers of publicly owned businesses have considerable discretion over the quantity, quality, price and distribution of output. The managers generally have little incentive to increase the asset values that underlie the business. If performance is measured and rewards are determined by quantity of output, management has an incentive to maximise output regardless of marginal revenue and cost.

Recognising such limitations, governments may attempt to improve corporate governance by specifying objectives, giving agencies independence from short-term ministerial direction and improving incentives for managers. However, such reforms may not solve the underlying problems—the incentive structures in the public sector and the conflicts of interest between public sector managers, consumers and taxpayers. In the 1960s and 1970s the UK government issued three detailed White Papers on objectives and performance criteria for the major nationalised trading businesses, but the enterprises rarely adopted the recommended marginal cost pricing principles. Kay and Thompson (1986, p. 20) concluded: 'it is now apparent that at no time did this [the economic literature on nationalised industries] have any important influence on what nationalised industries actually did'. A more recent World Bank review of corporate reforms in state-owned electricity utilities in Mexico, New Zealand, the Philippines and South Africa (Irwin and Yanamoto, 2004) concluded that relationships between government and the utility will always be difficult while the government remains the utility's owner, but that corporate reforms do improve public utility performance.

Commonly public enterprises set prices too low, often under government pressure, and under-invest because they are under-funded and under-capitalised. The World Bank (2004a) reports similar practices and outcomes in developing economies. In wide-ranging international surveys, Hodge (2000) and Megginson and Netter (2001) show that, following

privatisation, firms tend to increase output, investment and capital efficiency. In Australia, both Telstra and the Commonwealth Bank working in competitive markets became much more customer focused after privatisation. On the other hand, in the late 1980s and early 1990s, the State Banks of Victoria and South Australia racked up billion dollar losses due to poor capital allocations and did not survive as independent institutions.

Overall, the evidence suggests that, with competition, private firms are more likely to produce the quantity, kind and quality of goods that consumers want than are public enterprises. Returns to shareholders and managers encourage private firms to supply market demands. Most firms face the threat of takeover. Many face international competition. Actual or potential competition ensures that prices are related broadly to costs. The need for survival and the goal of large profits spur a constant search for new products. Even in Eastern Europe, where institutional support for markets is often weak, privatised firms have reportedly been more innovative than state-owned enterprises (World Bank, 1996).

However, two major caveats are in order. First, the global financial crisis that erupted in 2007 revealed that private firms, especially those in the financial sector, created incentives for managers and traders to take risks, were liable to act inefficiently and dishonestly, and make costly errors. Many went bankrupt, thus contradicting the idea that private firms invariably make better investment and management decisions than public firms.

Second, post-privatisation outcomes are not necessarily efficient in uncompetitive markets. Private monopolies, such as Sydney Airport Holdings which owns and operates Sydney airport, may over-price and under-supply services. Energy prices on Australia have also soared since privatisation. To deal with market dominance issues, the private firms may be subject to detailed regulatory control. In such cases, a publicly owned business may produce as efficient a quantity and range of goods as would a private firm.

Productive efficiency

The evidence regarding productive efficiency is a little clearer. As we saw in Chapter 16, public services are often supplied at lower cost when outsourced. Several studies of the performance of public and private enterprises show similar results. In Australia, Domberger and Piggott (1986) found that private firms provided banking, urban bus services and airline services more cheaply than public enterprises. Internationally, Shirley and Walsh (2000) recorded that, out of 16 studies involving alternative ownership in competitive markets, 11 found that private ownership improves labour and total factor productivity, while five studies found no difference. In a survey of privatisation studies, Megginson and Netter (2001) concluded that private firms achieve large productivity improvements. In a World Bank study of 1200 utilities in 71 developing and transition economies, Gassner *et al*. (2008) found that privately operated utilities substantially out-performed state-run ones in operational performance and labour productivity. This indicates that cost savings are due to productivity and not just to paying lower wages to private sector workers. Rosen and Gayer (2014, p.66) also tend to the conclusion that private firms operate faster and at lower costs, but they qualify the conclusions by suggesting that they may produce lower quality services.[3]

However, analysts disagree on the extent to which productivity improvements should be attributed to changes in ownership as distinct from changes in competition. Vickers and Yarrow (1991) found that a competitive market structure had more effect on production costs than did ownership. On the other hand, Boardman and Vining (1989) examined the performance of the 500 largest manufacturing and mining companies in the world outside the United States, all operating in a competitive environment, and found that the private firms

[3] Some studies, for example Bhattacharya *et al*. (1994), find that the public sector is more efficient than the private, but these studies are a minority.

were significantly more productive and profitable than public agencies. In Australia, when public airlines or banks competed with private ones the private firms generally performed more efficiently. Where large productivity and cost differences have been found in other cases (e.g. in urban bus services, cleaning services, refuse collection and housing construction), it is unlikely that all differences are attributable to market structure.

A major reason why private firms usually produce goods at lower cost is the stronger incentive to reduce costs. Managerial salaries are usually related to profits. In a competitive market, bankruptcy and takeovers eliminate high-cost firms. Also, there is more monitoring and disclosure of information in private markets than in public ones. The incentives to reduce costs are lower when private firms enjoy significant market power. As the economist John Hicks (1935) remarked: 'the best of all monopoly profits is a quiet life'. But, in a monopoly market, when the market valuation of a company falls, takeovers occur and managers lose their jobs. This threat constrains managerial 'slack'. On the other hand, publicly owned businesses tend to have inflexible structures for labour, marketing and finance that are not conducive to efficient management. They also face little threat of termination. In public ownership there are no hard budget constraints.

One factor that arguably favours public production is the lower cost of capital. Private firms often pay several percentage points more for equity capital than government does. Quiggin (2010) argued that this represents a real cost saving because the public sector is more efficient at managing and diversifying risk than are private capital markets. Caution is needed here because government's power to borrow at low rates reflects in large part its capacity to tax households rather than superior risk management. To compare the real costs of capital, it is necessary to account for the deadweight loss of taxation, which is usually estimated to be at least 20 cents in the dollar (see Chapter 27). The high cost of equity in the private market reflects non-diversifiable systemic risks. Government agencies are also liable to systemic market risk, though perhaps less so than are private firms.

Concluding comments

As Rosen and Gayer observe (2014, p.66), there seems to be little systematic evidence on the cost differences between public and private production.

In public agencies, unclear objectives, ad hoc political directions, principal–agent issues and inflexible working structures tend to create inefficient processes. Even when markets are not initially competitive, privatisation of a public business may lead to more competition in the market, although this depends on how the privatisation is managed. Importantly, where there is competition for a market the government may extract efficiency gains from a competitive tender. The problem is then, with long-term businesses, how to prevent monopoly practices from creeping back into the operations of the successful firm.

If a competitive outcome is not feasible, the advantages of private operations are much smaller. There is still competition in the capital market if the private firm is inefficient. But a dominant private firm is likely to be regulated. In this case, the efficiency benefit of privatisation (if any) depends on the relative efficiency of public and private monopolies and the efficacy of the regulations.

Distributional Effects

Privatisation can have significant distributional impacts. It may affect prices, wages and the returns to owners of capital. For example, there is often concern that private ownership may:

- increase the prices of basic services;
- lower real wages of employees and lower employment; and
- create capital gains for investors at the expense of taxpayers.

Prices of services. For various reasons, government agencies often provide basic services at low prices to some or all households. These prices may rise after privatisation occurs and the constraints and incentives change. Alternatively, a public business may cross-subsidise some consumers. In Australia, publicly owned telecommunication businesses (Telstra and Australia Post) for many years used profits from urban services to subsidise rural services. These subsidies are likely to continue with Australia Post and the National Broadband Network Corporation. Privatisation threatens this subsidy because private firms may enter the profitable sector, drive down prices there and not supply the unprofitable part of the market.

To avoid these outcomes government can regulate the prices charged by private firms if it so wishes. However, price regulation is complex (as seen in Chapter 14) and most subsidy programs involve substantial transaction costs. Another strategy is to provide subsidies to service suppliers who lower their prices or to households by way of vouchers. When government subsidises the supply of a service, competition for provision of the service can ensure that the cost of the subsidy is minimised. When government subsidises consumers, consumers can choose the service of preference. Thus, the supply of subsidised services can be efficient and does not depend on public ownership.

Wages and employment. To maximise profits or simply to survive, private firms aim to minimise unit costs. Thus, privatisation may result in lower wages and employment in the privatised business, for example by contracting out work or by substituting capital for labour. After the privatisation of Japanese National Railways in 1983, the number of employees fell from 420 000 to 275 000 in three years. Most studies find that, following privatisation, employment falls in the privatised enterprise (Megginson and Netter, 2001). However, the longer-run labour outcome is less clear. If a privatised enterprise increases productivity and sales relative to the counterfactual (continued public ownership), real wages and employment may rise after a time.

From an economy-wide perspective, privatisation may redistribute and even increase employment rather than reduce it. Saunders (1985) found no relationship between total employment and the size of the public sector in OECD countries. It is generally more efficient to pay labour the value of its marginal product and to compensate low-income earners through the tax transfer payment system than to pay premiums to employees in the public sector. Departures from productivity-related payments tend to result in misallocation of labour and to reduce GDP and consequently incomes available for distribution. If privatisation is efficient but reduces the wages or other entitlements of workers in a previously state-owned business, government can provide compensation if it so wishes. For example, some UK privatisations allowed employees to acquire shares in the privatised organisation on favourable terms.

Investors and taxpayers. The third concern is that privatisation may create excess benefits for investors. In the 1980s the UK government underpriced the sale of several businesses. The share prices of Associated British Ports, British Aerospace, British Telecom and Cable and Wireless more than doubled, relative to market movements, after their float. Less than 4 per cent of the population benefited directly from the underpricing of public assets, although others benefited indirectly through their investments in financial institutions. In Australia, the share price of the Commonwealth Bank appreciated far more than the market index after floating in 1996. In these cases, shareholders gained considerably at taxpayer expense. On the other hand, investors who purchased the second and third tranches of Telstra shares have experienced substantial capital losses.

Shares in privatised businesses are underpriced when (1) returns to the privatised entity turn out to be higher than expected or (2) government wants certain revenue from the sale and is unwilling to auction the shares. In both cases new shareholders are rewarded for bearing the

risk and taxpayers are losers. Government may achieve a higher price by selling shares by tender to the highest bidders. But it may lose from not having a guaranteed, under-written, price. And sometimes share prices fall post-privatisation as profits are less than expected.

Of course, an initial fair market price along with a widespread shareholding may not prevent another company, even one controlled by a single person, taking over the privatised enterprise. Entrepreneurs are often willing to pay a premium over the established market price to obtain control of a company. Although privatisation usually increases competition, it creates the potential for concentrated private ownership of major enterprises and acquisition of wealth in a few hands unless government has a strong competition policy.

Broader social issues

Turning to broader social issues, libertarian advocates of privatisation claim that the markets are important not only because they allocate resources efficiently but because they contribute to liberty. They argue that state control over private consumption is coercive and illiberal. They advocate privatisation to reduce the power of the state over its members.

On the other hand, social democrats may argue that collective provision of basic social goods is preferable to the individualistic forms of behaviour encouraged by private markets. They may view public enterprises as enhancing communal interest. For example, many people argue that privatisation of health care services diminishes the sense of community. In Australia, as in most European countries, there is strong support for public health and education services, as well as for national transport and communications, because these services are felt to be important elements of the national social infrastructure.

Evaluating Changes in Ownership

Using the cost–benefit criterion, a change in ownership is **efficient** if the net social benefit (NSB) is positive. This means that the net benefits of production under new ownership must exceed those under existing ownership. The NSB of a privatisation can be expressed as:

$$NSB = \sum_{i=1}^{m}(WTP_P - RC_P) - \sum_{j=1}^{n}(WTP_S - RC_S) + \Delta TPE \qquad (18.1)$$

where WTP represents the prices that consumers are willing to pay for goods, RC denotes the resource (opportunity) costs of production, TPE is any third-party effects, Δ denotes change, subscripts P and S represent private and state-owned businesses respectively, $i = 1 \ldots m$ private goods and $j = 1 \ldots n$ publicly provided goods.

Clearly, if the same output is supplied under public and private ownership, NSB reflects simply productive efficiency. Equation 18.1 allows for different goods, or different quality goods, to be produced after privatisation. Importantly, estimates of NSB should be based on the real opportunity costs of factors of production, not on nominal wages paid to labour or other nominal expenses where these differ from real opportunity cost. If workers of equal productivity are paid less in the market than in the public sector, this is a distributional impact but not an efficiency gain. All values should be estimated over a relevant period and discounted into present values.

The net efficiency change must equal the sum of changes to all affected parties. Thus, NSB must equal the sum of consumer and producer surpluses (for existing and new owners of capital and for labour) plus any changes in third-party effects. Thus,

$$NSB = \Delta GY + \Delta SY + \Delta LS + \Delta CS + \Delta TPE \qquad (18.2)$$

where GY is government income, SY is private shareholder income, LS is labour surplus and, CS is consumer surplus. Of course, one or more of these values may be negative, notably LS. Again, all values should be in discounted present values.

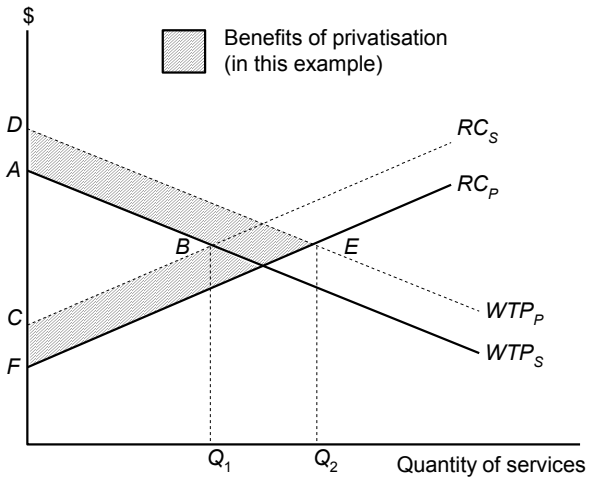


Figure 18.1 Estimating the net social benefit of a change in ownership

Both Equations 18.1 and 18.2 require estimates of the demand for goods (willingness-to-pay values) and opportunity costs. Equation 18.2 requires and provides more information, including distributional information that may interest policy makers. Either way, it is necessary to forecast the outcomes with a privatised business and those that would occur under continued public ownership.

These ideas are illustrated in Figure 18.1. This diagram shows separate willingness-to-pay (demand) curves for the services produced by a state- and privately-owned business (WTP_S and WTP_P respectively). There are also two resource cost curves (RC_S and RC_P respectively). There are no *TPEs* in the figure. In this case, the NSB of privatisation is the difference between areas *DEF* and *ABC*. Of course, NSB is not always positive.

Note that Equations 18.1 and 18.2 are efficiency equations. Neither equation includes any equity weighting. As discussed in Chapter 8, there is no technical basis for introducing weights into cost–benefit analysis. However, policy makers may give more weight to some costs and benefits than others.

Distributional effects. If a public business is sold into a competitive capital market, government may receive most of the expected increase in profits in capitalised form via the sale price, including profits from any expected increases in prices. If it is a state government sale, the Commonwealth government may share the gains via increased taxes. On the other hand, new shareholders benefit if government underprices the shares or if future profits exceed expectations at the time of sale.

Turning to labour, if labour in publicly owned enterprises receives economic rents (paid higher wages than in the private sector), it may lose income when the business is privatised. Labour employed by a private firm is unlikely to receive significant economic rent.

Consumers may benefit from privatisation because private firms have more focus on consumer demands than do public agencies. Thus, more services may be available. The net benefit of these services is the consumer willingness to pay for them less the costs of producing them. On the other hand, a private firm may lower quality of service or raise prices to appropriate some of the ownership benefits. Also, the distribution of consumer benefits may change with changes in cross-price subsidies unless regulations prevent this.

Financial issues. When government sells a business, it gives up a productive asset and foregoes the related income stream in return for a financial asset. The net worth of government rises if the capital revenue from the sale exceeds the present value of the income stream forgone.

However, an increase in government's net worth does **not** itself signify an increase in economic efficiency. The financial value of a business may rise with the sale of a business because the private firm can raise prices more easily or employ labour at lower wages. These pecuniary changes may not reflect real efficiency benefits. Conversely, government could give a business away and lose substantial net worth, but society could benefit from the efficiency gains of privatisation, with the gains shared between the new owners and consumers.

Governments may sell assets to reduce a cash budget deficit or to finance creation of new assets without raising taxes, public borrowing or interest rates in the short run. In effect, government is financing its operations by selling equities instead of bonds (or raising taxes). Where government sells an asset to create a new one, the process is sometimes called "asset recycling".

Generally, privatisation is justified by efficiency benefits, not by short-term revenue gains. If the revenue is used to fund a recurrent budget deficit rather than invested, taxes and borrowing rise over time as the income stream from the public business is no longer available. If the change in ownership does not change the productivity of the business, the sale of a productive public asset at its full market price has the same effect on net public worth, taxes and interest rates as does the sale of a government bond of similar value. If privatisation increases productivity and government appropriates this gain via the sale price, the sale increases net public worth and reduces government's revenue-raising requirements. But this does not change the total gain from privatisation. It simply divides the gain between the public and private sectors.[4]

However, this case rests on the assumption that the government can borrow or equivalently that the public is indifferent between holding shares in a privatised business and holding government bonds. Where government is over-indebted and cannot borrow, it may have to sell productive assets. Greece is an example.

Public-Private Partnerships

In recent years governments have combined increasingly with private firms to finance, develop and manage public infrastructure, especially for roads and transport infrastructure but also for utilities, hospitals, schools and even prisons. Araujo and Sutherland (2010) report that capital investment in PPPs in OECD countries rose from around $10 billion per annum in the second half of the 1990s to well over $20 billion per annum between 2002 and 2007. About a third of the PPPs were in the United Kingdom.

Such privately financed deals are not new. In the 19th century most of the infrastructure of France was constructed using private equity and, by 1860, infrastructure companies accounted for 70 per cent of the total capitalisation of the Paris stock exchange. However, a feature of today's arrangements is that government often agrees to pay for some, or all, of the outputs provided. For example, government may pay for bed days in a hospital or for prisoner days in a prison or pay a shadow toll for each user of a highway. Thus, tax revenues pay ultimately for some or all use of the facility. Often the infrastructure is constructed and services are supplied by the private sector for an agreed period at the end of which the facility becomes public property (see Table 18.1).

[4] If government revenue is increased, there may be a second order gain in efficiency if marginal tax rates (and the deadweight loss of taxation) can be reduced.

Other key features of PPPs are that they are typically contracts with a single company or consortium and they are long-term contracts often over 20 to 30 years or even more. This distinguishes them from the large numbers of contracts that government makes each year or even each month with private providers.

Benefits of PPPs. PPPs aim to achieve efficiency benefits from competition and private sector involvement in construction and operations while retaining more public control than can be achieved by regulation of private enterprises. In addition, ownership and management structures may be devised with the aim of ensuring that tasks are undertaken and risks are borne by the parties best able to do the work and bear the risk. Thus the operators have incentives to provide efficient design, construction and maintenance. Proponents of PPPs argue that this introduces economies in construction and more cost-effective services. A PPP structure may also provide some financing to a public project that would not otherwise be available when the provider of the finance has more confidence in the ability of a private company to manage the service in a financially viable manner. However, government can borrow to finance most projects given its power to raise revenue and repay loans via taxation.

Infrastructure Partnerships Australia (2007) employed consultants to review 21 PPP projects and 33 traditional projects. The report found that the cost overrun on contracted PPP projects of $4.9 billion was only $58 million compared with a cost overrun of $673 million on $4.5 billion of traditional projects. They also found that the PPP projects were completed on average 3.4 per cent ahead of time whereas the traditional projects were completed on average 23.5 per cent behind time. In addition, as measured by the amount of public data, the PPPs were far more transparent than the traditional projects. In the UK, the National Audit Office (2003) found that only about 20 per cent of PPP projects were over budget or over time compared with just over 70 per cent of traditionally procured projects.

Costs of PPPs. On the other hand, there are also several costs associated with PPPs. One is the high cost of finance. The cost of private equity, the most risky element of the funding package, may be 6–7 percentage points above the government borrowing rate. This is an economic cost required to compensate parties for bearing risk. The private sector bears regulatory risk (the possibility that government may change regulations) and it may be less efficient at managing risk than governments which can spread risk more widely over projects and taxpayers. However, as noted above, when public loans are repaid from taxes the borrowing rate understates the true economic cost of capital.

A second potential cost of PPPs relates to networking and economy of scale issues, especially when a business is a natural monopoly. When infrastructure is part of a network, breaking up the network into separate entities may involve operating diseconomies or high transaction costs. For example, there may be costs associated with operating parts of the urban rail network on a separate basis or with breaking up a vertically integrated operation into separate operations. In these circumstances private parties cannot accept market risks without protection from government, which effectively shifts some risks back to government.

Thirdly, because of the length of most PPP contracts, it is difficult to write precise and enforceable long-term contracts that clearly specify output requirements. When circumstances change, the government has to negotiate changes with a single major partner. As Araujo and Sutherland (2010) observed, governments are highly susceptible to hold-ups by private partners re-opening negotiations or to simple hold-ups when government seeks any variation in the terms of the contract. It is difficult to introduce competition during the life of the PPP. A lawyer in a large construction company observed recently to the writer that ‘the legal section makes all the profits’. What he meant was that company bid very leanly for its initial contracts and made most, or all, of its profits on the variations.

Fourthly, government is almost always taking on some contingent liability. If the private operator fails to perform according to the contract, the government frequently has to honour the creditors and ensure that the promised services are available to the public.

A final cost of PPPs arises from the secret bilateral negotiations between the government and a private firm, which are a common feature of PPPs. Often the final agreed contract follows months of negotiation and contains hundreds of clauses that are not in the initial tender. The private firm aims to pass the risk back to the government, usually by obtaining revenue guarantees or by restricting competition. Thus, the operators of the Sydney Harbour tunnel were guaranteed public revenue. The Sydney cross-city tunnel was constructed with private finance on the basis that numerous alternative routes would be closed. The Sydney M2 freeway was built with a government promise not to allow any new competitive infrastructure. Almost always these negotiations result in some economic inefficiency or, as Adam Smith would have described it, in some form of 'conspiracy against the public interest'.

Summary. Public–private partnerships of some kind are a feature of numerous government activities. They can be efficient, save millions of dollars in cost and provide services to the public that they would not otherwise receive. However, the private sector requires a high return on capital. There are also many practical pitfalls in negotiating effective contracts. Therefore, while some PPPs are very successful, the general public is not always a net beneficiary of these deals.

Summary

- Over the last 30 years, governments worldwide have sold many businesses into private ownership. Governments are also drawing increasingly on the private sector to finance, construct and operate infrastructure that was traditionally financed, constructed and operated by the public sector.
- Incentives in the public and private sector are such that the transfer of businesses into private ownership is likely to be efficient when markets are competitive. The evidence supports this. The transfer of assets has generally resulted in an increased range of services and a lower cost of supply.
- However, when markets are not competitive all forms of ownership are potentially imperfect. If a business has natural monopolistic characteristics, the competitive parts may be separated and privatised. But if the whole business is privatised, regulation is generally necessary. In this case, what matters is the optimal combination of ownership and regulation.
- Ignoring third-party effects, privatisation is efficient if the sum of producer and consumer surpluses (including impacts on new and old owners, consumers and labour) is positive. An increase in the net worth of government is not a sufficient or a necessary condition for an efficient outcome.
- Privatisation may have significant distributional effects including lower wages and employment in the privatised firm and increased prices. However, privatisation may not reduce total employment.
- Also, if government wishes to provide subsidised services to selected groups, it can make community service payments to suppliers of the service or provide financial support direct to consumers.
- Public–private partnerships can be efficient, reduce costs and provide services to the public that they would not otherwise receive. However, the private sector requires a high return on capital and there are many practical pitfalls in negotiating effective long-term PPP contracts.

Questions

1. What incentives might make private firms more efficient than state-owned enterprises? Why is market structure important to incentives?
2. The Australian government sold Telstra, the major telecommunications company in the country. How would you analyse the efficiency and distributional effects of this sale?
3. Irving Fisher (the Fisher Separation Theorem) postulated that the investment and financing decision can be regarded as independent of each other. Is this a correct view?
4. Do post-privatisation share prices tell us anything about the nature of the privatisation?
5. Suppose that government can fund a project by issuing bonds at 6 per cent (public funding) or equity finance on which lenders require an average 10 per cent rate of return (private funding).

   Why might government prefer private funding to public funding despite the higher cost of capital? Does the interest rate differential represent a real economic cost of the project?
6. If government sells a business and thereby avoids a budget deficit, is this a benefit of privatisation?
7. Are financial savings to government a measure of economic efficiency?
8. Public hospitals provide a range of medical, diagnostic and accommodation services. How might government decide on the optimal provision of these services by public employees or by private contracts? What kind of outcomes might occur?
9. If private contractors can run prisons and asylum detention centres more cheaply than can the government, does this mean that these prisons and detention centres should be run by the private sector?
10. Is public ownership necessary for an equitable supply of basic household services?
11. The Australian government has recently established a public company, the National Broadband Network Corporation, to provide a national broadband infrastructure with services to all homes and businesses in Australia. Most of the construction work is being done by private companies. Does this mean that the NBN is really a PPP?
12. What kind of contingent liabilities may a government incur in a PPP arrangement?
13. If a private company partner in a public private partnership goes into receivership, does this necessarily mean that the project developed by the PPP is inefficient?

Further Reading

Araujo, S. and Sutherland, D. (2010) *Public Private Partnerships and Investment in Infrastructure*, OECD Economics Department Working Papers No. 803, OECD, Paris.

Chan, C., Forward, D., Roper, H. and Sayers, C. (2009) *Public Infrastructure Financing*: *An International Perspective*, Productivity Commission Working Paper, Productivity Commission, Melbourne.

Grossman, S.J. and Hart, O.D. (1986) 'The costs and benefits of ownership: a theory of vertical and lateral integration', *Journal of Political Economy*, 94, 691–719.

King, S. and Pitchford, R. (1998) 'Privatisation in Australia: understanding the incentives in public and private firms', *Australian Economic Review*, 31, 313–328.

Megginson, W.L. and Netter, J.M. (2002) 'From state to market: a survey of empirical studies on privatisation', *Journal of Economic Literature*, 39, 321–389.

Quiggin, J. (2005) 'Public-private partnerships: options for improved risk allocation', *Australian Economic Review*, 38, 445–450.

Quiggin, J. (2010) 'Privatisation', pp. 174–205 in *Zombie Economics*, Princeton University Press, Princeton, NJ.

Rosen, H. and T. Gayer (2014) *Public Finance*, 10th ed., Chapter 4, McGraw-Hill, New York

Vickers, J. and Yarrow, G. (1991) 'Economic perspectives on privatisation', *Journal of Economic Perspectives*, 5, 111–132.

World Bank (2004) *Reforming Infrastructure: Privatisation, Regulation and Competition*, World Bank and Oxford University Press.

Chapter

19

Economic Infrastructure: Transport

It is good to be out on the road, and going one knows not where.

John Masefield, *Tewkesbury Road*

Use and Pricing of Transport Infrastructure ♦ Investment in Transport Infrastructure ♦ Role and Regulation of the Private Sector in Transport

There are many forms of economic infrastructure. In this chapter, we discuss transport as an example of the economic issues arising in the use, supply and management of infrastructure. With any infrastructure, there are two key issues: how to use the existing infrastructure and how to augment it.

Using transport infrastructure efficiently is essentially a pricing issue. When traffic volumes are low, use of infrastructure is non-rival and efficient prices may be low. However, prices should allow for externalities such as congestion, accidents, noise and air pollution. On the other hand, when capacity is limited, high prices are efficient. But introducing equity, low public transport prices may reduce social inequalities and exclusion (Stanley *et al*. 2011).

Regarding investment, standard cost-benefit principles apply. But special issues arise as transport infrastructure may provide economic benefits that are not reflected in user benefits.

Turning to management, most railways, ports, airports, pipelines and some major roads are natural monopolies. These market failures provide a strong case for government involvement. Nevertheless, there may be a role for private investment or management.

In this chapter, we discuss each of these issues in turn: use of existing infrastructure and the role of pricing; investment in transport infrastructure; and the role and regulation of the private sector in supplying and operating transport infrastructure.

Use and Pricing of Transport Infrastructure

Following our core principle for efficient use of capital, use of transport infrastructure is efficient when the net social benefit (NSB) from use is maximised. NSB is simply:

$$NSB = \sum_{i=1}^{m} B_i - \sum_{j=1}^{n} C_j \qquad (19.1)$$

where B is benefits, there are $i = 1\ldots$m benefits, C is costs and there are $j = 1\ldots n$ costs. Most benefits accrue to users of the infrastructure though there may also be third-party benefits. NSB is maximised and infrastructure is used efficiently when the marginal benefit of use equals the marginal opportunity cost inclusive of external costs.

Prices signal how transport infrastructure can be used most efficiently. General pricing principles for public services were described in Chapter 17. We summarise below the main points with special reference to transport infrastructure before describing some applications in the transport sector.

General pricing principles

The default principle for pricing public services is short-run marginal cost (SRMC) pricing, i.e. setting user charges equal to the marginal cost of the service (see Box 19.1). This maximises the NSB from use of infrastructure. If use is non-rival and there is no marginal cost of use, there would be no user charge. If there is a cost of use, including externalities such as congestion costs, the user charge should reflect the marginal social cost of use.

However, this assumes either no significant funding shortfall to cover fixed costs or that a funding deficit can be funded at little deadweight cost. If these assumptions do not hold, a transport agency may be required to meet a revenue target, such as full cost recovery. In this case, it would raise the extra revenue with minimum distortion of transport activity by adopting price mark-ups or multi-part tariffs as described in Chapter 17.

Price mark-ups may also be appropriate if competitive services are priced above SRMC. It would be inefficient to price public services at marginal cost and attract custom away from a private operator that can supply similar services at lower cost. Equation 19.7 in Box 19.1 provides the pricing formula that maximises NSB in these circumstances.[1] The price mark-up for the subject service depends on the size of the mark-ups and the cross-price elasticities of demand for the substitute services. Pragmatically, if private services are close substitutes (the cross-price elasticities are high), long-run marginal cost (LRMC) pricing is generally appropriate for the public transport service.

Road user charges

There are various kinds of road user charge. The purest form is a charge per vehicle kilometre for road use at a regulated time. Closely related charges are road tolls and cordon charges that permit access into a specified area at regulated times. Other charges include area licences that permit entry into an area over a longer period or fuel taxes.

The alternatives to road user charges are generally less efficient ways to ration road space because they do not focus on the spatial and timing problems of congested roads. Selective toll roads and cordon charges encourage drivers to choose less efficient routes or less preferred destinations. Fuel taxes are not related to use of congested roads. Other charges such as vehicle registration fees or sales taxes may contribute substantially to road construction and maintenance, but they are fixed charges and not charges for road use. Fixed charges may have a minor effect on vehicle ownership, but have little effect on road use.

Following our previous discussion, the efficient pricing principle is that the road user charge should reflect the marginal social cost of use. If a road is not congested and use is non-rival, efficient use requires no charge unless there is road damage or negative external effects (such as safety risk or air pollution). Most passenger vehicles do minimal pavement damage. The damage increases at the third or fourth power of axle weight. Given that the axle weight on a car is typically about 450 kilograms and the axle weight on a truck is typically some 8000 kilograms, a truck axle causes about 6000 (18^3) times as much road damage as a car axle. Trucks are also involved in a disproportionate number of fatal accidents. Thus, there is a strong case for road user charges for trucks even when there is no congestion.

[1] This is equivalent to Equation 17.2.

Box 19.1 Pricing principles for transport services[2]

Suppose that net social benefit (NSB) from a transport service is expressed as:

$$NSB = \int_0^Q P(X)\,dX - Q.AC(Q) \qquad (19.2)$$

where $P(X)$ is the inverse demand curve, X is a variable of integration, Q is the volume of trips and $AC(Q)$ is average cost which varies with trips made. The first term on the right hand side is customer willingness to pay for services (the area under the demand curve) and the second term is the cost of the service, which is here borne by the service provider.

To maximise NSB, the first derivative of Equation 19.2 with respect to Q must equal zero. Rearranging the result:

$$P = AC + Q.(dAC/dQ) \qquad (19.3)$$

Equation 19.3 states that the price of a service must equal the sum of the average cost of servicing a user and the change in this cost from servicing an extra user.

Now suppose that, as with a road, the users provide and pay for their own services and bear these costs. NSB can now be expressed as:

$$NSB = \int_0^Q P(X)\,dX - Q.AUC(Q/L) - LCC(L) \qquad (19.4)$$

where AUC is the average cost to a user of operating a vehicle on the road, L is the number of units of capacity (such as lanes on a highway), CC is the average amortised cost to provide a unit of the facility and the other variables are as before. User costs rise with congestion and so depend on the ratio of (Q/L).

To maximise NSB from use of the infrastructure, the level of capacity is held constant and the first derivative of Equation 19.3 with respect to Q is set to zero. Rearranging the result, the following is obtained:

$$P = UC + Q.(dAUC/dQ) \qquad (19.5)$$

The efficient road user charge is equal to the second term on the right-hand side of the equation. This is the change in average user cost as a result of serving an additional user, which is the marginal cost at the efficient level of road use.

If these efficient prices do not cover total cost, Ramsey's formula to raise sufficient revenue at least cost (see Chapter 17) is:

$$(P_i - MC_i)/MC_i = k/\eta_{di} \qquad (19.6)$$

where P_i is the mark-up price for customer i, MC_i is the marginal cost of serving customer i, η_{di} is the compensated price elasticity of demand for customer i and k is a constant determined by the revenue required to meet the budget.

Price mark-ups may also be appropriate in second best cases where competitive transport suppliers charge in excess of marginal cost. The efficient price here is:

$$P_i = MC_i - \Sigma_j\,[\eta_{ij}/\eta_{di}).(Q_j/Q_j).(P_j - M_{Cj})] \qquad (19.7)$$

where i is the subject service, j is the competitor(s) service, η_{ij} is the cross-price elasticity of demand for good j with respect to the price of good i and the other variables are as above.

When roads are congested, efficient road use requires that the user charge should reflect the congestion cost caused by the marginal road user. This is demonstrated in Figure 19.1 overleaf. The generalised cost (GC) of a trip is shown on the vertical axis. GC is the sum of costs borne by the trip-maker, including the cost of travel time, trip time reliability, vehicle operating costs and any road tolls. The demand for trips rises as GC falls. The marginal private cost (MPC) per trip, measured in terms of generalised trip costs, increases when congestion occurs. When congestion occurs, the marginal social cost (MSC) of a trip exceeds MPC because each vehicle entering the traffic stream imposes a cost on other vehicles, which increases with traffic volume.[3]

In the absence of road pricing or controls on road use, there will be Q_M trips per period at an average trip cost of \$$D$. But at this traffic level, MSC exceeds MPC. The efficient quantity of trips is Q_E, the point at which the marginal travel benefit equals MSC, where the average trip cost is \$$G$. The deadweight loss of congestion equals area BCF. This is the difference between the marginal social cost of the excess trips and the benefits of these trips. This can be avoided by setting a road user charge equal to the marginal external cost *at the efficient level of traffic* (equal to BH in Figure 19.1). Ideally the charge would vary with changes in congestion over time.

---

[2] The presentation in Box 19.1 is based on Gomez-Ibanez (1999). This presentation does not include externalities.

[3] For ease of exposition, other externalities of trip making are ignored here.

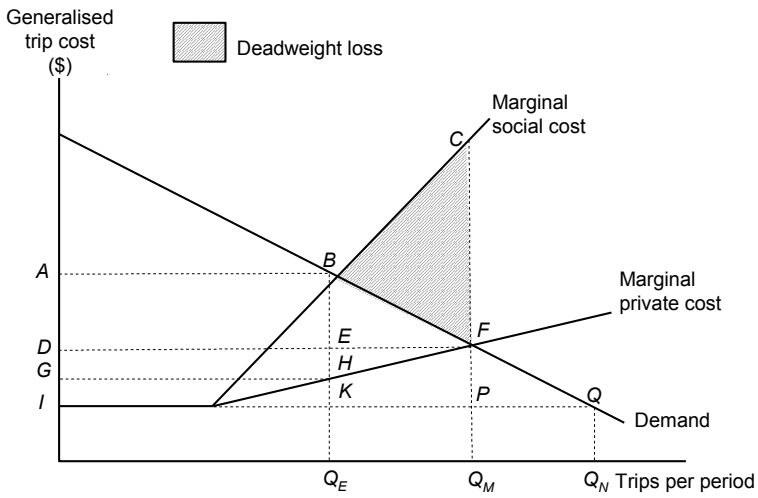


Figure 19.1 Traffic congestion costs

Efficient pricing has two main benefits. First, it eliminates excess congestion costs equal to area *BCF*. Second, it may defer the need for new infrastructure. When existing capacity is used inefficiently, there is greater need for increased capacity. In relation to Figure 19.1, suppose an investment in road capacity increases road use from an inefficient level Q_M to Q_N and that there is no congestion at Q_N. The benefits would equal areas *DFPI* + *FPQ*. On the other hand, if road use is efficient, road use increases from Q_E to Q_N. The benefits of the increased road capacity now equal areas *GHKI* + *BQK*. Although there is a greater increase in road use, the benefits from the increase in capacity are lower because the existing capacity is better utilised. Accordingly, new investment may be deferred. On the other hand, if inefficient road use is endemic, increasing capacity does produce the benefits shown above, albeit as a second-best solution.

However, road user charges present two challenges. One is to introduce a network-wide electronic tagging (GPS) system. Selective charging for roads encourages inefficient road use as some road users choose inferior free roads. The second challenge is political. If there is high congestion, the revenue from road user charges (area *ABHG* in Figure 19.1) may greatly exceed the benefits to continuing road users (area *DEHG*). This does not affect the NSB from a road user charge because, in cost–benefit accounting, the revenue is a transfer payment rather than a resource cost. However, the excess of revenue over benefit may create a political problem unless a popular strategy can be devised such as using the funds for public transport or reducing vehicle registration fees.

Prices for public bus and rail services

Pricing features prominently in debates about public bus and rail services. This reflects concerns about access, large deficits and efficiency. A central issue is again whether prices should be set equal to short-run marginal cost or perhaps even below SRMC or to recover all costs. Here we consider three arguments for SRMC or even sub-SRMC pricing: scale economy, second-best and equity arguments. We then note some arguments against subsidies.

Where there are fixed costs and some scale economies, efficient use of bus or rail would require SRMC pricing. However, the efficiency benefit of SRMC may be small for two main reasons. First, scale economies may not be large and marginal cost may not be much lower than average cost. Gomez-Ibanez (1999) estimated that the marginal cost of rail service was

usually at least 70 per cent of average cost. For buses, including congestion effects of buses, marginal social cost may exceed average financial cost. Second, efficiency gains from SRMC pricing are usually small because demand for most bus and rail services is price inelastic. Typically, the direct price elasticity for urban public transport is about –0.3 (Goodwin, 1992).

Turning to second-best issues, the NSW Independent Pricing Tribunal (2015) proposes that rail and bus services should be priced at below average cost when private road users do not pay for the marginal social cost of road use especially in peak hours. Substitute services should be under-priced to attract road users and reduce the negative externalities of road use. However, this argument should be treated cautiously. In many places, private motorists pay high fuel excise changes, road tolls and parking charges which may partly, or wholly, substitute for road user charges. Importantly, the cross-price elasticity of demand with respect to private vehicle use is usually low. When a rail or bus operator lowers the price of their services, they attract patronage mostly from other public transport modes and some new passengers (which creates a deadweight loss); they may attract few road users. When roads are under-priced, raising road user charges is more efficient than a second-best solution (reducing bus and rail fares) and, if feasible, should be preferred.

Equity considerations may support transport subsidies, including pricing below SRMC. Many bus and rail users have low incomes. And some trips, to work or medical services, may be viewed as necessities. However, if government wishes to subsidise some public transport users, travel vouchers to targeted individuals would be more efficient than general trip subsidies. Vouchers allow trip makers to choose their preferred mode. Also, they do not distort price relativities for most users. Reducing transport fares below marginal cost for all users is not efficient and may not be equitable.

On the other hand, there are arguments for full-cost recovery pricing of public transport services. Loss of revenue may reduce quality of service unless government supplies adequate deficit funding. Management disciplines are hard to maintain. If subsidies are provided, they may be absorbed into higher cost structures rather than improved services. Moreover, revenue deficits must be financed from tax revenues that themselves have deadweight costs.

In summary, SRMC pricing for rail or bus services is generally efficient because it maximises use of services. Second-best and equity considerations may justify below-SRMC pricing in some places. However, low prices may encourage inefficient use. Also, SRMC pricing may create management issues and generally involves raising taxes to fund deficits. Where services are competitive and cross-price elasticities are high, competitive neutrality (efficiency) principles suggest public supply should be based on full cost recovery.

Ideally pricing policy would be based on estimates of the benefits and costs of alternative pricing structures. Glaister and Lewis (1978) estimated optimal second-best fares for London rail and bus services separately for peak and off-peak services. Allowing for operating costs, external pollution and congestion costs of all modes, own-price and cross-price elasticities of demand, they found that peak fares for London rail and buses were too low and that off-peak fares were too high. Similar exercises could provide useful information to decision makers elsewhere. However, this is a complex exercise.

Charging for airports

The principles for efficient use of airports are similar. However, their application involves different issues.

Airports provide various runway, terminal and related services. Airport managers recoup expenses by landing fees, rental charges for use of terminals and parking charges which reflect monopoly powers. Landing fees are usually based on aircraft weight. This may be reasonable in off-peak periods. Aircraft weight reflects potential runway damage, although this is small for all aircraft, and may be viewed as a proxy measure for the number of passengers and amount of freight using the terminal facilities.

However, critically, in peak hours runway capacity is the limiting factor on airport capacity. Weight-based costs are not then an efficient basis for airport charges. When an aircraft lands or takes off in congested conditions, it imposes delay costs on other traffic. Slow aircraft, usually small aircraft, impose more delays and larger costs because they reduce the landings and take-offs that can be achieved in peak periods. When one aircraft displaces another, with no change in congestion, the real cost of using a time slot is the value of the slot to the displaced aircraft. If the displacement increases delays for other aircraft, the efficient charge is the sum of the price the displaced aircraft would pay for the slot and the additional congestion cost. Accordingly, the first step towards efficient use of scarce runway capacity is to establish charges that reflect these costs. An auction for time slots would ensure that the airline that places the highest value on a landing or take-off slot will obtain this slot.

These peak-hour landing fees may exceed average runway and terminal costs. However, from an efficiency perspective these fees should rise with demand for runway use until the excess demand for runway capacity justifies the cost of a new runway (see below). Efficient pricing produces an efficient use of existing capacity and creates signals for efficient investment in additional capacity.

A further issue is the monopoly position of most airports. This enables airport authorities (public or private) to extract economic rents from airlines and air passengers. This may warrant some form of price regulation (see the discussion below).

General conclusions

Marginal social cost pricing is an important efficiency principle. When transport facilities are not fully used, this pricing principle encourages use of the spare capacity. The inclusion of externality costs (see Box 19.2 on p. 331) ensures that trip makers consider the full costs of their travel. When transport facilities are heavily used, and especially in congested conditions, marginal social cost includes congestion costs. These prices allocate scarce capacity efficiently and provide an important signal for investment.

On the other hand, full cost recovery pricing ensures efficient allocation of resources when substitute services are charged at full cost, is competitively fair, encourages efficient management and avoids taxation funding of deficits.

Investment in Transport Infrastructure

Investment in infrastructure is efficient when the expected present value of the benefits of increased capacity exceeds the present value of the costs, inclusive of externalities. Benefits typically include cost savings for existing users of the infrastructure and benefits to new users. More precisely, the most efficient level of capacity is achieved when the present value of an extra unit of capacity just equals the cost of providing that unit. However, marginal adjustments are difficult to achieve when investments are lumpy (indivisible) as they are with road lanes, rail tracks and airport runways. Efficient timing is achieved when the net benefit of increased capacity exceeds the opportunity cost of capital employed.

Efficient investment is linked to efficient pricing. Low prices may create excess use of infrastructure before additional capacity is justified. Figure 19.2 shows existing capacity (S_1) and new capacity (S_2) with a lumpy asset such as an airport runway. Charging at SRMC, there may appear to be sufficient demand for a new runway even with D_1 level of demand. Actually, a new runway will be financially viable only when demand rises to D_3. In this case the efficient price with existing capacity would be P_3. However, a new runway could be justified with less demand than D_3, but with more demand than D_2. At a market clearing price for the new runway slightly below P_2, some users could gain consumer surpluses. If these surpluses were large enough to offset the loss to the airport authority, the net social benefit would be positive. Of course, the evaluation should allow for shifts in demand over time.

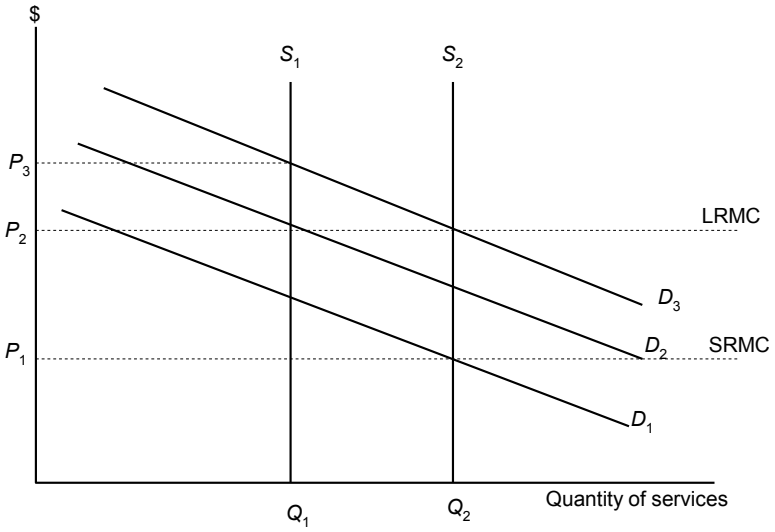


Figure 19.2 Pricing and investment

We describe below how to evaluate road, rail and airport infrastructure along with some examples. Slightly different issues occur in each mode.

Evaluation of road investment

The major costs of a road upgrade or new road are land, property construction and maintenance costs. There may also be environmental costs. On the other hand, nearly all the benefits are associated with lower road user costs, especially travel time savings.

As shown in Figure 19.3a overleaf, the demand for road trips between any two points in a period is a function of the generalised trip cost (GC). By far the largest component of GC is travel time. The value of travel time was discussed in Chapter 8 (Box 8.3). The UK Department of Transport (2010) and Hensher (2011) provide more detailed discussions of the value of travel time savings. GC also includes vehicle operating costs (VOC); these are the marginal VOC associated with the trip, which are largely fuel costs. User costs also include road user taxes and charges, if applicable.[4]

Figure 19.3a shows two sets of generalised cost (GC_1 and GC_2) representing user trip costs before and after a road upgrade. As trip cost falls from GC_1 to GC_2, trips increase from T_1 to T_2. Approximating demand by a linear schedule, the change in consumer surplus (ΔCS) is:

$$\Delta CS = T_1 (GC_1 - GC_2) + (T_2 - T_1)(GC_1 - GC_2)/2 \qquad (19.8)$$

Equation 19.8 includes gains to existing users and estimated benefits of generated trips. The latter include trips diverted from other routes, modes and destinations, as well as new trips.

However, Equation 19.8 does not allow for changes in producer surplus. Generalised cost includes transfer payments, such as fuel excise and road tolls that do not consume resources. Panel (b) introduces 'resource cost' schedules (RC_1 and RC_2) before and after the road upgrade respectively. As the term implies, resource costs reflect the use of resources. This includes travel time and vehicle operating costs, but not transfer payments. The shaded area in

[4] In principle, generalised cost should include an allowance for accident costs. However, these costs are often treated separately in road evaluations.

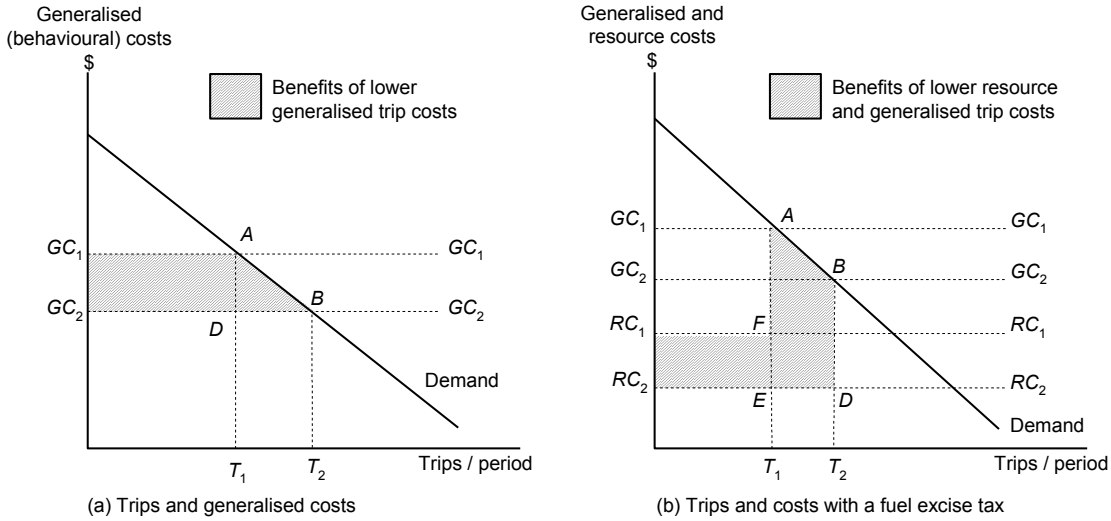


Figure 19.3 Valuation of road user benefits

panel (b) shows the net benefit of lower trip costs. The area equal to T_1 ($RC_1 - RC_2$) shows the savings in resource costs for existing trips.

Area *ABDE* is the benefit for generated trips. This equals the amount that trip makers would be willing to pay for the generated trips less the resource costs incurred. Including changes in producer surpluses, the benefit can be expressed as:

$$\Delta CS + \Delta PS = T_1 (RC_1 - RC_2) + \Delta T (GC_2 - RC_2) + \Delta T (GC_1 - GC_2)/2 \qquad (19.9)$$

where ΔT equals ($T_2 - T_1$).

When a road is part of a network, these benefits are estimated over the network or at least the relevant parts of the network. Ignoring externalities, the estimated present value of surpluses (summed over all relevant network links) would be compared with the present value of the capital and operating costs of the improved network. Introducing a toll just for the new road would not change RC_1 or RC_2 but it would raise GC_2 and so reduce generated trips and the benefits from the new road. In practice, economic evaluations are often based on estimated cost savings for a given set of trips across a trip network. The benefits then include only the first term on the right-hand side of Equation 19.9 and exclude any benefits from generated trips. On the other hand, externalities such as traffic noise and air pollution costs may also be included in the evaluation (see Box 19.2 and case study below).

Cost-benefit example

A cost–benefit analysis (CBA) of a 3.4 km road tunnel (the Lane Cove Tunnel) connecting two motorways in north-west Sydney exemplifies some of these issues.[5] The cost and benefits and overall results are summarised in Table 19.1. The evaluation was based on forecast trips over the regional road network with and without the proposed tunnel over 29 years following a five-and-a-half-year construction period. No residual value was included.

[5] This summary is based on *Lane Cove Tunnel and Associated Road Improvements*, Working Paper 14, *Economics*, prepared by Sinclair Knight Merz (2001) for the NSW Roads and Traffic Authority.

Box 19.2 Environmental values in transport studies

There is a strong case for factoring the costs of environmental impacts into (1) the use of transport infrastructure via pricing and (2) investment decisions via cost–benefit analysis. These impacts include air pollution, climate change due to the emission of greenhouse gases, noise, impacts on nature, soil and water deterioration, severance and visual intrusion. However, valuations of damage vary with study and circumstances. Two estimates of the costs per passenger car vehicle km in cities from air pollution, climate change and noise are cited below.

Austroads (2003), a major Australian reference, draws on two international studies. The estimates by Bickel *et al.* (2006) are averages of results found for seven European cities. These indicate lower but still substantial costs. Costs per vehicle km will vary with fuel consumption which depends especially on the type of car and the amount of stopping and starting. The estimated costs per heavy goods truck are about 20 times the costs per passenger vehicle.

Estimated average costs of environmental damages in cities: cents per vehicle km by passenger cars

	Austroads (2003)	*Bickel et al. (2006)*
Air pollution	2.37	0.50
Global warming	1.69	0.80
Noise	0.79	0.35

Table 19.1 Economic evaluation of Lane Cove Tunnel: summary of results ($m in 2001–02 prices)[a]

	2002	2003	2004	2005	2006	2007	2008	---	2035
Costs									
Capital	10	31	232	271	194	77	0		0
Recurrent	0	0	0	0	0	7	7		8
Total cost	10	31	232	271	194	84	7		8
Benefits: savings in									
Vehicle operating costs						67	69		111
Travel time						211	215		294
Vehicle accidents						8	8		9
Environmental costs						27	28		37
Total benefit						313	320		451
Net benefit	–10	–31	–232	–271	–194	229	313		443
Discount rates	4%	7%	10%						
NPV ($m)	463	2711	1646						
BCR	6.6	4.7	3.5						

(a) The full results included annual estimates for each year between 2008 and 2035.

Source: Sinclair Knight Merz (2001).

As is common in such studies, but not very satisfactory, forecast trips were a function of fixed land uses and did not allow for any generated trips. In effect, the CBA estimated only the fall in total trip costs for a given overall trip matrix. To model choice of road, a $2 toll was assumed. All costs and benefits were estimated in 2001–02 prices. Costs and benefits were discounted at a real discount rate of 7 per cent, with 4 and 10 per cent rates used in sensitivity tests. Capital costs include property take, project development and construction.

The benefits include savings in vehicle operating costs, travel time, vehicle accidents and environmental costs across the regional road network due to lower congestion. Savings in vehicle operating costs exclude tax components and were estimated separately for trucks and cars. Savings in travel time were estimated separately for people travelling privately or for work purposes and for commercial vehicles and in peak and off-peak hours. Savings in environmental costs were based on reductions in emissions of NO_X, particulates and CO_2 and

on reductions in noise impacts on households. Drawing on various studies, the report valued total savings in environmental costs at 4.2 cents per vehicle kilometre (vkm), which included air pollution costs at 1.4 cent per vkm, CO_2 emissions at 2.25 cents per vkm and noise costs at 0.55 cents per vkm.

Not surprisingly, given the strategic nature of this tunnel in the Sydney transport network, the study reported high estimated net present values and benefit–cost ratios. The positive results were not affected by alternative estimates for key variables such as construction costs, trip forecasts or values of travel time savings. The report did find that the proposed road toll, which would not apply to alternative routes, would reduce traffic using the tunnel by over 25 per cent, but it did not indicate how the toll affected the estimated net present value.

Evaluation of rail or bus services

The evaluation of rail or bus services is similar in many ways to evaluation of road services. Will user and third-party benefits justify the incremental capital and operating costs of the improved service? However, two special problems arise. One arises when the quality of the transport service improves in say comfort (reduced crowding) or frequency. There is often no market price measure of the value of this improvement. In this case, stated preference surveys (see Chapter 11) may be needed to determine what users would be willing to pay for this quality improvement.

Related evaluation issues arise with mode switching. Again, the appropriate valuation approach is to attempt to measure what mode switchers are willing to pay for the improved service. Evaluations often simply compare the costs of door-to-door trips by different modes on the assumption that the modes provide equivalent service levels. This tends to underestimate the benefits of private car travel, such as freedom to choose the time of travel,

Ideally user benefits of rail and bus services should be valued, like other services, by the amounts that people are willing to pay (WTP) for them. This avoids use of ad hoc premiums (or discounts) for service features such as frequency or comfort of service.[6] Suppose that WTP values for a rail and bus trip and marginal vehicle operating costs are as follows:

Mode	*WTP value ($)*	*Marginal cost ($)*
Rail	9.0	7.0
Bus	7.0	4.0

If passengers switch from rail to bus, there is a $3 saving in operating cost per passenger. If passengers value the services of the two modes equally, this represents the net benefit of the switch in mode. However, if passengers would be willing to pay $2 extra for rail, a switch to bus represents a net benefit of only $1. When a new mode is less attractive than an existing one, savings in resource costs inclusive of travel time overestimate the benefits of the new mode. Conversely, if the new mode is more attractive, savings in resource costs underestimate the benefits.

In general, where other modes set price (P) equal to marginal cost (MC), the net benefit of a new service is the difference between the amount that a user is willing to pay for it and the marginal resource cost (RC) of the service. If $P > MC$ in a substitute mode, the net benefit of a trip diverted to say rail from another mode is the difference between the benefit of the rail trip and the benefit of the trip on the substitute mode:

[6] However, in the absence of data on willingness-to-pay, WTP values may have to be derived from estimates of 'door-to-door' travel times and other trip features. In any case, trip forecasts may be modelled on 'door-to-door' travel times and other trip features.

$$NB_{DR} = (WTP - RC)_R - (WTP - RC)_0 \qquad (19.10)$$

where NB_{DR} is the net benefit of the trip diverted to rail and subscripts R and O represent rail and other modes respectively.

Figure 19.4 shows the benefits of trips due to an improved rail service. In panel (a) demand for rail trips increases from D_1 to D_2 and rail trips rise from Q_1 to Q_2. Panel (a) also shows the consumer surpluses of existing and generated rail users and the operating producer surplus (increase in net revenues) when the fare exceeds marginal operating costs. Note that panel (a) does not include any fixed costs associated with the improvement in services.

Panel (b) shows the shift in demand for bus trips as the new rail operates. The shaded area in panel (b) shows the surplus lost by bus operators as bus trips fall from B_1 to B_2. The net benefit is the surplus gained in the rail sector less the surplus lost in the bus sector.

Similar losses of producer surplus could occur with generated trips, when individuals transfer to rail from a non-transport activity. However, most non-transport sectors are assumed to have competitive markets with prices close to marginal cost. Thus, the benefits of generated trips are similar to those of diverted trips in panel (a). Deductions are rarely made for any loss of surplus in other sectors.

Box 19.3 shows possible costs and benefits associated with investment in rail services. Clearly the importance of the effects varies with the context. Freight traffic is critical to the Adelaide to Darwin rail service via Alice Springs, but of little importance between Sydney and Canberra. However, the benefit to freight, like that to passengers, is better estimated by what customers are willing to pay for the freight service itself rather than by ad hoc estimates of the values of travel time savings or other service attributes for freight traffic.

Evaluation of airport location

New airports are major economic investments, can transform the local region and create major third-party impacts. In 1971 the Commission of Inquiry into the Third London Airport produced a pioneering cost–benefit study of potential new airport sites for London.

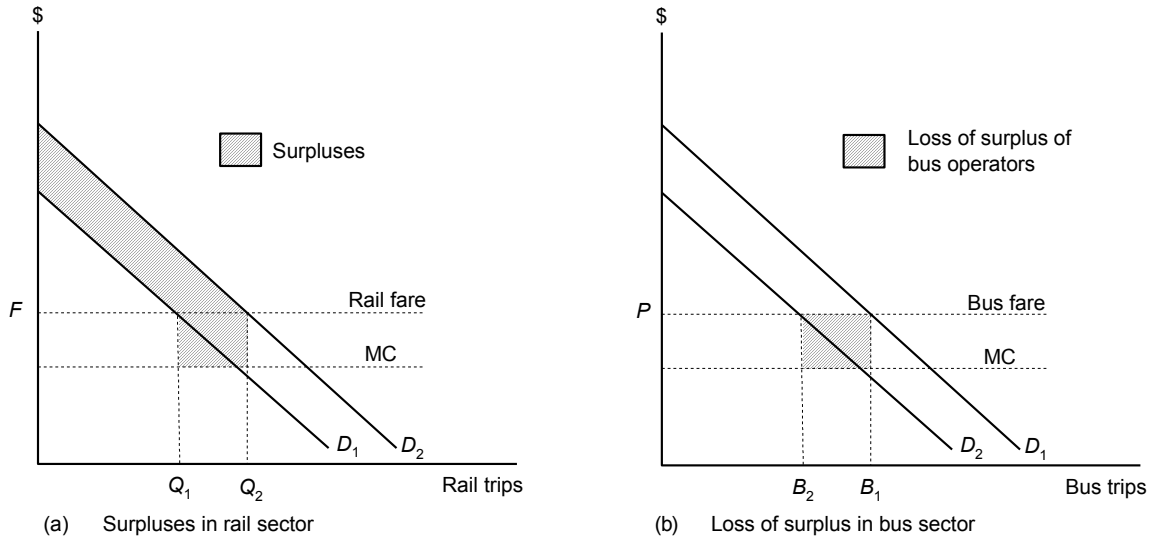


Figure 19.4 Benefits of improved rail service

Box 19.3 Possible costs and benefits of investment in new rail services

Possible costs	*Possible benefits*
Land and property costs	Passenger willingness to pay for services
Construction costs (rail infrastructure)	Freight user willingness to pay for services
Construction costs (rail stations)	Reductions in transport accidents
Costs of rolling stock	Reduced congestion on roads
Fixed rail operating costs	Net reduction in noise impacts
Variable rail operating costs	Net improvement in air quality
Rail maintenance costs	Net reduction in greenhouse gas emissions
Lower surpluses of other transport services	Savings from deferring other transport infrastructure
	Development benefits (very occasionally)

Actually, the Third London Airport study (TLAS) was essentially a cost-minimisation exercise. The study aimed to find a new four-runway international airport close to London that would, with the other two major airports (Heathrow and Gatwick), minimise the costs of the proposed three-airport system for the forecast air traffic (passengers and freight). The costs included property take, site development, airport construction and operating costs, aviation flight costs, the costs of relocating defence and civil airports, costs of providing access to the airports, costs of access for air passengers, noise costs and costs of urbanising areas around the new airport. The study departed from cost-minimisation only in so far as one of the three-airport options provided superior access to air travellers and was forecast to generate more air trips, which provided a small offset against costs of that option.

Despite its technical quality, the TLAS was a practical failure. None of the four sites on its recommended short list has been developed as an airport. Instead, Heathrow and Gatwick airports along with Stansted (then a small airport) have expanded into a three-airport system.

In 1972, the NSW government commissioned a similar cost–benefit study for a second Sydney airport. The aim was to find the best site for an international two-runway airport with long wide-spaced parallel runways for international aircraft and a cross runway and Badgery's Creek was recommended. Over the next 40 years, several more studies, but no new airport. Only in the last few years has the Australian government announced that it will finally develop a second Sydney Airport. Where? At Badgery's Creek!

Political factors play a large part in decisions about airports which can affect thousands of households. These households cannot be easily compensated for an adverse airport location. However, there are also instructive economic issues. These relate to the nature of the problem to be solved, the pricing of airport services and the timing of new infrastructure.

Both London and all Sydney airport studies suffered from two related flaws. They failed to identify the infrastructure problem, namely a lack of capacity in peak hours. In Sydney, there was excess demand for runway space in peak hours because a quarter of the peak traffic were slow-moving, small aircraft providing local services, which create congestion and slow down faster and larger aircraft. Only a fifth of the flights were international. Sydney then required extra runway capacity to service these small aircraft. It did not need a new two-runway international airport. In the 1970s, London likewise needed some extra runway capacity, but not a new four-runway airport.

Related to this, the TLAS and most Sydney studies failed to consider pricing and timing issues. Peak-period pricing sorts out the aircraft that need to land or take off at a central airport in peak hours from those that can move to other airports or hours. Until there is considerable excess demand, the amounts that airlines are willing to pay for extra peak hour

services are usually well below the costs associated with a new airport. Incremental investment is justified when there is sufficient excess demand *given efficient peak pricing*.

The optimal timing for infrastructure occurs when the service benefits exceed the benefits of deferment. A project that is viable in the long run should not necessarily be started today. Suppose that K is the capital cost of investment and r is the annual rate of return on capital. The benefit from deferring an investment for a year is rK. Investment should be deferred until $b>rK$, where b is the benefit of the infrastructure service in that year. Figure 19.5 illustrates the principle. In this case, the project should open in year 10.

Wider economic benefits

In recent years, there has been extensive literature on a suite of impacts not included in the standard traditional economic appraisal of transport projects, which have the generic title of wider economic benefits (WEBs). These fall into six main categories:

1. Agglomeration economies: the impacts of increased employment density on productivity
2. The impacts of transport infrastructure on labour supply
3. The value of business travel time savings in imperfectly competitive markets
4. The impact of transport infrastructure on competition
5. Benefits arising from dependent (or induced) development, and
6. Possible further economic impacts of investment in transport infrastructure.

The first four WEBs were identified in the influential paper on WEBs published by the UK Department for Transport (2005) and have been the focus of most analysis since then. WEBs 5 and 6 above were discussed as possible additions to the standard transport appraisal in the 1980's and 1990's and have re-surfaced. We discuss these WEBs briefly in turn.

Agglomeration economies occur if a transport project increases employment densities and this raises productivity (and income) per worker. This may happen, though transport projects can also disperse employment. Contentiously, it is sometimes argued that lower transport costs increase "effective density" without actual changes in employment density and that this also raises productivity. However, this form of agglomeration economy is implausible and not strongly supported by evidence.

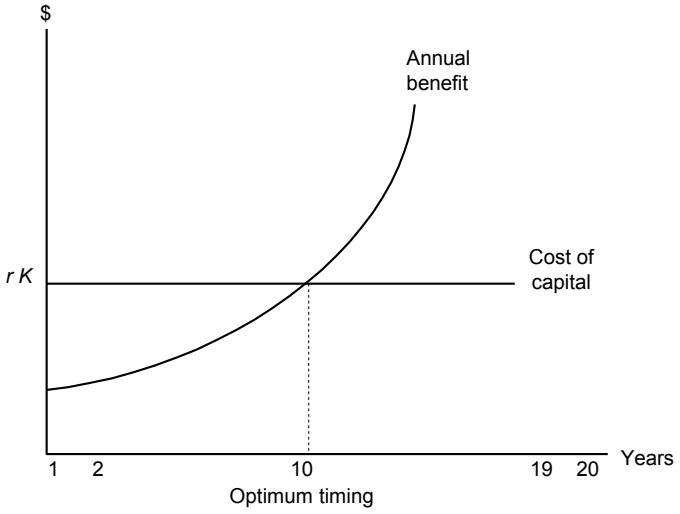


Figure 19.5 Optimal timing of infrastructure

Labour supply. Transport projects may encourage labour supply by reducing commuting costs. Technically, the labour benefit is equivalent to the consumer surplus of new trips shown in Figure 19.3 and Equation 19.9. However, there may be some additional tax revenue that is not picked up in Equation 19.8.

Value of output. In imperfectly competitive markets, goods are sold with a mark-up over factor cost. To allow for this, the value of business time saved may be marked up by 15 or 20 per cent over the wage rate.

Increased competition. UK DfT (2005) suggested that transport projects could increase the competitiveness of cities linked by the new infrastructure. However, this is regarded as implausible where transport networks are already well developed.

Induced developments. If new transport infrastructure induces non-marginal development, such as a resource project in a rural area or major new housing in an urban area, this may create producer surplus that is not captured in travel benefits. This surplus is the market revenue from the new project less the sum of all related public and private costs.

Impacts on the economy. All expenditure has an opportunity cost. It is generally concluded that investment in transport infrastructure does not have special additional effects on employment, productivity and output that are not captured by the standard method of evaluation.

However, these are strongly contended positions, in part because they have major impacts on whether transport projects pass standard cost-benefit criteria. Suggested extra references are Dobes and Leung (2015), Douglas and O'Keefe (2016) and UK DfT (2016a, 2016b and 2016c).

Role and Regulation of the Private Sector in Transport

Because transport infrastructure is often a natural monopoly, government has traditionally financed, owned and operated most such infrastructure. Nevertheless, private firms may be able to supply and operate some of the infrastructure more efficiently than the public sector.

There are many examples of private involvement in transport infrastructure and services and these examples take many forms. In its fullest form, government cedes to the private sector the right to finance, build, own and operate the infrastructure. For example, a private company (Eurotunnel) owns and can operate the Channel Tunnel without a time limit. The UK Government has also privatised most rail services in the UK while maintaining ownership of the rail infrastructure. In some cases, for example French autoroute operations, the operation of the transport infrastructure is a joint public–private venture.

In Australia, several major airports and most shipping ports are privately owned. In Sydney, several bus and ferry services are run privately under contract with the state government. Likewise, most intra-urban and coastal ferry services in Queensland are run by private companies under contract with the Queensland government. And there are many privately-run toll roads in Melbourne, Sydney and Brisbane.

We discuss below some of the generic issues in public-private partnerships and some regulation issues of private firms in the transport sector.

Roles of public and private sectors

As discussed in Chapter 18, the effectiveness of public–private partnerships (PPPs) depends on the circumstances, the structure of the arrangements and the nature of any regulations. In principle, PPPs should facilitate explicit objectives, clear and detailed contracts and competition in supply and use of infrastructure. However, poor management of a complex process can have poor outcomes, especially when the private agency is a *de facto* monopoly.

Theory and evidence suggest that productivity is likely to rise under PPPs. Private firms have more incentive than public agencies to complete construction faster (as revenues start earlier) and to a higher quality (because this affects patronage). In Sydney, the Harbour Tunnel and major motorways (M2 and M5) were constructed much faster under PPPs than was expected. There have also been savings in the order of 20 per cent in construction and operating costs of major roads in Australia, the UK and France (Bureau of Transport and Communications Economics, 1996; Quinet and Vickerman, 2004). While some savings may have been achieved by lower wages or inferior working conditions in the private sector, almost certainly there were savings from enhanced productivity.

Two related issues are the allocation of risk and the cost of finance. Here the arguments for, or against, PPPs are less clear, but they may influence the preferred structure of a PPP. Risks may be political, technical or commercial. Political risk is the risk that government will change regulations or introduce competition to the infrastructure element under concession. Technical risk relates to construction and operating costs. Commercial risk is market risk.

Government is the only party that can deal with political risk. It can also spread risks over millions of taxpaying households. On the other hand, the private sector is generally more skilled at dealing with technical and commercial risk than government agencies. Also, markets can spread risk to parties that are willing to take it on, sometimes because the new transaction provides diversification and even reduces portfolio risk. In a well-structured PPP, the party that can best deal with the risk should bear it.

One way that markets deal with risk is to charge a higher price of capital. Private financing costs are often several percentage points higher than government's borrowing rate. The actual differential depends on the extent to which government underwrites an infrastructure project with support of various kinds. However, the real cost of public borrowing depends on the extent to which repayments are funded from taxes. If funded by taxes the deadweight loss is often at least 20 per cent of revenue raised. This deadweight loss may offset partly or wholly any differential in the nominal cost of capital.

The main potential cost of a PPP is the concession of a monopoly service to private interests with poor contract design and regulations that fail to safeguard the public interest. Government may concede to a private firm the right to own and operate an infrastructure asset before it finalises the contract. In other cases, as with the East-West cross-city Sydney tunnel, to maximise the financial return to government, government agreed to close some existing roads and thereby protect the private firm from route competition. This reduced the economic benefits from the investment.

Critically, contracts are often renegotiated without the benefit of competitive pressures. In a study for the World Bank, Guasch (2000) found that, of over 1000 concessions awarded in Latin America and the Caribbean, over 60 per cent were renegotiated within three years, nearly always at the initiative of the concession holder. Eighty per cent were in transport or water sectors. In many cases contracts were renegotiated not because of changes in economic conditions but because the concession holder had submitted a low bid to gain the concession in the expectation of renegotiation. This means that the most efficient firm may not have won the contract.

Managing transport infrastructure monopolies

The main public policy issues arising in management of privately owned infrastructure are vertical integration, conditions of access, pricing and quality of service, such as frequency. The issue of vertical integration relates to whether the owner should also provide the services that use the infrastructure. This issue arises most acutely for rail services and pipelines. Issues of access and pricing are related as it is generally agreed that service providers should have access to essential infrastructure, but the sticking point is the price.

Vertical integration and access. There are three main options for the relationship between the owner and users of the infrastructure.

1. Full vertical integration of infrastructure and services provided with one owner-operator.
2. Partial integration with both infrastructure owner and other operators using the services.
3. Structural separation where all users of the infrastructure are independent of the owner.

Vertical integration (option 1) may be the most efficient industry structure when there are economies of scope as well as of scale. Economies of scope exist when a single agency can supply the infrastructure and the services using the infrastructure at lower unit cost than can separate agencies. Some rail services may be an example of this. However, it is difficult to prove this without market testing. To ensure that services are provided efficiently it is generally desirable to allow competition among potential users and to ensure as far as possible that they compete on equal terms. Thus option (2) is usually preferred to option (1).[7]

However, if the infrastructure owner competes against outside service suppliers, the competition may be unfair and hence inefficient. Accordingly, a regulator is required to monitor that the terms of access are efficient (and not contain monopoly profit to the owner) and that all competitors have access to the network on equal terms to each other and the owner. However, the regulator will have to rely on information on costs and contracts from the regulated infrastructure operator.

Complete structural separation (option 3) sidesteps an owner's incentive to deal unequally with agencies tendering for use of the infrastructure. These considerations have led several European countries and the state of Victoria in Australia to break up vertically integrated rail systems and introduce competitive bidding for the right to run rail services over the network.

Price regulation. Given the natural monopoly character of most of these private operations, the charges are generally regulated.[8]

Whichever industry structure is adopted, there are efficiency and equity arguments for regulation of charges for use of the infrastructure. In off-peak periods, when there is underutilisation of capacity, user charges should reflect the marginal costs of use. When infrastructure use is at or close to capacity, the owner should be generally be encouraged to charge market-clearing prices which ensure that those who place the highest value on use at that time gain access. Overall, charges may have to be modified to allow for a fair return on equity capital. Also, fairness considerations are likely to influence the regulated prices.

As we saw in Chapter 14, the case for regulating infrastructure user charges or other service provision is based on the presumption that the regulator (either government or an independent regulator) will determine a more efficient and fairer set of charges than would the monopoly owner. Detailed fare and service regulation is highly complex. Also, regulation may stifle service innovation and deter investment. Thus, there are potential costs as well as benefits from regulation of monopoly-owned infrastructure.

In principle, road tolls are straightforward. However, there are significant equity implications because of the wide range of road users. For rail, the major policy issue is the amount of access to allow private operators. Once access is agreed, the allocation of access may be determined by an auction process. Rail operators may in turn be subject to separate price and quality controls. However, few governments regulate prices for access to airport runways, the critical monopoly feature of airports.

---

[7] Regulation may involve service regulation. For example, Eurotunnel is required to provide a minimum level of service at all times of the day and year.

[8] In Queensland, the government reimburses some ferry operators for agreed operating costs and takes the revenue risk rather than the operator taking it.

Summary

- In this chapter, the transport sector exemplifies important general issues for economic infrastructure, much of which is supplied by natural monopolies.
- The key issues are (1) how to use existing transport infrastructure efficiently and equitably and (2) how to achieve an efficient level of transport infrastructure.
- Efficient prices create efficient use. Use of infrastructure is efficient when the marginal social benefit of use equals the marginal social cost. However, equity may also be important.
- When infrastructure is under-utilised, short-run marginal cost pricing encourages efficient use. However, cost recovery may be important for efficient management and competitive neutrality. When infrastructure is heavily used, market-clearing prices ensure that those who most highly value access can gain it.
- Investment in infrastructure capacity is efficient when the estimated present value of the benefits of increased capacity exceed the present value of the costs, inclusive of third-party benefits and costs.
- Optimal capacity is achieved where the marginal benefit of extra capacity just equals the marginal cost.
- To achieve optimal timing, new capacity should be available when the net benefit from use over a period such as a year exceeds the opportunity cost of capital employed.
- Some recent literature contends that transport infrastructure produces wider economic benefits that are not captured in standard appraisals, but this is currently an open question.
- Government traditionally financed, owned and operated most transport infrastructure. However, competitive private sector involvement can often provide services more cost-effectively. These benefits may not be achieved by a private monopoly supplier.
- Two main issues arise with privately owned infrastructure. One is whether the owner should also provide services that use the infrastructure. Where this is allowed, private operators should generally be allowed access to the infrastructure on a competitively neutral basis.
- Second, there are generally efficiency and equity reasons for regulating both the services provided and the prices charged by private monopolies. However, detailed regulation is complex and may stifle service innovation and deter investment. Public policy must determine where the balance lies.

Questions

1. What are the basic principles for efficient use of transport infrastructure?
2. What are the general principles for efficient investment in transport infrastructure?
3. Why is efficient pricing of services a key link between efficient use and efficient investment in transport infrastructure?
4. Would these efficient pricing and investment principles apply equally to the other major infrastructure networks (water, power and telecommunications)? If not, why not?
5. How might equity or social factors influence the use of transport infrastructure?
6. How might equity or social factors influence investment in transport infrastructure? Give examples.
7. How would you determine the role of the private sector in the ownership and operation of shipping, ports and airports?
8. The text distinguishes between political and market risks for PPPs and suggests that government should be responsible for the former and private firms for the latter. Is this distinction practical?
9. Suppose that the construction cost of a new road is $100 million and that the opportunity cost of capital is 7 per cent. Construction takes one year and the road has a life of 30 years. The estimated benefits of the road would be $5 million in year 2 and would rise by 4 per cent per annum.
   i. What would be the net present value of the road if it was constructed today?
   ii. What would be the optimal year in which to construct the road?
   iii. If constructed then, what would be the net present value of the road?
10. Is it possible to account efficiently for environmental factors in transport pricing and investment decisions?

Further Reading

Clark, H. and D, Prentice (2010) 'Reforming taxes and charges on Australia road transport' pp. 389–414 in C. Evans, R. Krever and P. Mellor (eds), *Australia's Future Tax System: The Prospects after Henry*, Thomson Reuters, Sydney.

Douglas, N. and B. O'Keeffe (2016), 'Wider Economic Benefits – When and if they should be used in evaluation of transport projects', *Australian Transport Research Forum Proceedings*.

Gomez-Ibanez, J.A. (1999) 'Pricing', pp. 99–135 in J. Gomez-Ibanez, W.B. Tye and C. Winston (eds), *Essays in Transportation Economics and Policy, A Handbook in Honour of John R. Meyer*, Brookings Institution Press, Washington DC.

Guasch, J.L. (2000) *Concessions: Bust or Boom? An Empirical Analysis of Ten Years of Experience in Concessions in Latin America and Caribbean*, World Bank Institute, Washington DC.

Hensher, D.A. (2012), 'Transport Economics: An Overview', pp. 1-48 in D. Hensher (ed.) *Major Works in Transport Economics*, Vol.1, Routledge, UK.

Hensher, D.A. (2011) 'Value of Travel Time Savings', pp. 135–158 in A. de Palma, R. Lindsey, E. Quinet and R. Vickerman (eds), *Handbook in Transport Economics*, Edward Elgar, London.

Small, K.A. (1999) 'Project Evaluation', pp. 136–177 in J. Gomez-Ibanez, W.B. Tye and C. Winston (eds), *Essays in Transportation Economics and Policy, A Handbook in Honour of John R. Meyer*, Brookings Institution Press, Washington DC.

Newbery, D.M. (1995) 'Pricing and congestion: economic principles relevant to pricing roads', pp. 345–372 in T.H. Oum et al. (eds), *Transport Economics: Selected Readings*, Korea Research Foundation for the 21st Century, Seoul.

UK Department for Transport (2016), *Wider Economic Benefits Appraisal*, TAG Unit A2.1.

Part

7

Social Welfare and Income Re-distribution

Chapter

20

Poverty and Inequality

Annual income twenty pounds, annual expenditure nineteen nineteen six, result happiness.
Annual income twenty pounds, annual expenditure twenty pounds ought and sixpence, result misery.

Charles Dickens, *David Copperfield*

Measurement of Income ♦ Distribution of Income and Wealth in Australia ♦ Measures of Poverty ♦ Measures of Inequality ♦ Trends in Poverty and Inequality ♦ Causes of Unequal Incomes

Reducing poverty and inequality in society are fundamental functions of government.[1] However, to develop appropriate welfare policies, we need first to establish the meaning, measures and causes of poverty and inequality. Therefore, in this chapter we start by discussing basic income concepts and describing the distribution of income and wealth in Australia. The next two sections discuss measures of poverty and inequality. The last two sections describe trends in poverty and inequality and their causes.

It should be noted that we focus in this chapter on income measures of poverty and inequality. Lack of income is the prime indicator of poverty and inequality of income is a major welfare and policy concern. There is also an extensive literature on both deprivation measures of poverty including poor health and vulnerability (Saunders and Naidoo, 2009) and on poverty as a lack of capabilities and freedoms (Sen, 1997). These issues are fundamentally important. We pick up on some of these related social issues in the following chapters.

Measurement of Income

Income is a common word but a complex concept. Issues include the components of private income, the period over which income is measured, the unit of observation, and the effects of government benefits and taxes.

Economists usually adopt the comprehensive Haig–Simons definition of personal income: this is the value of market goods that an individual could consume in any period without any change in their wealth. However, this definition does not include non-market goods, such as health or leisure, that may also contribute greatly to individual welfare.

The three basic sources of **private income** are labour, capital and gifts (see Box 20.1). The most important components of labour income are wages and salaries, employer supplements (mainly pensions and, to lesser extent, fringe benefits) and income from own business. These

[1] This statement reflects mainstream European and Australian cultures. It appears less supported in some other countries including in the United States.

Box 20.1 Sources of private income

Income from labour	*Income from capital*
Wages, salaries and cash supplements	Property rents net of costs
Own business enterprises, self-employment, partnerships, etc.	Interest and dividends
	Private pensions
Other employment supplements, pension contributions and fringe benefits	Real capital gains and losses from investments
	Imputed rents from physical capital
The underground (cash) economy	*Other private income*
Household production of services	Bequests, gifts, alimony, etc.

components usually make up about two-thirds of total income as shown by the national income accounts (ABS, Cat. No. 5206.0). Other labour income (usually unmeasured) includes income from the cash economy and household production of goods.

Income from capital includes some easily measurable elements, such as rental property income, interest income and dividends, and private pensions. Greater measurement difficulties arise with capital gains and imputed rents. Capital gains, whether realised by sale of assets or not, are part of the Haig–Simons definition of income because they contribute to changes in net worth and could be spent on consumption with no decline in net worth. Such gains should be measured in real terms, that is, after allowing for changes in prices. For example, if the value of an individual's assets rises from \$100 000 to \$106 000 and the inflation rate over this period is 4 per cent, the real increase in the value of the assets is \$1923.

$$\$106\,000/1.04 - \$100\,000 = \$1923 \qquad (20.1)$$

On the other hand, losses in asset values should be deducted from income. In practice, most estimates of income exclude changes in asset values.

Imputed rent is the market value of services from consumer durables, such as houses, cars, household contents and so on. The imputed rent for housing services is the net rent (gross rent less expenses) that a homeowner would pay for a comparable dwelling unit. These net rents are included in national income statistics. In Australia, total income from dwellings owned by persons, rented and owner-occupied, accounts for about 8 per cent of national income. Half of all households in the lowest income quintile in Australia (mainly elderly households) own their housing outright and obtain significant imputed rent.

Thus, the economist's concept of income extends well beyond the narrow definition of income in the national accounts. However, most estimates of personal income fall well short of the comprehensive Haig–Simons definition. For example, imputed rents, cash payments and gifts are difficult to track and rarely included in estimates of personal income.

Income may be measured over any period. A common period is a year, but this is an artificial convenience. For interpersonal comparisons, there are strong arguments for using lifetime rather than annual income. Incomes often start low, rise, and then decline with age. People may have similar incomes over their lifetimes but very unequal incomes at any point in time. Hence inequality estimates based on annual income are generally higher than those based on lifetime incomes. However, lifetime incomes are difficult to measure. Also, a person with little income at any point in time may need assistance now because he or she cannot borrow against future prospects. For many needs, the relevant period is the short run.

The choice of income unit. This is another important issue. To measure needs, the household unit is generally preferred to the individual because resources are usually shared within households. Individual consumption usually depends more on household income than on

personal income. Suppose that two adults in a household earn $100 000 and $30 000 per annum respectively. Individually, one would be well off and the other poor. But a two-adult household with an income of $130 000 would usually be regarded as well off (see Table 20.1).

To make comparisons of income between households of different size and composition we need to define an equivalised household unit (EHI). In 2009-10, of the then 8.4 million households in Australia, 26 per cent were couples with dependent children, 26 per cent were couples only, 25 per cent were lone persons and 23 per cent were other households. While larger families have greater expenditure needs, economies of scale mean that, as household size increases, the cost of maintaining a given standard of living per person falls. Larger households and households with children need less income per capita to achieve given consumption standards. The aim is therefore to estimate the expenditure that households of different size and composition will need to make to obtain a given level of welfare or its material equivalent (a given set of goods).

To estimate EHI, the Australian Bureau of Statistics (ABS) adopts the OECD scale. This allows 1.0 point for the first adult, 0.5 points for each additional person aged 15 and over, and 0.3 points for each child under 15. EHI is derived by dividing total household income by the sum of the equivalence points allocated to the household members. Thus, EHI is an indicator of capacity to consume, not a measure of actual income.

Including government benefits and taxes. Government benefits and taxes substantially affect the final distribution of income. Household income is reduced by personal income taxes and by indirect taxes that are passed on in higher prices to households. On the other hand, household income is increased by cash benefits such as age pensions and family payments. Some other forms of government expenditure provide an indirect benefit to households, for example expenditure on health and education benefits certain groups of households.

Figure 20.1 shows the main relationships between various measures of income. Gross income is the sum of private income and government cash benefits. Disposable income is gross income less direct tax. Final income includes indirect benefits, for example benefits due to public health expenditures, and excludes indirect tax.

Distribution of Income and Wealth in Australia

Table 20.1 shows gross household income per week of Australian households in 2015-16 and equivalised disposable household income on a weekly basis in 2013-14 and 2014-15.[2] Gross income is regular cash receipts before tax. This includes wages and salaries, earnings from own businesses, property income, interest and dividend payments, private pensions and government cash payments (including allowances, benefits and pensions). Thus, gross income is broader than private market income. Disposable income is gross income less income tax and Medicare levy.[3]

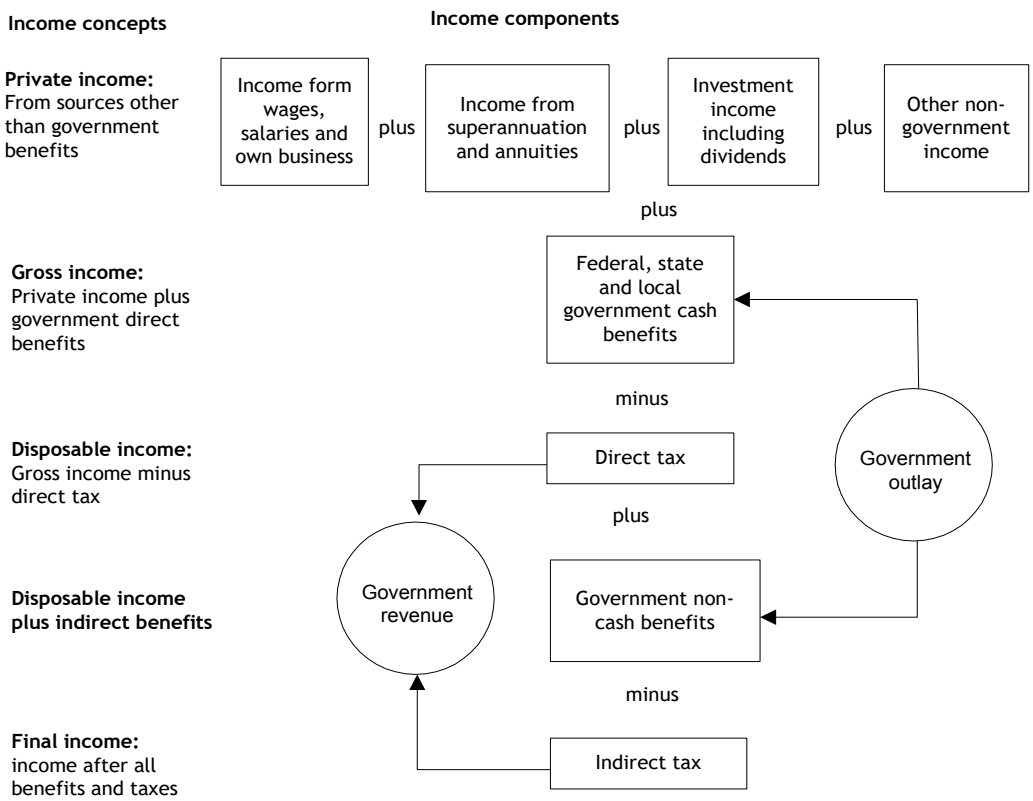
As shown in Table 20.1, in 2015-16, mean gross household income was $2109 per week. Gross income at the 90-percentile level was 8.8 times income at the 10-percentile level. However, the 90/10 ratio for equivalised disposable household income was 3.9. This reflected the impacts of household size and taxation on household incomes.

We should also note that income here does not include imputed rents. The ABS (2011) estimated that the average value of imputed rents was $57 per week per household and would add 7 per cent to mean equivalised income. Because many low-income elderly households own their houses, inclusion of imputed rents reduced the 2007–08 P90/P10 equivalised disposable household income ratio from 4.29 to 3.76.

---

[2] ABS 6523.0 does not provide a similar quintile breakdown of gross household income.

[3] These results are based on samples of around 20,000 households.



Source: ABS (2005), *Household Expenditure Survey, Australia: Summary of Results, 2003-04*, Cat. No. 6530.0

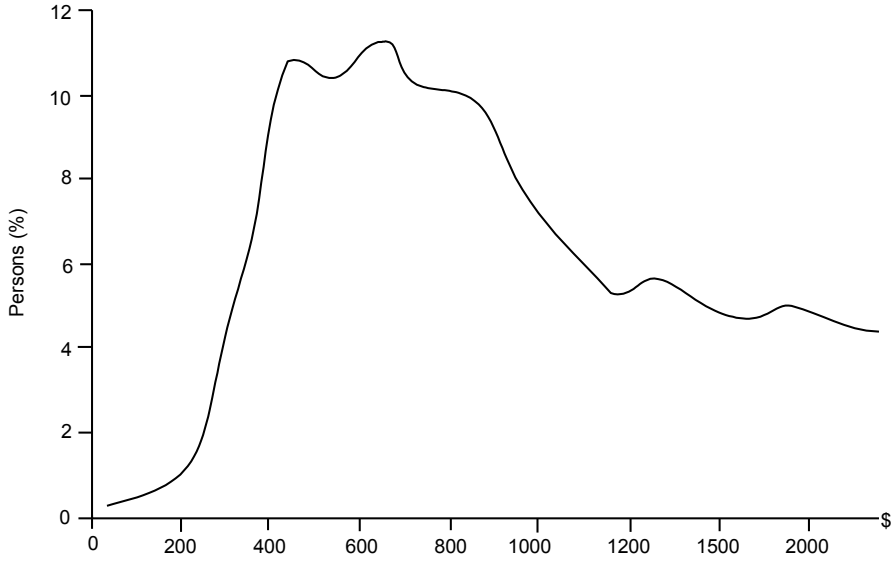
Figure 20.1 Income concepts and components

Table 20.1 Mean household income per week

Percentile boundaries and ratios	*Unit*	*Gross household income 2015-16*	*Equivalised disposable household income 2013-14*	*Equivalised disposable household income 2015-16*
P10	$	471	415	436
P20	$	701	511	523
P50	$	1,616	844	853
P80	$	3,097	1,308	1,371
P90	$	4,171	1,688	1,705
P90/P10	Ratio	8.88	4.07	3.91
P80/P20	Ratio	4.42	2.56	2.62
Means				
All households	$	2,109	998	1,009
Couple with children	$	3,083	1,011	1,033
One parent with children	$	1,427	687	705
Couple only	$	1,986	1,113	1,097
Lone person under 44[a]	$	1,135	1,009	1,107
Lone person 65 and over	$	661	607	629

Source: ABS (2017) *Household Income and Income Distribution, Australia*, 2015-16 Cat. No. 6523.0 and ABS (2015) *ibid.*, 2013–14.

a) Gross household income is under 35, equivalised disposable household income is 25-44



Source: ABS (2009), *Household Income and Income Distribution, Australia, 2007-08,* Cat. No. 6523.0.

Figure 20.2 Distribution of equivalised disposable household incomes per week in Australia in 2007–08

Figure 20.2 shows an estimated full distribution of equivalised disposable household incomes per week in 2007–08. This illustrates the distribution of resources available to households after allowing for differences in household composition.

Wealth is distributed more unequally than income. ABS (2017) estimated that, in 2015-16, the wealthiest 20 per cent of households in Australia held 62 per cent of total household net wealth, with an average net worth of $2.9 million. On the other hand, the 20 per cent of households with least wealth held only 0.8 per cent of total household net wealth, with an average net worth of $39 000.

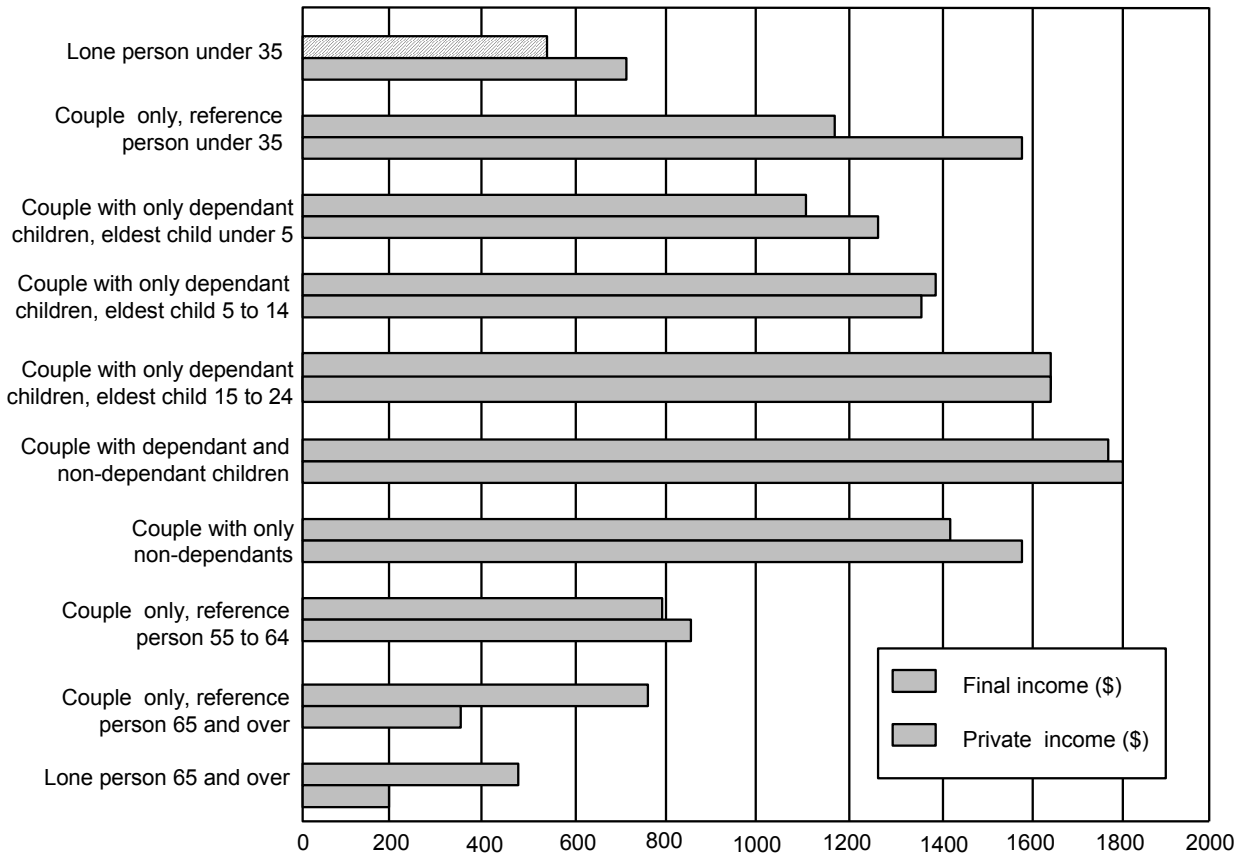
However, wealth is correlated with age as much as income. Because about half of all wealth is held in housing and another quarter in pension funds, many low-income households have significant wealth. Table 20.2 shows mean household net worth in relation to equivalised household disposable income. Households in the lowest EHI quintile held 75% of the average EHI net wealth. On the other hand, households in the highest EHI quintile held 43% more household wealth than the average EHI household.

The ABS (2007) has also estimated how government benefits and taxes affected the distribution of income in 2003-04. In this study, the ABS allocated 60 per cent of taxes on production and slightly more than 60 per cent of expenditure to households.

Table 20.2 The distribution of household income and wealth in 2015-16 ($)

	Equivalised disposable household income					
	Lowest	*Second*	*Third*	*Fourth*	*Highest*	*All*
Mean disposable equivalised household income	694	898	909	1,018	1,446	1,009
Median disposable equivalised household income	601	803	808	930	1,223	853
Mean household wealth	601,000	803,000	808,000	930,000	1,223,000	853,000

Source: ABS (2017) *Household Income and Income Distribution, Australia,* 2015–16 Cat. No. 6523.0



Source: ABS (2005), *Household Expenditure Survey, Australia: Summary of Results, 2003-04*, Cat. No. 6530.0

Figure 20.3 Life cycle groups: average weekly private and fixed income 2003–04

Table 26.3 (page 467) shows the effects by income quintiles for equivalised households. Government interventions greatly increased the shares of the three lowest quintiles and reduced the shares of the top two quintiles.[4]

ABS (2007) also showed the impacts of government benefits and taxes on the incomes of 10 household units in 2003–04. The mean results are shown in Figure 20.3. The largest recipients were singles and couples over 65. The major net taxpayers were singles and couples under 35 with no children.

Measures of Poverty

Poverty may be defined in absolute or relative terms. An **absolute definition of poverty** implies a fixed standard that does not change over time, which is independent of the general standard of living of the community. Such a definition of poverty typically focuses on the basic necessities for life and health. Rowntree's (1901) pioneering study of poverty in the UK based the poverty line (the income below which people were in poverty) on basic allowances for food, clothing, housing and heating. To make international comparisons, the World Bank

Absolute poverty standard
A fixed and unchanging level of poverty

[4] The ABS estimates, like most such studies, take observed initial market incomes as given. However, fiscal arrangements affect behaviour and factor incomes and so influence initial market incomes. Some general equilibrium models allow for this (Cordes and Watson, 1998).

has adopted an international poverty line of US$1 a day per person in 1985 prices. According to this measure, in 2000, 49 per cent of people in sub-Saharan Africa, 32 per cent in South Asia and 15 per cent in East Asia and the Pacific were poor (World Bank, 2006). However, this is a measure of extreme poverty and higher dollar values are sometimes adopted.

Relative poverty
Poverty is defined relative to the standard of the relevant society

A **relative definition of poverty** is based on the prevailing standards of society. As Adam Smith remarked, by the necessities of life 'I understand not only the commodities which are indispensably necessary for the support of life but whatever the custom of the country renders is indecent for creditable people, even of the lowest order, to be without'.[5] James Wolfensohn, a recent head of the World Bank, reportedly remarked that 'anyone in the world who lacks access to the Internet is poor'. A relative measure of poverty is reasonable in that, as a community's income rises, its capacity to raise the income of the poor increases. On the other hand, a relative definition of poverty presents a moving target. Some people who were rich yesterday are poor today.

The related concepts of needs and poverty have been defined and estimated in various ways (Piachaud, 1993). A common method is the budget standards (basket expenditure) method. Under this method, the analyst estimates the income required to purchase a basket of goods deemed necessary for a minimally adequate standard of living for households of different sizes. If the poverty line is regarded as an absolute standard, the poverty line should rise with a price index for goods purchased by very low-income households. More often it is viewed as a relative standard and rises with increases in the local standard of living.

To get round the the difficulty of establishing such baskets of goods, poverty is often defined simply in terms of relative income. OECD studies of poverty, such as OECD (2008), typically deem as poor anyone in an equivalised household with a disposable income of less than half of the median equivalised household disposable income (EHDI). In this case the poverty line simply moves over time with median EHDI. This means that if the distribution of income does not change, the proportion of persons in the population who are deemed to be poor does not change. *With a relative standard, poverty reduction can occur only with a more even distribution of income*.

There are two main societal measures of poverty. One is the headcount ratio. This is simply the proportion of the population below the designated poverty line. The second is the size of the poverty gap—the amount by which low incomes fall below the poverty line. One measure of the poverty gap is the sum of all individual poverty gaps as a proportion of total income. This indicates the amount of redistribution necessary to eliminate extreme poverty.

Household poverty lines in Australia

In his often-cited Harvester wage judgement (1907), Justice Higgins drew on a small sample of household budgets to determine that an amount of seven shillings per day was 'appropriate to the normal needs of the average employee regarded as a human being living in a civilised society'. Likewise, the Royal Commission on the Basic Wage chaired by Justice Piddington (1920) tied the wage to 'the cost of living according to reasonable standards of comfort for the typical family'.

In more recent times the main measure of poverty in Australia has been the Henderson Poverty Line (HPL) based on the Henderson Commission of Inquiry into Poverty (1975). The inquiry set the poverty line for a reference two-adult two-child household (with rental housing) at an after-tax income equal to the value of the basic wage plus child endowment (as it was then). This was equal to 56.5 per cent of seasonally adjusted average weekly earnings. Henderson observed that, as a proportion of average earnings, the HPL was similar to the poverty lines used in several other countries. Henderson described this poverty line as 'an

[5] Smith, A., 1776, p. 691.

austere low level' and considered that anyone below it would be 'very poor'. Those with an income less than 20 per cent above it would be 'rather poor'.

The Henderson Inquiry also applied equivalence factors to nine other household units to allow for variations in needs. To develop household equivalence scales, Henderson drew on a 1954 survey of household budgets in New York. Extra costs were allowed for households where the head was working to reflect commuting costs and loss of household output. Henderson also estimated lower poverty lines for home owners who do not pay housing rents. The Institute of Applied Economic and Social Research (Melbourne University) updates the poverty lines in line with movements in seasonally adjusted household disposable income per capita. Table 28.3 shows the poverty lines in the March quarter 2011.

The Henderson poverty lines have drawn criticism. This is inevitable in that they are based on value judgements. One criticism is the false appearance of precision. A few dollars above or below the HPL does not represent a real difference between poverty and non-poverty (as Henderson well knew). A more cogent criticism is the reliance on expenditure patterns in New York in 1954 to indicate equivalent family units in Australia many years later.

Partly in response to concern about the stringent nature of the HPL, the Social Policy Research Centre (University of New South Wales) estimated low-cost budget standards for a range of Australian households, with a low-cost budget standard defined as 'one designed to meet basic needs at a frugal level while still allowing social and economic participation consistent with community expectations'. The study found that most low-income households require an income that is 10 to 20 per cent higher than the equivalent HPL budget. This reflected the budget standards adopted by the study which related, as the authors recognised, to 'low-income' households rather than to a poverty standard (Saunders, 1999).

As a practical matter, the most common benchmark of poverty in Australia and elsewhere today is EDHI of less than half the median EHDI. This is a crude measure in that it generally does not account for home ownership or journey to work expenses.

Table 20.3 Henderson poverty lines in September quarter 2017 ($ per week)[a, b]

	Including housing costs		*Excluding housing costs*	
Income unit type	*Head in workforce*	*Not in workforce*	*Head in workforce*	*Not in workforce*
Couple	684.57	587.78	500.69	403.8
Couple, 1 child	822.89	726.1	622.38	525.59
Couple, 2 children	961.21	864.42	744.07	647.28
Couple, 3 children	1099.53	1002.73	865.76	768.97
Couple, 4 children	1237.85	1141.05	986.11	889.31
Single person	511.75	414.05	344.4	247.61
Single parent, 1 child	656.99	560.1	473.01	376.22
Single parent, 2 children	795.21	698.42	594.7	497.91
Single parent, 3 children	933.53	836.73	716.39	619.6
Single parent, 4 children	1071.85	975.05	838.08	741.29

(a) Based on seasonally adjusted household disposable income per head per week for the September quarter 2017 of $819.09.

(b) All figures refer to income after tax.

Source: Melbourne Institute of Applied Economic and Social Research, *Poverty Lines: Australia, 2017.*

Measures of Inequality

Equality, like welfare, has many dimensions such as health and environmental goods as well as income. However, for most practical purposes, estimates of inequality are based on measures of the dispersion of incomes. These may be based on individual, household or equivalised household income and may be gross or disposable income.

Pictures of frequency distributions, like Figure 20.2, illustrate the dispersion but do not measure it. Here we discuss some descriptive measures of inequality.

As we have observed, a common practice is to divide the population into percentile groups, such as quintiles or deciles, and to show the proportion of income accruing to each group. But this does not provide an overall measure of income variance or inequality.

The variance (V) of a distribution provides a simple summary measure of dispersion:

$$V = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \tag{20.3}$$

where y_i is the income of the i^{th} individual (or household), $\bar{y}$ is average income and there are $i = 1...n$ individuals. The variance takes into account the whole distribution. Any redistribution that brings an individual's income closer to the mean reduces V. However, the variance does not satisfy the principle of independence of scale. The level of inequality, as measured by the variance, changes when the scale in which income is measured changes. If everyone's income doubles, there is no change in inequality but V quadruples.[6]

The coefficient of variation (C) avoids this problem because it normalises the square root of variance on average income.

$$C = \frac{V^{0.5}}{\bar{y}} \tag{20.4}$$

Thus, C is independent of the level of income. However, this measure has the disadvantage that any transfer from a higher to a lower income individual has the same effect on C regardless of the income level of a recipient. Box 20.2 discusses inequality measures and two principles of transfer.

The variance of the logarithm of income (VL) avoids this issue because it gives greater weight to transfers to lower incomes.

$$VL = \frac{1}{n}\sum_{i=1}^{n}(\log y_i - \log \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}[\log(\frac{y_i}{\bar{y}})]^2 \tag{20.5}$$

VL also has the advantage that it is invariant to the absolute level of income (unlike V). However, like V and C, VL considers only differences in incomes from the mean and it squares those differences. Such procedures are arbitrary.

Much the most commonly cited measure of inequality is the Gini coefficient (GC). This is illustrated in Figure 20.4 on page 352. The Lorenz curve shown in the Figure shows the cumulative proportion of total income received against the cumulative proportion of the income recipients (individuals or households) from the poorest to the richest. In Figure 20.4 the dotted lines show that 40 per cent of the poorest in the society received 20 per cent of the total income. If incomes were completely equal, the Lorenz curve would follow the 45° line. If one person earned all income, the curve would follow the horizontal axis and then the vertical axis. The GC is defined as:

$$GC = \frac{\text{the area between the Lorenz curve and the diagonal}}{\text{the total area under the diagonal}} = \frac{A}{(A+B)} \tag{20.6}$$

[6] These and following comments on measures of inequality draw on Barr (2004, Chapter 6).

Box 20.2 Inequality measures and two principles of transfer

A critical feature of an inequality measure is that its value should fall when income is transferred from a rich person to a poor person, without an interchange of their initial position on the income ladder. This is called the Pigou–Dalton condition of rank preserving transfer.

If $I(y)$ is an inequality measure and there is a small transfer of money from i (a rich person) to j (a poor one) then

$$dI(y) = \frac{\partial I(y)}{\partial y_i} dy_i + \frac{\partial I(y)}{\partial y_j} dy_j$$

$$= dy[\frac{\partial I(y)}{\partial y_i} - \frac{\partial I(y)}{\partial y_j}] < 0$$

where $dy = dy_i = dy_j$ (by construction).

This condition can be tested by differentiating the first four inequality measures (V, C, VL and GC) discussed in the text.

$$\frac{dV}{dy} = \frac{2}{n}[y_j - y_i]$$

$$\frac{dC}{dy} = \frac{y_j - y_i}{n\bar{y}\sqrt{V}}$$

$$\frac{dVL}{dy} = \frac{2}{n}[\frac{1}{y_j}\log(\frac{y_j}{\bar{y}}) - \frac{1}{y_i}\log(\frac{y_i}{\bar{y}})]$$

$$\frac{dGC}{dy} = \frac{j-i}{n^2\bar{y}}$$

These equations show that all these inequality measures satisfy the Pigou–Dalton transfer principle. Because $y_i > y_j$, *V*, *C* and *VL* satisfy this condition.

Another commonly supported transfer principle is that the measure should be sensitive to the kind of transfer. Specifically a transfer from a richer to a poorer person deserves more weight at the lower end of the income scale than at the higher end. Although both gainers are poorer than the losers, the gainer at the lower end is poorer than the gainer at the upper end.

Neither *V* nor *C* satisfies this principle: a transfer of a given amount of money from a high-income individual to a medium-income individual has the same effect as a transfer from a medium-income to low-income individual. Thus *V* and *C* are described as weakly sensitive to transfer.

On the other hand, *VL* can be strongly sensitive to transfer. That means a \$1 transfer from an individual with \$10 000 to someone with \$9900 would reduce *VL* by less than would a \$1 transfer amount from an individual with \$600 to someone with \$500. However, the sign of $\frac{dVL}{dy}$ depends on the nature of the function $\frac{1}{x}\log x$. The first differentiation of this function is $(1 - \log x) / x^2$. This is positive if $x > e$ (i.e., 2.718) and negative if $x < e$. If $\frac{y_i}{\bar{y}} \leq e$, a transfer from i to j will decrease inequality. But if $\frac{y_i}{\bar{y}} > e$, the same transfer will increase inequality. This means that, if the poorer of the two individuals has an income at least 2.718 times the mean income, *VL* will not satisfy the condition of transfer.

GC is calculated by arranging people in ascending order of income. A poorer person's rank is lower than the richer. Therefore, $j < i$. This implies that *GC* also satisfies the Pigou-Dalton transfer principle. However, the sensitivity depends on the relative ranking of the persons concerned. It does not depend on the income difference of the people who are affected by the transfer or on their position in the income ladder. Thus, for a given amount of money transfer, the Gini coefficient is most sensitive in the modal region.

where *A* is the area between the 45° degree line and the Lorenz curve and *B* is the area between the Lorenz curve and the x-axis. *GC* has two limiting values. If incomes are perfectly equal, A is zero and *GC* equals 0. If one person has all the income, *B* is zero and *GC* equals 1.0. Mathematically *GC* equals half of the arithmetic average of the absolute differences between all pairs of incomes, with the total normalised on mean income:

$$GC = \frac{1}{2n^2\bar{y}}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j| \qquad (20.7)$$

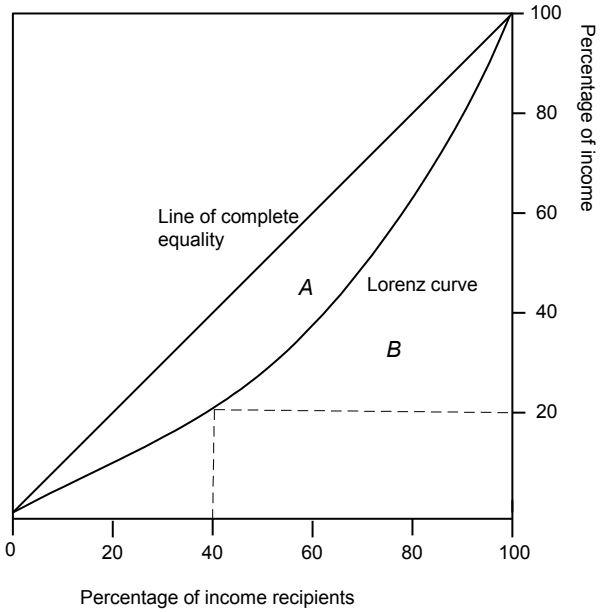


Figure 20.4 The Lorenz curve and the Gini coefficient

Sen (1997) provides an interpretation of the Gini formula as follows:

> in a pair-wise comparison the man with the lower income can be thought to be suffering from some depression on finding his income to be lower. Let this depression be proportional to the difference in income. The sum total of all such depressions in all possible pair-wise comparisons takes us to the Gini coefficient.

The Gini coefficient has several advantages. It is independent of the absolute level of income (as are *C* and *VL*). It avoids arbitrary squaring procedures. Also, it compares each person's income with the income of each other person rather than with mean income. Thus, it is an indicator of aggregate relative deprivation.

On the other hand, the *GC* has the disadvantage that, when a transfer of income is made from a rich to a poor person, the change in *GC* depends upon the relative rank of the persons concerned (i.e. the number of people between them). Thus, a transfer of income from a person with $80 000 to someone with $50 000 may have the same effect on the *GC* as a transfer of income from a person with $70 000 to someone with $60 000.

Like other measures of inequality, there are implicit value judgements. This can be understood by considering what occurs when Lorenz curves cross. The *GC* may be the same for both distributions, but one distribution may be preferred to the other. The preferred distribution depends on a value judgement as to whether inequality at the bottom or top end of the income scale matters more. The *GC* can be shown to be a weighted sum of individuals' incomes with the weights determined by the rank order position of the person by income level (Sen, 1997, p. 32). These weights are arbitrary and not necessarily consistent with any preferred social welfare function.

Actually, most descriptive measures of inequality incorporate mathematical functions, such as means, squares or weights, which may not reflect social values. Ideally these functions would reflect normative values, such that a reduction in an inequality measure could always be interpreted as an increase in social welfare. To interpret inequality measures in terms of

social welfare, it is necessary to integrate the measure of inequality with value judgements. The Atkinson index of inequality (Atkinson, 1970) exemplifies this approach. It is an attempt to estimate the welfare loss associated with departures from equal incomes. Formally the Atkinson Index (*AI)* is presented in Equation 20.8.

$$AI = 1 - [\frac{1}{n}\sum_{i=1}^{n}(\frac{y_i}{\bar{y}})^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}}, \ \varepsilon \neq 1$$
$$AI = 1 - \exp(\frac{\sum_{i=1}^{n}\log_e(\frac{y_i}{\bar{y}})}{n}), \ \varepsilon = 1 \quad (20.8)$$

In this formulation, ε may be taken as a measure of the concavity of the utility function (the diminishing marginal utility of income) or as an explicit measure of relative inequality aversion as in Chapter 7 (Equation 7.6).[7] The other variables are as before. $AI = 0$ if incomes are equally distributed. The value of AI rises to 1 as individual incomes diverge from the mean or as ε increases. Given that the value of ε must be chosen, AI is a normative measure of loss of welfare due to poor distribution rather than a purely descriptive measure of inequality.

In summary, all the above inequality measures, including simple descriptive deciles or quintiles, have been used to show or measure inequality. For example, in the OECD (2008, 2009) studies of income distribution in 13 countries cited below the authors used the *GC* along with minor variations of *V* and *VL* to estimate income inequality. As can be seen, the choice of measure can affect the inequality ranking. The *GC* is the most commonly used measure of inequality. The Atkinson index has also been used, though in this case the analyst must determine the equity parameter (ε).

Trends in Poverty and Inequality

Using as a criterion of poverty an EDHI of less than half the *mean*, Table 20.1 shows that around one in five Australian households were living in poverty in 2013-14 and 2015-16. Typically, government pensions and allowances constitute 90 per cent or more of gross household income for some 15 per cent of households and are the main source of gross household income for about 25 per cent of all households (ABS, 2011).

Saunders and Bradbury (2006) provide a more comprehensive analysis of trends in poverty. They estimated that, in 1998–99, 8.4 per cent of Australian households experienced severe hardship and 12.2 per cent experienced some hardship. Using a poverty line of one-half of the *median* equivalised gross household income, they calculated that the proportion of individuals in poverty rose slightly in the second half of the 1990s to reach about 13 per cent in the early 2000s. Note that median income is less than the mean which reduces the percentage said to be in poverty.

Adopting a similar poverty line, Rodgers *et al*. (2009) estimated that the poverty rate rose from 10.3 per cent in 1995–96 to 13.3 per cent in 2004–05. Reflecting rising living standards, they also noted that, judged against the 1995-96 poverty line, there were only 7.4 per cent below the poverty line in 2004–05. The percentages of poor people are also very sensitive to the choice of poverty line at 40, 50 or 60 per cent of median income.

---

[7] ε is the relative sensitivity to income transfers at different income levels. As ε rises, more weight is attached to transfers at the lower level of distribution and less weight to transfers at the top. This can be seen from

$$\frac{dAI}{dy} = \frac{y_i^{-\varepsilon} - y_j^{-\varepsilon}}{n\bar{y}^{1-\varepsilon}[1-AI]^{-\varepsilon}}.$$

The international picture on poverty trends in developed countries is not very clear (see Barr, 2004). Increases in wage dispersion and the rise in the retired elderly in the population in recent years have increased the proportion of people with relatively low gross incomes in many countries. However, to some extent, in the UK for example, this trend has been offset by increased government cash benefits.

Inequality

Findings on inequality are sensitive to the measure of income and the income unit, the period over which changes are estimated and the measure of inequality. Notwithstanding these issues, most studies of trends in inequality in Australia and elsewhere find that the dispersion of private earnings within countries has increased significantly over the last 30 to 40 years. However, the increased dispersion of private income has to some extent been offset by government taxes and grants as well as by social expenditures.

Many studies have confirmed the rise in market inequalities in Australia. Borland (1999) estimated that between 1975 and 1997 real weekly earnings of a male employee at the 25th percentile (ranked from the bottom) increased by 1.3 per cent, whereas earnings of an employee at the 75th per centile rose by 15.8 per cent. For female employees, the comparable figures were rises of 15.8 per cent and 31.2 per cent, respectively. Atkinson and Leigh (2007) found that the income share of the top 1 per cent and top 10 per cent of income earners was higher at the start of the 21st century than at any time over the past 50 years. The OECD (2011) reported that in 2008 the average income of the top 10 per cent of working-age Australians was 10 times higher than that of the bottom 10 per cent. This was up from a ratio of 8 to 1 in the mid-1990s.

Turning to disposable household income after transfers, the ABS (2011) reported that the P90/P10 and P80/P20 ratios for EDHI inclusive of government transfers rose significantly between 1994–95 and 2009–10. Over the same period, the estimated Gini Coefficient for EDHI income rose from 0.302 to 0.328. At 2016-16, the estimated GC was 0.323. This indicates that government taxes and transfer did not fully offset the increases in private income inequality that occurred. However, this does not take account of the distributional effects of government expenditures on goods and services.

Table 20.4 shows the share of disposable market income in 10 countries in 2000 and the changes between 2000 and 2014. In Australia in 2014, the top 20 per cent of individuals received just over 40 per cent of disposable market income whereas the poorest 20 per cent received only 7.5 per cent. These shares were broadly similar to those in the European countries (Germany, France, UK and Italy). In North America (the US, Canada and Mexico) inequalities were much larger. In the US, the top quintile obtained 50% of disposable market income and the lowest quintile a miserable 3.3%. In the Scandinavian countries (Norway and Finland) shares were a little less uneven. Table 20.4 also indicates that between 2000 and 2014, disposable market incomes became more unequal in Australia. They also became markedly more unequal in the US and Canada. In Italy, Norway and the UK, the incomes became less unequal.

Table 20.5 provides measures of inequality and poverty based on EHDI in 13 OECD countries between the mid-1990s and mid-2000s. The measures of inequality are the Gini coefficient (shown as 0 to 100), the squared coefficient of variation and the mean log variation (see notes to the table). Judged by the Gini coefficient, EDHI inequality rose in eight countries and fell in only five countries. Judged by the other two measures of inequality, inequality fell in only three and four of the 12 countries respectively (with no corresponding data available for the Netherlands).

The poverty level here is based on EDHI of less than 50 per cent of the mean. Using this measure, Australia has the third highest poverty rate after the United States and Japan. Of course, this poverty measure is in large part a reflection of inequality.

Table 20.4 Share of disposable income going to individuals by population quintiles

	Share of income by quintile, 2014					*Change in share of income by quintile, 2000 to 2014*				
	1st	*2nd*	*3rd*	*4th*	*5th*	*1st*	*2nd*	*3rd*	*4th*	*5th*
Australia	7.5	12.3	16.9	22.4	40.8	-0.2	-0.3	-0.7	-1.6	2.4
Canada[a]	4.1	9.6	15.3	23.8	47.2	-3.2	-3.1	-1.8	0.8	7.3
Finland	9.8	14.2	18.1	22.6	35.3	-0.2	0.1	0.5	0.6	-1.3
France	9.0	13.6	17.2	21.7	38.5	0.0	0.6	0.2	-1.3	1.5
Germany	7.5	13.5	17.6	23.0	38.3	-0.5	0.5	0.4	0.0	-0.5
Italy	6.8	13.0	17.6	23.2	39.3	0.5	1.2	0.8	0.2	-2.7
Mexico	4.5	8.7	12.8	19.3	54.8	1.1	1.5	1.0	0.3	-3.5
Norway	9.9	15.4	18.7	22.5	33.6	3.8	4.1	2.1	-1.5	-8.6
UK	7.8	12.8	17.1	23.0	39.3	0.4	0.7	0.8	1.0	-2.7
USA	3.3	9.0	14.8	22.9	50.0	-2.0	-1.8	-1.2	-0.1	4.9

Source: Aristei *et al.* (2009) and WIID (2016).

a) Canada is for 2011.

Table 20.5 Income inequality and poverty, mid-1990s to mid-2000s: equivalised household disposable income

Country/period	*Income inequality*				*Poverty*			
	Gini coefficient		*SCV[a]*		*MLD[b]*		*Poverty rate*	
	Final level	*Change %*	*Final level*	*Change %*	*Final Level*	*Change %*	*Final level*	*Change %*
Australia, 1993/94-2003/04	30	-2.0	39	4.0	17	-4.5	12.4	30.5
Belgium, 1995-2004	27	-9.7	30	-36.4	13	-51.7	8.8	-18.5
Canada, 1994-2005	32	12.7	59	90.9	18	25.9	12.0	34.8
Denmark, 1994-2005	23	6.0	60	162.0	10	13.6	5.3	6.0
Finland, 1995-2004	27	-10.3	81	233.3	13	44.4	7.3	49.0
France, 1990-2004	28	-3.8	37	-43.2	14	-52.5	7.1	4.4
Germany, 1994-2004	30	6.4	45	38.9	16	18.5	11.0	20.9
Italy, 1993-2004	35	1.4	110	88.4	24	0.0	11.4	-19.7
Japan, 1994-2004	32	20.8	41	38.5	20	58.7	14.9	84.0
Netherlands, 1994-2005	27	6.7	na	na	Na	na	7.7	-13.5
Norway, 1995-2004	28	9.4	46	50.8	16	22.1	6.8	-15.0
Sweden, 1995-2004	23	-23.3	65	199.5	10	-9.1	5.3	-17.2
United States, 1995-2005	38	10.5	81	83.7	29	32.4	17.1	0.0

(a) The SCV (squared coefficient of variation) index is the sum of the squared deviations of the income of each household from that of the population mean, divided by the square of mean income.

(b) The MLD (mean log deviation) index is the average of the log ratios of the income of each household to the mean income.

Sources: OECD, (2008, 2009).

Causes of Unequal Incomes

It is important to understand the causes of poverty and inequality as a basis for developing policies. Prevention of unequal incomes may well be regarded as preferable to the cure of income redistribution. Indeed, Australian governments have often intervened in factor markets with the aim of preventing or at least mitigating poverty and inequality. Accordingly, this chapter concludes with a brief review of the main causes of poverty and inequality.

Extreme poverty (lack of income) is explained principally by lack of earnings.[8] This in turn nearly always arises from unemployment, limited part-time employment or non-participation in the labour force. Non-participation is recorded formally as the proportion of people of working age (defined as 16 to 65 years of age) who are not working. But, in practice, of course it includes all those outside this age bracket who are not working, and notably the numbers of aged people with low savings.

The causes of income differentials are also a function of labour earnings but more complex as they also depend to some extent on life cycle effects and wealth.[9] Neville (1995) estimated that income changes over individual life cycles (from low to medium or high income and down again) account for a third of the inequality in disposable incomes and for a greater proportion of inequality in gross incomes.

Overall about two-thirds of private income derives from labour with the balance coming from capital including home ownership. Differences in income from capital may significantly exacerbate differences in earnings. However, as we have seen, in Australia the distribution of wealth is only weakly correlated with income at any point in time because wealth is correlated with age while earnings fall with age after a certain point. Therefore, we focus here on differences in earnings, which are in turn a major cause of differences in wealth.

Human capital and other individual factors. Individual earnings are a function of several interacting variables, notably ability (intelligence, health, strength and dexterity, and motivation), family background, education and training, and chance. Market structures and discrimination may also be important factors. Sorting out the impacts of these variables on earnings is complicated by the interactions between them.

As we saw in Chapter 12, human capital acquired by education, training and experience is a significant driver of earnings. Mincer (1974) estimated that educational capital accounts for between one-half and two-thirds of the variations in personal earnings. Much more recent studies, including those cited in Chapter 12, show that investment in a year of education at almost any level from Year 10 schooling to university graduation typically increases annual earnings by 7 to 12 per cent. Thus, three extra years of education increases earnings on average by between 25 and 30 per cent. Importantly, most recent studies (for example Card, 1999; Miller *et al*., 2006; Leigh and Ryan, 2008a; Barrett, 2012) take considerable care to sort out the different effects of education, ability and family background, to estimate the impacts of additional education.

Other important determinants of earnings are effort and the disutility of work for which compensation is required. To some extent, earnings differentials reflect hours worked. Also, workers are compensated for dangerous, or less pleasant, occupations. Conversely, workers who choose attractive occupations (for instance, in the leisure industry) usually accept lower earnings than they could earn elsewhere. In so far as earnings differentials reflect compensation for work effort and disutility, the differentials may be regarded as justified.

Markets and earnings. Market earnings are also affected by the nature of markets. Industrial concentration, market power, firm size and unionisation can have important effects on industry wage differentials. Managers and other workers in protected or monopolistic occupations with restricted entry can gain economic rents. Several studies have shown that labour unions play an important role in wage determination. Freeman (1991) reported that

---

[8] Addictive, excessive or indiscreet consumption may also cause poverty.

[9] In the history of economic thought there are two approaches to explaining income distribution. One approach, originating with Ricardo and called the functional distribution of income, seeks to explain the income share of the major factors of production. The other approach, stemming from Pareto and called personal income distribution, deals with the distribution of income among a set of economic units (individuals or households).

less unionised countries had higher wage inequality in the 1980s. Glittleman and Wolff (1993) found that industrial wage inequalities increased most with a low level of unionisation. In Australia, Borland (1999) found that income dispersion was lower among union workers than among non-union workers.

Some commentators (such as Goos and Manning, 2007; Goos *et al*., 2011) argue that labour markets are segmented into primary markets where skilled persons work and secondary markets where unskilled workers work. Workers in the primary market are said to enjoy high wages and fringe benefits, stability in employment, promotion possibilities and favourable (often regulated and protected) working conditions. By contrast, the less regulated secondary labour market offers low wages, few fringe benefits, little job security and little opportunity for promotion. Although some workers in the secondary market, such as students and mothers of young children, may eventually enter the primary labour force, other workers in this sector with little education, poor work history or a minority racial position may find exit is difficult. Minimum wage and other labour legislation may provide some base support in this sector.

Gender inequality is another important dimension of earning differences. Women working full time earned on average only 85 per cent of incomes of male workers in 2004 (National Institute of Labour Studies, 2006). Women were under-represented in full-time employment and over-represented in part-time employment, where they accounted for 38 per cent and 67 per cent, respectively, of the workforce. Explanations include lower educational attainment and work experience of women, partly due to interrupted career paths. Also, women may receive less on-the-job training than men because employers have less incentive to invest in female workers. In some markets, wage discrimination may be a significant factor.

Factor price equalisation theorem
Given free trade and other conditions, prices of factors of production will be equalised between trading partners

Increasingly the openness of an economy's markets determines incomes. The **factor price equalisation theorem** (Samuelson, 1948) shows that, given unrestricted free trade, identical technology and no transport costs between countries, prices of factors of production will be equalised between trading partners.[10] Countries export goods that make intensive use of their abundant factor(s) and import goods that are intensive in the use of their scarce factor. In the absence of transport costs, the trade in goods and services results in equal wages for equivalent labour even if factors of production cannot move between countries. Reduced commodity protection lowers the price of factors of production used intensively in production of that commodity. Therefore, wages in developed economies are influenced by the wages in the less developed countries with which they trade. This is mitigated when countries produce untraded goods, production technology differs between countries, and there are significant transport costs between markets. However, these constraints have been worn away by recent technical changes and the factor price equalisation theorem is a major explanator of recent changes in income distribution across and within countries.

In summary. Poverty is due primarily to non-participation in the workforce or to unemployment. Unequal earnings reflect many factors, including life-cycle position, ability, family background, investment in education, compensation for effort and the disutility of work, the nature of markets, and sometimes racial and gender discrimination. Unequal income from capital also contributes to income inequality.

Causes of changes in income differentials

Changes in income differentials *within countries* can be explained partly by an increased share of capital and a lower share of labour in the overall division of national income.[11] Atkinson (1997) showed that increases in real interest rates, dividends and share prices contributed significantly to increased inequality in the UK in the 1980s and 1990s. In

[10] This is also known as the Heckscher–Ohlin model of international trade.
[11] The global distribution of incomes is discussed in Chapter 34.

Australia, the share of gross operating profits (before tax) rose from 23 per cent of total factor income in 1997 to 28 per cent in 2010. Over the same period, the share of wages and salaries fell from 57 per cent of total factor income to 53 per cent (ABS, Cat. No. 5206.0).[12] In Australia, over the long run share market prices have tended to rise by more than labour incomes as profits increased faster than wages, but this depends partly on the period chosen for analysis. The significant rise in real residential property prices over the last 20 years has also been a factor in wealth and income distribution.

Turning to labour incomes, the increased dispersion reflects the demand and supply of labour. Earnings dispersion increases when demand for high-skilled workers increases relative to supply and/or demand for low-skilled workers falls relative to supply. This has occurred in the UK and the United States where labour demand has shifted towards skilled labour and away from unskilled labour (Atkinson, 1997). In Australia, the kind of work has also changed. Between 1981 and 1991, the ratio of professional workers to production workers in Australia increased by 92 per cent in manufacturing and by 52 per cent in all non-agricultural sectors (Tyers *et al*., 1997). Concurrent with these changes, the supply of relatively unskilled female and youth workers in the labour market has increased.

However, changes in the labour market are complex. Workers with tertiary qualifications have experienced a decline (on average) in relative earnings in Australia, probably due in part to the large increase in university-educated workers (Borland, 1999). This may also reflect a (possibly related) decline in standards of undergraduate degrees. On the other hand, demand has increased for people in prime age groups, perhaps reflecting a demand for experience. There has also been increasing earnings dispersion within occupational groups. This is often attributed rather unsatisfactorily to unobserved skill characteristics.

However, we need to do more than attribute changes to shifts in demand and supply for individuals with various skills. What causes these shifts? The four main explanations are: technical change, increased trade, offshoring of production and domestic regulatory changes. To some extent these changes are inter-related and all are related in some degree with globalisation. Gaston and Rajaguru (2009) provide time series econometric analysis to support the hypothesis that globalisation has increased inequality within Australia.

Technical change, such as changes in information and communications technology (ICT), has increased the demand for high-skilled workers and their market reach. OECD (2011) cites several studies that consider technical change, especially in ICT, is the key driver of increased inequality within countries, but acknowledges that it is hard to distinguish technical change from other drivers, especially offshoring that is driven by digitalisation.

In the 30 years to 2007, the share of global trade in world GDP increased from one-third to one-half (IMF, 2007). The standard theory of international trade discussed above suggests that trade integration increases inequality in developed economies. In these economies, the income of skilled workers will increase because these workers are overall in low supply while the income of less skilled workers will fall as they compete with imports produced by a large supply of low-skilled workers in less developed economies. With limited trade, less-skilled developed economy workers can earn premium wages because they work with more equipment, better technology and more skilled co-workers than their counterparts in developing economies. As trade increases, they can no longer obtain this premium. In Australia, increased imports have reduced the demand for locally produced textiles, clothing and footwear, many household goods, motor vehicles, and chemical and petroleum products. However, OECD (2011) reports that empirical work finds that international trade has had small, if any, effects generally on income inequality within countries.

Outsourcing of production, via foreign direct investment or simply by managing part of the production process such as call centres or data management, has also grown greatly in the last

[12] In OECD countries overall, wages and salaries account for 75 per cent of household income of working-age adults (OECD, 2011).

20 years. OECD (2011) finds that this has probably contributed to unequal earnings in developed countries more than has trade in final goods.

Deregulatory policies, whether related to globalisation or independent of it, have also tended to increase earnings dispersion in Australia and elsewhere (OECD, 2011). In Australia, over the last 20 years there has been significant reduction in industry protection and in employment protection legislation, more decentralised enterprise bargaining and less industry-wide bargaining (though this has been reversed somewhat recently), de-unionisation of the labour market and real reductions in the minimum wage. Gaston and Rajaguru (2009) found that de-unionisation has increased labour income inequality in Australia. Elsewhere, earnings inequality has increased most in countries like Canada, the UK and the United States that have relatively deregulated labour markets. It has changed less in countries like Sweden and Germany with more regulated labour markets.

In summary. Increased income inequality within many countries reflects both increased inequality of earnings and an increased share of income from capital. The four main forces of increased inequality have been technical change, increased trade in goods and services, the internationalisation of production and deregulation of many economies. Education of less skilled workers is the main counteracting force and policy.

Driven by globalisation, many of the factors that have increased earnings dispersion within countries and created its related social problems are likely to continue. However, while globalisation may increase earnings inequality within countries, it may reduce worldwide inequalities (see Chapter 34).

Summary

- An individual's income is the sum of his or her consumption and change in net wealth in a given period.
- Private income is derived from labour and capital. Labour income includes wages and salaries, own business income, employment supplements, cash income and the value of household services.
- Income from capital includes property rents, interest and dividends, pensions, capital gains and imputed rents from physical assets. Direct and indirect government benefits and taxes are also components of final income.
- The household is the most commonly used income unit because households share consumption. To allow for differences between households, equivalent household units are required.
- The poverty line is the level of income required to provide a minimally adequate standard of living for an income unit. It is usually about half the median income of the equivalent household unit in the relevant economy.
- Measures of inequality include the Gini coefficient, the coefficient of variation and the variance of the logarithm of income. Each measure embodies normative assumptions and has advantages and disadvantages, so that welfare conclusions should be drawn carefully.
- The number of people described as in poverty is sensitive to the determination of a poverty line and the inclusion or exclusion of housing and journey to work costs.
- Using half the median equivalised household disposable income as a criterion of poverty, about 15 per cent of Australian households would be poor before receipt of any government benefits.
- In recent years, private income has become more unequal in many OECD countries, including the United States and Canada especially and to some extent in Australia.
- The most important cause of poverty is lack of employment. Inequality reflects major dispersions in both labour and capital income.
- Market earnings depend principally on education, but also on position in life cycle, ability, family background, and compensation for effort and the disutility of work. Market structures can also influence earnings.
- Increased income inequality within most OECD economies reflects an increased share of income from capital and greater inequality of earnings. The major drivers of earnings inequality have been technical change, increased trade, internationalisation of production and deregulation of economies.

Questions

1. Adam Smith defined the necessities of life as 'not only the commodities which are indispensably necessary for the support of life but whatever the custom of the country renders is indecent for creditable people, even of the lowest order, to be without'. Discuss this statement with reference to the different notions of poverty.
2. How are household equivalence scales derived? What kinds of problems arise?
3. Are poverty lines useful? If not, then how can an economist usefully describe the amount of poverty in the community?
4. The Gini coefficient is the most commonly used measure of inequality. What advantages does this measure offer over other measures? What problems might arise in using this measure?
5. What is the evidence that incomes have become more unequal in Australia over the past twenty years?
6. How may the international movement of labour between a poor and a rich country affect income inequalities within and between the countries?
7. All definitions of (i) poverty and (ii) inequality involve value judgements. Discuss this statement giving examples.
8. Are racial minorities and women discriminated against in labour markets? What would be the definition of discrimination, possible reasons for discrimination and evidence for such discrimination?
9. Consider the data contained in the table below. For the three distributions of income, A, B and C, plot the corresponding Lorenz curves. Which distribution is the most equal? Which distribution is the least equal? In each case explain your answer.

	Proportion of total income going to each quartile			
Distribution	*Lowest quartile*	*Second quartile*	*Third quartile*	*Highest quartile*
A	4	12	30	54
B	12	16	34	38
C	8	18	30	44

10. The table below lists the weekly gross incomes of the persons in two countries, A and B.
    i. Determine the degree of inequality using the statistical measure of variance for countries A and B. Which country has a more equal distribution of income?
    ii. Using the coefficient of variation approach, determine the degree of inequality for each country. Which country has a more equal distribution of income?
    iii. Explain any differences that arise above. What do these differences suggest about inferences drawn from any measures of inequality?

Person	*Income ($) Country A*	*Income ($) Country B*
1	120	40
2	430	800
3	40	350
4	620	160
5	86	860
6	1020	460
7	340	380
8	90	760
9	200	450
10	780	1200

11. Calculate the Gini coefficients of males and females separately living in Ballarat and Bendigo:

Ballarat		*Bendigo*	
Male	*Female*	*Male*	*Female*
$60	$60	$30	$10
$70	$60	$30	$60
$80	$90	$130	$120

Now calculate the Gini coefficient for all males and for all females in both cities. What do you observe?

12. What are the main causes of increased earnings inequalities in OECD countries?

Further Reading

Aristei, D., Bacchettini, V. and Calzoni, G. (2009) 'Income Distribution, Growth and Well-being: Evidences for Selected OECD Countries', *Argumenta Oeconomica*, 22, 5-54.

Atkinson, A.B. (2015) *Inequality: What can be Done?* Harvard University Press, Cambridge, M.A.

Australian Bureau of Statistics (2017) *Household Income and Income Distribution, 2015-16*, Cat. No. 6523.0, AGPS, Canberra.

Barr, N. (2004) *The Economics of the Welfare State*, Chapter 6, Oxford University Press, Oxford.

Creedy, J. (1996) 'Measuring income inequality', *Australian Economic Review*, 29, 236–246,

Creedy, J. (1998) 'Measuring poverty: an introduction', *Australian Economic Review*, 31, 236–246, 82–89.

Gaston, N. and Rajaguru, G. (2009) 'The long-run determinants of Australian income inequality', *Economic Record*, 85, 260–275.

Hindriks, J. and Myles, G.D. (2006) *Intermediate Public Economics*, Chapter 12, MIT Press, MA.

OECD (2008) *Growing Unequal?: Income Distribution and Poverty in OECD Countries*, OECD, Paris.

OECD (2011) *An Overview of Growing Income Inequalities in OECD Countries: Main Findings*, OECD, Paris.

Saunders, P. and Bradbury, B. (2006) 'Monitoring trends in poverty and income distribution: data, methodology and measurement', *Economic Record*, 82, 341–364.

Saunders P. and Naidoo, Y. (2009) 'Poverty, deprivation and persistent poverty', *Economic Record*, 85, 417–432.

Rodgers, J.R., Siminski, P. and Bishop, J. (2009) 'Changes in Poverty Rates during the Howard Era', *Australian Economic Review*, 42, 300–320.

Sen, A.K. (1997) *On Economic Inequality*, Chapters 2 and 4, Clarendon Press, Oxford.

Chapter

21

Social Welfare by Market Intervention

The progress of human society consists in the better and better apportioning of wages to work.

Thomas Carlyle, *Past and Present*

Labour Market Operations ♦ Regulation of Labour Markets ♦ Regulating Immigration ♦ Interventions to Support Industry ♦ Interventions to Support Consumers ♦ Concluding Comments

Government can aim to achieve distributional objectives in two main ways. One way is by market intervention. Government can intervene in factor markets, usually with the aim of supporting low income workers but sometimes to protect the earnings and capital of other groups in the community. Government can also intervene in product markets, either to support some producer groups or to reduce the prices of goods consumed by low income households. In both cases government may intervene either with regulations or by fiscal instruments (usually subsidies of some kind). The second way to achieve distributional objectives is to redistribute the incomes produced by markets, usually (but not always) to create a more equal distribution of income after taxes and benefit transfers.

In this chapter we examine market interventions. In the first half of the chapter we discuss the operation and regulation of labour markets, which are the main source of income inequality. This includes a brief discussion of regulating immigration. Various other methods to support low income earners are discussed in the next two chapters. We then discuss ways to assist specific industries by restricting domestic competition, price supports and import controls. Finally we discuss interventions designed to assist consumers, including measures to improve housing affordability.

As we will see, regulating markets nearly always involves a deadweight loss (DWL). This does not mean that regulation is necessarily undesirable. It may be desirable if the distributional gain is great enough and there is no other, lower cost way to achieve this gain.

Labour Market Operations

Before we intervene in a market we need to understand how it works and its strengths and weaknesses. Accordingly, we first examine how competitive labour markets work and then consider how market failures and other weaknesses may create a case for government intervention.

Abstracting from workers' occupational preferences, labour markets are efficient when each worker works at the margin in their most productive occupation. In other words, labour is allocated efficiently when the value of a worker's marginal product (VMP) is as high in

their current employment as in any alternative employment. This requires in turn that each worker's wage equals their VMP. If the wage is higher or lower than VMP, too much or too little labour will be supplied to that industry respectively. These efficiency conditions require that a given type of labour is paid the same wage in each industry, and indeed in each firm. Formally,

$$VMP_{AX} = VMP_{AY} = w_A = w_{AX} = w_{AY} \qquad (21.1)$$

where A represents labour of type A, X and Y represent different industries (or firms) and w_A is the market wage rate for labour of type A.

Value of marginal product
The addition to output from a unit increase in factor input times the value of the extra output

These conditions are achieved in perfectly competitive labour markets. In such markets, labour demand and supply (which allows for occupational preferences) determine the equilibrium wage and the quantity of labour for each class of labour. Figure 21.1a shows the equilibrium market wage (w_A) for type A labour. Given this market wage, the labour supply curve to each small competitive small firm, in panels (b) and (c), is horizontal. Each firm is a wage taker. A firm cannot attract labour at a lower wage and has no incentive to pay a higher wage. These labour supply curves show each firm's average and marginal wage cost (MWC). Because MWC is constant, average wage cost (AWC) is also constant. Each firm employs labour up to the point where the marginal revenue product of labour (MRP) equals MWC. MRP is the addition to revenue from a unit increase in a factor input. Similar conditions apply to all firms in each industry employing similar labour. Perfect competition in the product market ensures that MRP equals VMP. This is because price equals marginal revenue in competitive markets. Firms employ labour until $MRP = MWC = VMP$.

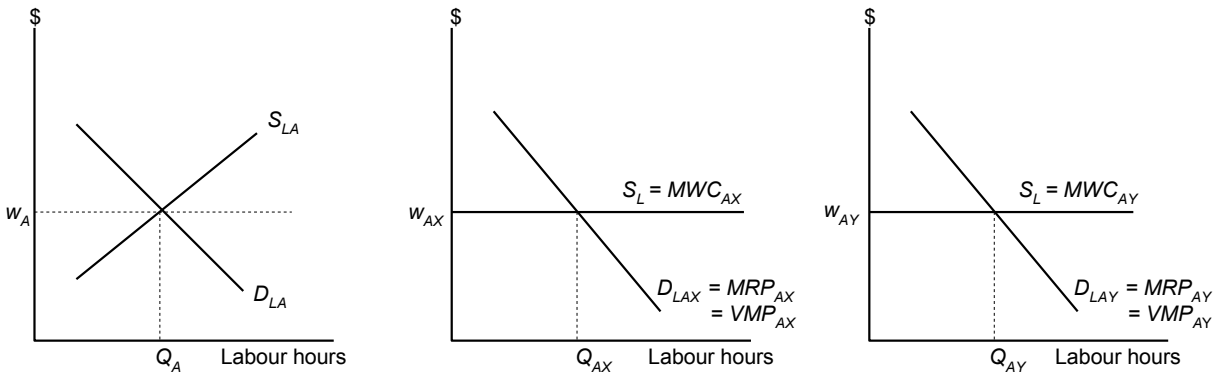
Average wage cost
Total wage divided by the number of workers

Marginal wage cost
The cost of employing an extra worker

Marginal revenue product
Addition to revenue from a unit increase in factor input

Thus, perfectly competitive labour markets are efficient. Labour is paid a wage that reflects both the value of the worker's marginal product and their occupational preferences. Given human capital endowments, the equilibrium wage produces the maximum value of output at any point in time. The market also provides an incentive for workers to invest in human capital and increase their productivity.

Competitive labour markets also compensate for differences in working conditions with regard to safety, comfort, location and so on. As we saw in Chapter 15, workers trade off wages for amenity. The market wage for jobs with onerous conditions must rise sufficiently to compensate the marginal worker employed for the disutility associated with the conditions.



Source: Derived from McConnell, Bruce and Macpherson (1999).

Figure 21.1 Efficient labour allocation in perfect competition

Efficiency and equity. Competitive markets may be viewed as fair as well as efficient in that each worker is paid the value of their marginal product and compensated for any disutility of working conditions. However, because wages reflect human capital, workers with low levels of human capital (innate or acquired) will be less productive and earn less income. Efficient labour markets can produce very unequal outcomes.

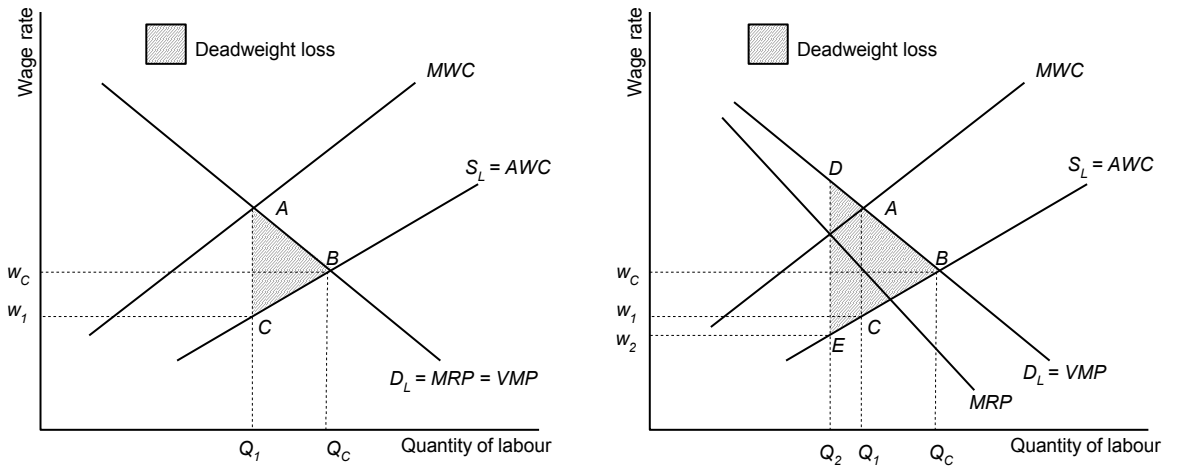
Market failures

The major market failure in labour markets is imperfect competition. This may reflect the market power of large firms or of worker unions. A key feature of a large firm is that it faces an upward sloping labour supply curve. Unless the firm can discriminate between workers, it must pay a higher wage to attract marginal workers and pay all workers this higher wage. In this case, MWC exceeds AWC.

In Figure 21.2a, the MWC schedule sits above the AWC schedule, which is also the labour supply curve. The firm maximises profit by equating MRP with MWC. Given a demand for labour (D_L), and assuming in this case that the product price is fixed (say internationally or nationally) so that MRP = VMP, this implies a wage of w_1 and employment of only Q_1 workers (compared with a wage of w_c and employment of Q_c in a competitive market). The DWL is the difference between the value of marginal product forgone and the opportunity cost of labour, which is given by area *ABC*.

Suppose now that the single buyer of labour is also a single seller of the product and that the price falls with increased output, the MRP schedule sits to the left of the labour demand curve. This is shown in panel (b). A profit-maximising firm now sets MRP = MWC. The wage falls to w_2 and employment falls to Q_2. On the other hand, DWL rises to area *DBE*.

Turning to the supply of labour, workers often combine with the aim of raising wages and improving working conditions. Typically a union establishes a minimum wage at which members will offer labour. In a competitive labour market, this raises wages, reduces the labour employed and cause deadweight losses. In Figure 21.3a, a union sets a minimum wage of w_u. The labour supply curve now becomes $w_u ABS_L$. The employer equates MRP with MWC, the wage rate rises from w_c to w_u and labour employed falls from Q_c to Q_u. Workers who retain their jobs have higher wages. However, there is a DWL due to the lost surpluses of employers and workers, which is represented by area *ACD*.



(a) Monopsony: with product prices set by the market

(b) Monopsony and monopoly: product price falls with firm's output

Figure 21.2 Wage rate and employment with a large firm

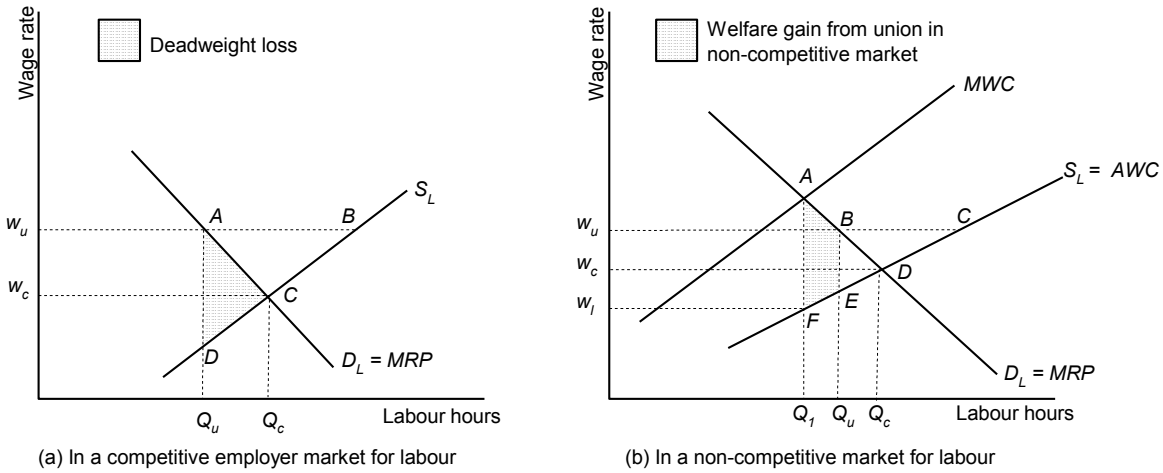


Figure 21.3 Effects of a union on wages and employment

On the other hand, if a large firm is effectively a wage maker, a union that sets minimum wages can have remarkably beneficial effects. In Figure 21.3b, the firm's demand for labour is shown by the downward-sloping curve (D_L), and the initial supply of labour by an upward-sloping curve (S_L). The MWC is again higher than AWC or the labour supply curve. Note that we assume here as in Figure 21.2a that the firm is a price taker in the product market. A profit-maximising firm will employ Q_1 labour (where MRP = MWC) and pay a wage of w_1. Now, if a union sets a minimum wage of w_u, the firm will employ Q_u workers. Although employment remains below the competitive equilibrium (Q_c), union action here increases both employment and the wage level! There is a welfare gain equal to area *ABEF*.

Two other labour market failures may be observed. One is a lack of information about working conditions. This may mean that wages for similar work or conditions may differ. Perhaps more important, there may be asymmetric information about workplace safety. As discussed in Chapter 15, this is an important reason for government regulation of the labour market.

Second, discrimination on the basis of race or gender may create barriers to employment. This is inefficient because labour is misallocated and productivity falls. It is unfair because discrimination between workers raises the wages of those who are protected from competition and lowers the wages of those whose choice of occupation is reduced.

Finally, labour markets may adjust slowly to wage differentials. Barriers to labour mobility may be geographical or institutional. Geographical barriers exist when movement costs offset any wage gains from moving, thus maintaining geographical wage differentials. Institutional immobility occurs when workers' associations place barriers to the movement of workers between jobs or when workers' long-term financial security depends on their loyalty to a firm. These labour supply constraints may prevent wages from adjusting in the short run to demand and supply conditions but they do not reflect market failures.

In conclusion, various market failures may warrant intervention in the labour market. But the prime concern is likely to be the unequal distribution of earnings which can occur in any market, including competitive and efficient markets. Whether intervention in markets or other polices are the best way to deal with this is a major, old and ongoing policy issue.

Regulation of Labour Markets

Most governments set rules for employment contracts. These rules are designed to protect employers and employees. In Australia, employee protection is more common. Examples are protection against unfair dismissal, leave and holiday arrangements, maternity arrangements and termination payments. Also, industrial tribunals (courts) at federal or state level regulate the length of a working week, penalty rates, minimum periods for working shifts and other such matters.

In addition, governments set rules for workers' associations. For example, government may set rules concerning union membership, the right to hold secret ballots, the right of workers to combine and negotiate as a group, the right to stop work or to strike, the right to operate secondary boycotts and so on. All such regulations may affect labour market outcomes.

Consider, for example, a regulation imposing employment termination conditions and redundancy payments on employers. Figure 21.4 illustrates possible effects in a competitive market. If the regulation does not change the behaviour or MRP of workers, the demand for labour curve is unchanged. However MWC rises with the cost of the regulation from MWC_1 to MWC_2. Given that profit-maximising firms equate MRP with MWC, employment will fall from Q_1 to Q_2, the wage rate will fall from w_1 to w_2, and the cost of employment to the employer will rise from w_1 to $(w_2 + r)$ where r is the cost of the regulation. Evidently the cost of the regulation may be borne in part by workers it is intended to assist. In addition, there would be a DWL given by area ABC.

However, other scenarios are possible. First, this analysis assumes no shift in the supply curve. If the new employment conditions are attractive to labour, the supply curve for labour will shift to the right. Indeed if workers value the benefits as equivalent to money wages, the labour supply curve would shift so far that the new MWC schedule inclusive of the cost of the regulation would be little different from the initial MWC_1. In an extreme case, workers would accept lower wage rates that directly offset any costs to employers and the MWC schedule with the regulation would be the same as without it. There would then be *no* change in employment and *no* DWL.

On the other hand, if the wage rate is fixed by government or judicial regulation then employers bear the whole additional cost of the regulation. The wage rate will not change, but there will be a greater fall in employment and a greater DWL than shown in Figure 21.4.

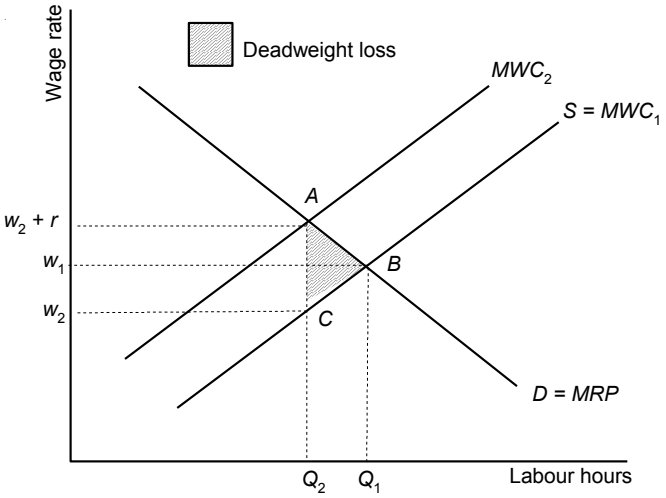


Figure 21.4 The cost of labour regulation in a competitive market

Box 21.1 Impact of dismissal cost on employment in Australia

In 2006, the Australian government introduced legislation to exempt businesses employing fewer than 100 workers from unfair dismissal laws, enabling small and even medium-sized firms to dismiss workers without special cause or cost. The government claimed that this would increase employment by 77000 workers.

Freyens and Oslington (2007) found that this was a major overestimate of the impact. They estimated that the costs of dismissal vary by cause from one month to four months' wage cost (depending notably on whether the dismissal is contested) and that redundancy costs are equivalent to four to five months' wages.

With an average tenure of 3.6 years for dismissed workers and 5.8 years for retrenched workers, dismissal costs average about 5 per cent of the total wage cost of a dismissed employee and about 7 per cent for a redundant employee. However, they observe that only 3 per cent of workers are dismissed and 14 per cent are declared redundant. Accordingly, dismissal costs amount to only 0.15 per cent of the total wage bill and redundancy costs to 1 per cent of the wage bill. Allowing a labour demand elasticity of 0.6, Freyens and Oslington estimated that exempting small firms from the direct costs associated with the unfair dismissal protection would create only an extra 6000 jobs.

In practice, matters may be still more complex. Workers are not homogeneous. Workers who are most protected by law may be the ones that employers are least keen to employ. Moreover, because regulations are hard to enforce in the informal sector, jobs will switch from formal to informal employment. However, the actual impact of termination regulations on employment is an empirical question. As shown in Box 21.1, some Australian research suggests that the effect of dismissal costs on employment is small. More recent experience reported in Box 21.3 (pages 369-370) suggests that these findings may be optimistic.

In imperfectly competitive markets, wage outcomes depend in part on the power of each party to extract a higher proportion of the economic rent from negotiations. The larger and stronger the employees' union, and the lower its liability for damages incurred, the greater the proportion of the economic rent it is likely to extract. The more power that the government allows to organised labour, the higher are the wages of organised labour likely to be. However, this may also be at the expense of employment.

Wage regulation

Wages may be regulated for any level of skill or kind of occupation. However, regulation of minimum wages for unskilled workers is the most common form of regulation and we focus on this below. Of course, similar analysis could apply to any level of wage regulation.

Figure 21.5a overleaf shows the demand and supply of labour in a competitive labour market with an equilibrium market wage (w_c). However, the wage regulator determines a minimum wage of w_r. This has no effect unless it is higher than w_c. Employment would fall from Q_c to Q_r and the DWL would equal area ACD. Part of this DWL is borne by workers who lose employment. Note that in this case the fall in employment depends entirely on the elasticity of demand for labour (in this case the elasticity of demand for unskilled labour). The labour supply schedule does not affect the employment outcome because there is no shortage of workers at the higher regulated wage. Le and Miller (2000) estimated that the Australian demand elasticity for labour is between –0.4 and –0.8. In a review of local and international literature, Lewis (2006) found that the demand elasticity is generally between –0.2 and –0.8. The DWL is low when labour demand is inelastic, but rises as the demand elasticity rises.

Panel (b) introduces an unregulated sector. Regulation in one sector causes labour to move to the unregulated sector, where the supply schedule shifts right from S_1 to S_2. Employment in this sector rises from Q_c to Q_2, but the wage rate falls from w_c to w_2. The area w_cEGw_2 represents a transfer from workers to employers. However, the DWL in panel (a) is now offset partly by the gain in producer surplus in the unregulated sector given by area EFG in

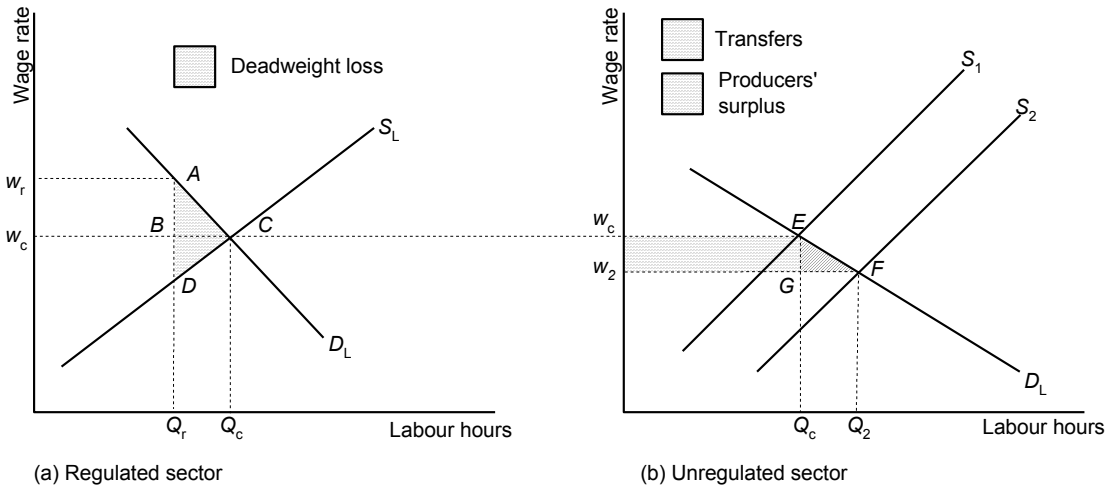


Figure 21.5 Effects of wage regulation in a competitive industry

panel (b). There is no labour surplus in panel (b)—the new employees in this sector have a net loss of surplus of area *BCD* in panel (a). Even if there is no change in total employment, there is a net DWL due to the inefficient allocation of labour.

However, minimum wage regulation may be efficient when employers have market power and are wage setters and so face an upward-sloping MWC curve. The argument is similar to the case of a union setting a minimum wage in an imperfectly competitive labour market, except that here government sets the minimum wage. As we saw in Figure 21.3b, minimum wage setting can increase both employment and the wage level. In this case, regulation would be efficient and may also be viewed as equitable.

International experience

Regulation of labour markets and employment outcomes vary considerably across OECD economies. In the United States the labour market is broadly deregulated and there is little welfare provision for the unemployed. There are some minimum wage provisions, but these are lower than in other less affluent OECD countries. However, as shown in Box 21.2, even these low minimum wage requirements have some negative effects on employment in the United States.

Box 21.2 Minimum wages and employment

In a major review of international labour market studies, Neumark and Wascher (2006) reported that most credible studies found that minimum wage regulations had a negative employment effect in the United States and other countries, especially for the least skilled groups. In a further study, Neumark and Wascher (2007) estimated that, in the United States, an increase in minimum wages of one per cent reduced the employment of Hispanics and Blacks under the age of 20 by 0.5–0.6 per cent.

The estimated negative effects for older minority group men and for young white men were smaller. They found little evidence of an effect of minimum wages on young female employment. The authors pointed out that the effects are likely to be smaller following the welfare reforms in the mid-1990s in the United States that require welfare recipients to work and the expansion of the earned income tax credits that encourage labour supply.

In continental Europe, labour markets are highly regulated and dismissal of workers is often difficult and expensive. Workers are less willing to accept the real wage reductions that have occurred in the United States. These conditions inhibit employment. For total employment, job creation has been more important than job losses. In the 1980s the United States lost 2 per cent of its jobs each month, while Europe lost only 0.4 per cent. Even so, between 1973 and 1994 the US generated a net increase of 38 million jobs while Western Europe reported no net increase. This suggests a trade-off. Maintaining higher wages for those in work may increase unemployment along with zero wages for those unemployed. The employed also pay higher taxes to finance social security payments. In the US, over three-quarters of working age people work; in Europe, the participation rate is under 70 per cent. Unemployment rates tend to be higher in Europe, even though a higher proportion of people of working age also leave the labour force, supported by government disability payments.[1]

In the UK and New Zealand, governments substantially deregulated the labour market and cut unemployment benefits in the 1980s. Unemployment rates fell in the 1990s. However, some unemployed persons moved to other benefits, such as disability benefits or pensions, or dropped through the safety net.

Since the mid-1980s, Australia has also moved from a highly regulated labour market to a much less regulated market notwithstanding some recent policy reversals (see Box 21.3). In the same period, employment has risen considerably and unemployment rates are lower than for 30 years. However, how much of this was due to labour market deregulation and how much to a global economic environment that has been very favourable to Australia remains to be sorted out. Also, earnings dispersion has increased significantly.

Box 21.3 Labour market regulation in Australia

More than most countries, Australia has used wage awards as a partial substitute for social welfare expenditure. In 1904, the Commonwealth established the Commonwealth Court of Conciliation and Arbitration to arbitrate on industrial disputes. In the path-breaking Harvester Case in 1907, Justice Higgins was required to arbitrate on the *Excise Tariff Act*, which provided tariff protection to firms who paid a fair and reasonable wage. Higgins determined that a fair wage was based on needs—the award should provide a minimum acceptable standard of living for a family (with a full-time employee, a wife and three children). For the next 90 years, most governments supported a needs-based minimum wage determined by industrial tribunals. Occupational benefits were also regulated. Awards required employers to pay for absence due to sickness and for long-service leave.

Labour market regulation in Australia had two key features: industrial (court-based) tribunals and centralisation. The tribunals made wage determinations based on needs, comparative wage justice and capacity to pay.

Comparative wage justice attempted to allow margins for skills; workers with similar skills and doing similar work for different employers or in different locations should receive similar wages. In the 1980s, wage awards were tempered by the view that the economy could not afford large wage increases.

But decisions remained centralised. The federal tribunal was the dominant industrial court. All courts made awards for wages and non-wage conditions for all members of a trade regardless of the industry they worked in or the location of the work. National awards were set for occupations regardless of local demand and supply conditions. Labour was overpriced where there was low demand for labour and underpriced where demand was high.

In 1993 the Labor government introduced enterprise bargaining, which represented a major shift to decentralised wage setting. Then in 1996 the Liberal-National Coalition government introduced the *Workplace Relations Act 1996*.

This Act reduced the power of the Australian Industrial Relations Commission (AIRC) to making awards for 20 matters (still covering most major concerns including pay, penalty rates, leave allowance, redundancy payments and so on). It encouraged development of enterprise-based collective agreements, known as Certified Agreements, which had to be consistent with AIRC determinations but allowed variations. Also, it allowed individual contracts, Australian Workplace Agreements (AWAs), whereby individuals could negotiate directly with employers without union involvement.

continued

[1] Some of the unemployment in Europe may reflect fixed exchange rates in the European Union.

In December 2005, the government established the Australian Fair Pay Commission to take over the wage-setting and adjusting functions of the AIRC. In setting minimum wages, the commission was to have regard to employment and competitiveness across the economy as well as an appropriate safety net for the low paid. AIRC retained its role in dealing with employment disputes.

In March 2006, the government introduced WorkChoices amendments to Australian labour law. This Act streamlined Certified Agreements and AWAs. Under the changes, AWAs had to meet only five minimum standards contained in the Australian Fair Pay and Conditions Standard: the minimum wage, four weeks annual leave (unless negotiated away), 10 days personal leave a year, 38 normal working hours per week, and unpaid parental leave for up to a year. The amendments facilitated dismissal of workers (exempting businesses with fewer than 100 staff) and allowed an AWA to override employment conditions in state or territory laws, except for occupational health and safety, workers' compensation or training arrangements.

The WorkChoices amendments proved politically unpopular and after the election of the Labor government in 2007, the new parliament passed the *Fair Work Act* in July 2009 which replaced the *Workplace Relations Act*. The main features of the *Fair Work Act* were increased worker protection from unfair dismissal (including for small firms previously exempt), 10 minimum employment conditions, rules governing industrial action, assistance for bargaining for low paid workers, widened provision for unpaid parental leave and rights to request flexible working arrangements.

These changes have increased the cost of employment. In 2010–11, 37 262 employees lodged cases of unfair dismissal or general protection as compared with 17 658 cases in 2009. The onus is on employers to disprove a claim and they often find it less costly to pay out claims than to contest them.

Assessments of Australian regulation of the labour market vary. When transport and communications constraints allowed a few firms to dominate markets, labour market regulation may have protected wages with possibly little impact on output or employment. Also, if tariffs were to provide economic rents, it was fair that labour should share the rent. However, in an open and competitive economy it is harder to protect labour by regulating employment conditions as the demand for labour is much more elastic.

Regulating Immigration

Given that wages depend on labour supply, restrictions on immigration of labour represent a possible strategy for maintaining or even increasing wage levels, especially restrictions on labour that can substitute most easily for local labour. However, immigrants with the skills to obtain employment in the local economy or the willingness to take on jobs that local workers do not want usually provide a net benefit to the local economy. The following simple example shows this

Consider the demand and supply conditions shown in Figure 21.6. The D_L curve shows the local demand for labour to service domestic consumption and exports. The domestic labour supply curve is S_D and there is a perfectly elastic supply of foreign labour given by the horizontal schedule S_F. With no immigration, there would be Q_1 local employment with wage w_1. Suppose now that limited labour immigration is allowed such that total employment rises to Q_2 and the wage falls to w_2. Employment of local workers would fall from Q_1 to Q_D and immigrants would take $Q_2 - Q_D$ jobs. However, the gross value of local output would rise by area ADQ_2Q_1 and there would be a net welfare gain to the local community of area $ABCD$. This gain is the sum of the reduction in costs due to using $Q_1 - Q_D$ imported labour which is shown as area ABC and the surplus associated with the $(Q_2 - Q_1)$ increase in low cost labour which is area ACD. These benefits arise because immigrants are willing to work at a lower wage than local labour. In this model, the fall in local wages from w_1 to w_2 is more than offset by the benefits of lower consumer prices and increased return to capital. Immigration is efficient but would reduce local wages and employment.

However this simple model does not account for the heterogeneity of the workforce or the flow-on effects of immigration. The increase in employment of labour and capital is likely to increase the demand for goods and services including demand for non-tradable goods like housing. Immigrant labour may also complement existing labour or capital and increase productivity and output including in export markets. These factors will increase the demand

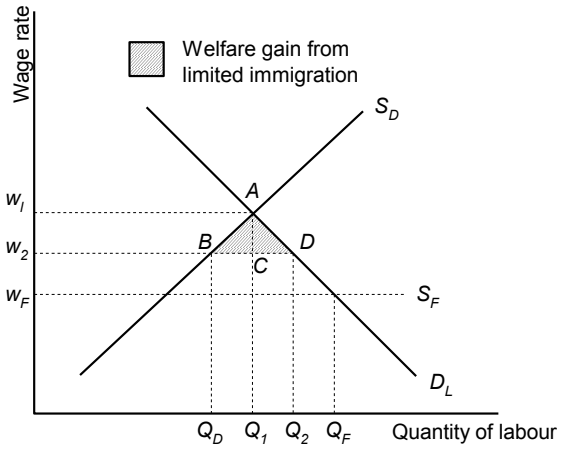


Figure 21.6 Partial model of immigration and the labour market

for labour and may increase local, as well as total, employment compared with the pre-immigration case.

Evidently, the impacts of labour immigration on wages and local employment depend on the circumstances, especially on skills that are not readily available locally. The effects depend also on the availability of capital to support labour and the flow-on effects on the demand for goods and hence for labour. A partial equilibrium analysis suggests that, although immigration initially reduces the wage rate, local households gain overall because lower prices and higher returns to capital offset the reduction in wages. However, modelling immigration effects is complex and a computable general equilibrium model is required to model the full effects of migration on the economy.

Interventions to Support Industry

Government may assist industries with financial support or by regulations. In both cases the prime objective is to increase labour incomes and returns to capital in the assisted industries. However, whereas taxpayers generally bear the cost of financial assistance to industry, consumers pay for regulations in the form of increased prices.

Table 21.1 overleaf outlines various industry assistance strategies. These include regulations to restrict entry into the industry, financial assistance to industry, price supports and protection from international competition. The table also cites some effects of these policies and examples.

The general stance of economists is that most assistance strategies, whether financial or regulatory, encourage resource misallocation and are inefficient. The key premise is that the competitive market understands market conditions and makes efficient investment and employment decisions. Assistance for selected industries encourages the employment of resources in relatively unproductive industries. It is an economic waste to subsidise labour to work in one industry when it could produce higher valued goods in another industry. It should be said that this is not the view of most Australian state jurisdictions which have active industry programs to avoid unemployment by supporting industry re-adjustment or to assist start-ups that the market is not financing (or only partially financing).

However, at the Federal level, the Australian Government has largely cut its traditional protection of manufacturing, notably ceasing all subsidies for motor vehicle manufacturers, which have now closed completely in Australia.

Table 21.1 Examples of industry assistance strategies

Industry assistance strategy	*Policy instrument*	*Policy effects*	*Examples*
Protection from domestic competition via quantity controls	Licensing of occupational groups Mandating product use	May provide quality control and protect consumers, but also reduces supply of services and raises prices	Licensing of professions, e.g. doctors, lawyers, architects, surveyors Mandating use of Australian produced ethanol-based petrol
Direct financial assistance	Direct subsidies Taxation relief	Encourages relatively inefficient producers and so misallocates resources	State government subsidies for businesses that are re-structuring Payroll tax relief for new businesses
Price supports for output	Guaranteed local prices	Misallocates resources to products in low demand	Support for dairy product prices and incomes of dairy farmers Support for sugar cane growers
Protection from international competition	Tariffs on imports, import quotas, domestic content requirements Subsidies for exports	Import restrictions reduce supply of goods and raises prices Export subsidies misallocate resources	Motor vehicle tariffs Clothing/footwear quotas Local content of TV programs

Chapter 15 described the DWL associated with licensing occupations and restrictions on the supply of services. Restrictions on entry into the taxi industry prop up the capital value of taxi licences but increase the costs of taxi services and do nothing for incomes of taxi drivers (for whom there is reduced demand). Abelson (2010) estimated that the DWL due to high taxi fares and poor service amounted to over $250 million a year in Sydney alone. Another issue is the mandating of ethanol use in petrol. Apart from possible environmental reasons, a major aim is support for sugar cane farmers and ethanol producers. This will almost certainly increase oil refinery costs and petrol prices and may reduce vehicle efficiency.

Mandating or guaranteeing producer prices in excess of market prices is another common policy especially for farm products. Indeed, guaranteed minimum prices for agricultural products in the European Union and the United States are responsible for more distortion of international resource use than any other factor. Friedman (2003) noted that direct US governmental support for farm crops, including wheat, cotton, rice, sugar and other crops, cost over US$21 billion in 1999. This did not include indirect supports such as output restrictions that cost consumers many billions of dollars more in increased prices. In both continents the major aim is to assist farmers, although environmental benefits are also claimed. There are fewer examples of price supports in Australia. For a long time dairy farmers were guaranteed a minimum domestic price, but this practice has been discontinued.

Figure 21.7 illustrates the DWL of mandated or minimum prices. In this figure, Q_1 and P_1 represent the market equilibrium quantity and price respectively. However, suppose that government mandates a higher price of P_2 and agrees to buy any surplus output. Market output would increase to Q_2. Government would hold ($Q_2 - Q_3$) stocks, which it could sell only if the price fell to P_3. The DWL would equal area DBF, which is the difference between the cost of supplying the product and its value to consumers. Alternatively, government could mandate a price of P_2 but not agree to purchase any surplus. In that case, producers would restrict supply to Q_3 and there would be a DWL of area ADE due to undersupply.

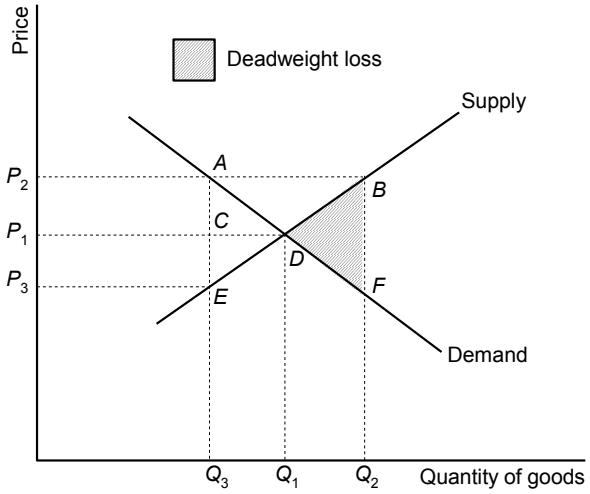


Figure 21.7 The deadweight loss of mandated producer prices

Import regulation

Regulation of imports by tariffs or by quantity (quota) restrictions are another form of industry protection. If the domestic market is competitive and quotas are auctioned or traded, for any given reduction in imports, the DWL is the same with quotas as with tariffs. Here we provide an analysis of the welfare loss with tariffs.[2]

Figure 21.8 shows the demand curve for a good that can produced locally or imported. It also shows three supply schedules: domestic supply (S_D), foreign supply without a tariff (S_F),

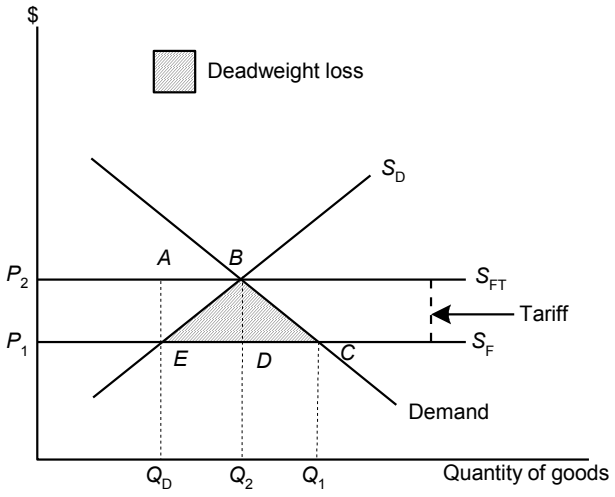


Figure 21.8 Deadweight loss of tariff protection

[2] The analysis here assumes that the importing country is a price taker (i.e. it is a small or medium economy). A large economy can reduce import prices and so can gain from imposing a tariff. This is not relevant to the subject here as the gains in this case accrue to government rather than to industry.

and foreign supply with a tariff equal to BD (S_{FT}). With no tariff, Q_1 units would be supplied at a market price of P_1. Local supply would be Q_D and imported supply would be ($Q_1 - Q_D$). With the tariff, supply falls to Q_2 while the market price rises to P_2. Local producers supply the whole market; there are now no imports. The total DWL equals area BEC. This is the sum of the loss of consumer surplus (area BCD) and the cost of the local resources used to meet demand compared with the cost of imports (area BDE). The area ABE is not a DWL because the gain in producer surplus offsets the loss in consumer surplus.

Interventions to Support Consumers

When the objective is to support producer incomes, government usually sets product prices above market rates. When the aim is to support consumers, especially those on low incomes, government typically mandates product prices below market prices.

Governments in many countries impose maximum price controls on commodities such as food, water, electricity and petrol. This is done most easily when goods are produced by public agencies and can be readily subsidised. It is harder to achieve by regulating the prices of private production. Unless these goods are also subsidised, this results in excess demand and chronic under-supply of these goods. Consumers suffer unless government maintains supply by paying higher prices separately to producers. But, if government does subsidise producers, there is a chronic strain on the government budget.

Housing rent and cost controls

Rent controls are a prime example of price controls that have been applied in many countries. The poor results have fully confirmed the predictions of economic theory. Figure 21.9 depicts the effects of rent control in the short and long run.[3] The main difference between the short and long run lies in the supply side. In the short run, the supply of rental housing is fixed. In the long run, capital can move out of housing (or not move into it) and the supply of housing is responsive to rents or, more precisely, to the rate of return on capital. The demand for

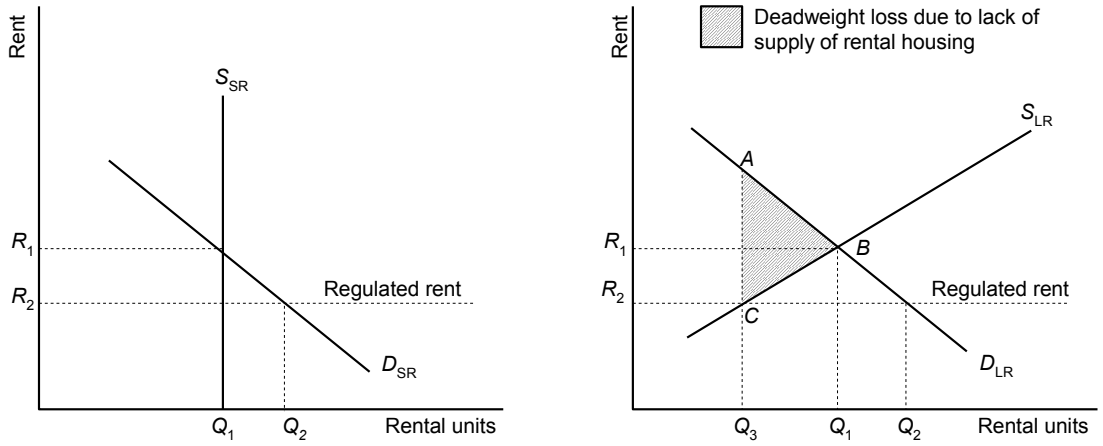


Figure 21.9 Effects of rent controls

[3] Figure 29.9 assumes a homogeneous standard of rental housing. Rent control is more complicated when housing standards vary greatly, as they generally do.

housing may also be more elastic in the long run as individuals can form more households in response to lower prices. Panel (a) depicts the short run. Rents fall from R_1 to R_2 (the regulated rent) and there is excess demand ($Q_2 - Q_1$) for rental accommodation. This may result in illicit black market arrangements that reward landlords. Alternatively, to accommodate the excess demand, occupancy rates rise so that the effective quality of the accommodation falls.

Panel (b) shows a greater shortfall of accommodation ($Q_2 - Q_3$) in the long run and the associated DWL. In addition to black markets and higher occupancy rates, while demand increases landlords will let the quality of their accommodation fall until the rate of return on their investment equals that in alternative investments. Thus, the rental unit in panel (b) is lower quality than the rental unit in panel (a). In many cities, including Sydney after the Second World War, New York, London, Alexandria and Mumbai, to mention just a few, rent controls led to substantial disinvestment in housing and to reduced and inferior housing stock.

Currently a popular proposal to deal with housing affordability is to require developers to provide part of their developments, say 10 per cent of the units, to community housing providers who will let out the units at 20 per cent below market rents. This is another form of rent control. In this case, the beneficiaries will be a few median income households (as low-income households cannot afford even discounted rents for new dwellings). It is not clear how the scheme will be managed, for example for changes in income status of the renter or for sub-letting. And it will probably have a minor disincentive impact on new construction and dwelling maintenance.

An example of cost controls. Traditionally Australian governments also attempted to reduce the cost of housing for owner-occupiers by regulating mortgage interest rates. This also exemplifies general problems of price controls. Up to 1985, the Australian government controlled the rate of interest that banks and building societies could charge on mortgages to house buyers. In effect, they were required to lend money to home buyers at below market rates. However, this had little effect on the real price of borrowing for home purchase. Because lending rates were controlled, the demand for funds exceeded the supply and the financial institutions had to ration their funds. They did this partly by requiring borrowers to have an account history with them before they could borrow. Borrowers had first to deposit funds with the institutions for a few years at low or zero interest rates. The institutions also restricted the amount lent on each property and thus reduced their risks. House buyers then borrowed the extra amount required in a secondary market at a high marginal rate. As so often occurs, the market found ways to nullify the effects of the regulations.

Effects of housing subsidies on prices and consumption

In Chapter 32 (Table 32.2) we will see that Australian governments heavily subsidise housing, especially owner-occupied housing, in an attempt to make it more affordable. Abelson and Joyeux (2007) analysed how the housing subsidy (net of taxes) affects house prices and consumption and the DWL. Some results are summarised here.

If the supply of housing is fixed, a demand subsidy of X per cent would simply increase house prices by X per cent. There would be no increase in housing consumption. Existing homeowners would receive a capital gain. New homeowners would pay a higher price for their houses and be no better or worse off than before. However, the analysis is complicated because the private rental sector also receives a small subsidy and because higher house prices may increase the supply of housing.

As usual, outcomes depend on demand and supply elasticities. Assuming plausible unitary demand elasticities in both homeowner and renter markets and a unitary supply elasticity, Abelson and Joyeux (*ibid*.) estimated that a 10 per cent subsidy to homeowners (with no subsidy to renters) would increase house prices by 4 per cent and the amount of housing

consumed by owner-occupiers by 5.8 per cent, but it would reduce the amount of housing consumed by renters by 13.5 per cent (due both to a switch to home owning and to the price effect on renting). The supply effect is critical. With limited supply response, a housing subsidy simply increases housing prices with little effect on housing consumption.

To estimate the DWL of the net housing subsidies, we need to compare what households are willing to pay for housing with the real opportunity costs of new housing. Drawing on the standard formula for DWL (see Equation 27.7, page 483) and assuming a competitive housing market, Abelson and Joyeux (*ibid*.) estimated that the DWL would vary from $0 per annum (if the supply of housing has is fixed) up to $450 million per annum if the supply is highly elastic. The efficiency loss arises because households consume goods that they would not consume in the absence of the subsidy.

Concluding Comments

Governments can intervene in markets in many ways to assist workers, industry or consumers. Market interventions designed to correct market failures can produce economic gains. Thus market interventions to support labour in markets where large firms fix wages or product prices can be efficient as well as fair.

However, most market interventions to redistribute factor incomes or to protect producers or consumers involve some loss of efficiency (deadweight losses). This holds for most regulations in labour markets, policies to regulate international movements of labour or goods, polices to assist industry and policies to protect consumers.

This does not necessarily rule out these policies. But economists will want to know, for any given redistribution of income, the efficiency costs (deadweight losses) arising from intervening in markets compared with a tax-transfer system that redistributes market incomes.

Summary

- In this chapter we review how government may regulate factor or product markets with the aim of improving distributional outcomes.
- Government can regulate labour markets in various ways, for example by setting the rules for employment contracts, establishing the conditions under which unions may operate, determining occupational and safety rules, or regulating wages.
- These regulations affect distributional outcomes and, when there are market failures, some regulations can also make markets more efficient.
- However, regulations of employment contracts, such as dismissal conditions or minimum wages, can also create distortions that affect the allocation of labour, reduce employment and result in a deadweight loss.
- Controls over immigration may protect some wages. But restrictions on immigration may reduce productivity, limit service provision and increase prices. The issues are complex and cannot be fully resolved without economy-wide modelling.
- Industry assistance policies include entry or quantity restrictions, financial assistance, price supports and regulations of imports.
- Such policies tend to have deadweight losses. Financial assistance distorts resource allocation. Price supports encourage oversupply of a good. For small or medium price-taking economies, tariffs have a deadweight loss.
- Governments sometimes mandate maximum prices in product markets to assist low-income households. Rent controls, a major example, reduce the quantity and quality of rental housing so that there is little, if any, overall gain for rental households.
- In summary, market regulations can improve distributional outcomes but often with significant efficiency costs. When there are such costs, policy makers need to know whether other policies could achieve the distributional objectives at lower cost.

Questions

1. Under what conditions may government regulation of minimum wages raise both wages and employment?
2. How do government regulations of labour unions affect the distribution of economic rents?
3. What are the main reasons for thinking that regulations of labour markets are likely to cause deadweight losses?
4. Suppose that government requires employers to pay mothers six months' maternity leave. Who bears the cost of this regulation?
5. How may the immigration of (a) skilled workers and (b) unskilled workers affect the earnings and welfare of existing Australian workers?
6. Explain why restrictions on taxi licences provide economic rent for owners of taxi licences but tend to reduce the earnings of taxi drivers.
7. Should farmers receive water at a below-market price? What are the possible efficiency costs?
8. The Australian government has recently (i) largely replaced individual Australian Workplace Agreements by contracts based on enterprise bargaining and (ii) re-introduced unfair dismissal regulations for medium sized businesses. The Department of Employment and Workplace Relations asks you to conduct an objective and fearless inquiry into the effects of these changes. How would you evaluate them?
9. In London, assessors of fair rents for rent-controlled apartments were required by law to base their assessments on rents set in comparable rent-controlled apartments and to assume that demand and supply for apartments were equal. Why is this assumption meaningless?
10. Australian government policy is to restrict use of regular unleaded petrol in order to promote the use of ethanol-based petrol. What are the likely effects?

Further Reading

Freyens, B. and Oslington, P. (2007) 'Dismissal costs and their impact on employment: evidence from Australian small and medium enterprises', *Economic Record*, 83, 1–15.

Friedman, L. (2002) *The Microeconomics of Public Policy Analysis*, Chapters 13–14, Princeton University Press, Princeton, New Jersey.

Lewis, P. (2005) 'Minimum Wages and Employment: A Reappraisal', Discussion Paper 05/2, Centre for Labour Market Research, University of Canberra, Canberra.

McConnell, C.R., Bruce, S.L. and Macpherson, D.A. (1999) *Contemporary Labour Economics*, Chapters 6 and 13, McGraw-Hill, New York.

Neumark, D. and Wascher, W. (2006) 'Minimum wages and employment: A review of evidence from the new minimum wage research', Working Paper 12663, <www.nber.org/papers/w12663>, National Bureau of Economic Research, Cambridge, MA.

Nicoletti, G. and Scarpetta, S. (2005) *Product Market Reforms and Employment in OECD Countries*, WKP(2005)59, OECD, Paris.

Productivity Commission (2011a) *Trade and Assistance Review 2009–10*, Productivity Commission, Melbourne.

Rosen, H.R. (1985) 'Housing subsidies', pp. 375–419 in *Handbook of Public Economics*, A.J. Auerbach and M. Feldstein (eds), Elsevier Science, North Holland.

Chapter

22

Social Welfare Provision

A decent provision for the poor is the true test of civilisation.

Samuel Johnson

Policy Objectives ♦ Social Welfare Provision in Australia ♦ Cash Transfers ♦ Transfers-in-Kind ♦ Social Insurance ♦ Incentive Effects of Social Welfare Programs ♦ Policy Implications

Social welfare provision is generally based on four pillars. These are the regulation of markets, cash transfers, transfers-in-kind and social insurance. Chapter 29 discussed market regulation. In this chapter we discuss the other three pillars.

Cash transfers are generally funded from consolidated revenue. In Australia, they total over 40 per cent of the Commonwealth government's own expenditure and about 9 per cent of GDP. **Transfers-in-kind** typically involve provision of free or subsidised goods such as health, food, housing or public transport. **Social insurance schemes** provide security for individuals who experience a specific adverse event, such as unemployment or sickness in the workplace. They may be publicly run or mandated private schemes. In some countries, such as the UK, public insurance schemes are a large part of the welfare program. In Australia, mandated private insurance is an important strategy for retirement.

Two further introductory points should be made. First, social welfare provision is not simply about relief of poverty, it is also about fairness. It involves fairness for many different types of persons, including persons with mental health problems, single parents and disabled persons, some of whom may not be poor as defined by poverty lines. Related to this, a social welfare system involves both expenditures and taxation. Indeed, cash transfers may be viewed as negative income taxes. In Chapter 27 we discuss the concept of an optimal income tax system, including negative income tax.

The chapter starts with a discussion of welfare objectives and a brief outline of the Australian welfare system. We then describe the major welfare pillars: cash transfers, transfer-in-kind and social insurance. The last part of the chapter discusses the critical issues of the incentive effects of welfare programs and the public policy implications.

Policy Objectives

Ideally, welfare policies would be designed to maximise social welfare as represented by a social welfare function. This approach is used to analyse optimal income redistribution (see Chapter 28; Kaplow, 2008). This formal approach clarifies and quantifies the nature of income trade-offs and choices. However, it is hard to articulate and narrow down many

diverse welfare programs and objectives into a formal social welfare function. For practical purposes, six looser welfare objectives may be identified.

1. **Poverty relief is a core aim of social welfare.** Most people believe that communities have a responsibility to protect all law-abiding members from severe poverty.[1] Others support poverty relief to reduce the negative externalities of poverty. On the other hand, there is also opposition to supporting individuals who can support themselves. In the words of Bill Clinton (1992, US Presidential campaign speech), 'No one who works full-time and has children at home should be poor anymore. No one who can work should be able to stay on welfare for ever'.
2. **Provision of economic security for all households.** The objective here is to protect individuals against major adverse events such as unemployment, sickness or disability and retirement when private insurance markets are inefficient or insufficient, even if these events do not cause severe poverty.
3. **Reduction in inequality.** The principle of vertical equity requires that income should be redistributed from people with more income and fewer needs to those with less income and more needs. This principle implies that benefits should depend on factors that determine needs, such as age, household composition and health status, as well as on income. Equality here is equality of welfare rather than simply equality of income. One consequence is that there are many potential categories of welfare recipients.
4. **Social integration** embraces social cohesion and the preservation of individual dignity. Social cohesion means developing and maintaining a compassionate and cooperative society, minimising alienation and social disorder. The preservation of individual dignity requires that welfare recipients should not feel, or be regarded as, less worthy members of society.
5. **Efficiency** requires that welfare programs should be provided at least administrative cost and with minimum deadweight loss (DWL). DWL occurs when programs create incentives that distort behaviour, notably withdrawing labour supply or reducing savings.[2]
6. **Good governance** requires that welfare programs be transparent, easy to understand, accessible and non-intrusive. Also, system abuse should be minimised. Benefits should be provided to those who are entitled to them, but not to the un-entitled. This may require detailed scrutiny of welfare applicants and recipients. Unfortunately, it is often hard to minimise abuse without an intrusive monitoring system.

As Creedy (2010, p. 104) observes, it is hard to obtain 'a precise specification of attitudes towards the basic aims of a tax and transfer system'. There is a multiplicity of welfare objectives. Each one involves normative judgements.

Three further observations may be made. First, there is the issue of the time period. Some people may be poor over a short period but well-off in the long run (over the life cycle). This applies to many tertiary level students and to some parents of young children. In this case the policy objective may become income smoothing rather than poverty relief. Second, in so far as people have income earning capabilities but choose a low income, they may be deemed not to need income support. Third, implementation of social security programs may require considerable information on personal behaviour. Thus, the contributions of individual programs to meeting these policy objectives (once agreed) may be hard to quantify. Nevertheless, these objectives provide some useful guides for policy making.

---

[1] Poverty relief programs may also be viewed as self-interested insurance against severe accidents that may happen to anyone. This is similar logic to Rawls' argument that society should maximise the welfare of the least well-off person because, *ab initio*, anyone could be in that position (see Chapter 7).

[2] The US literature typically refers to these disincentives as "moral hazards".

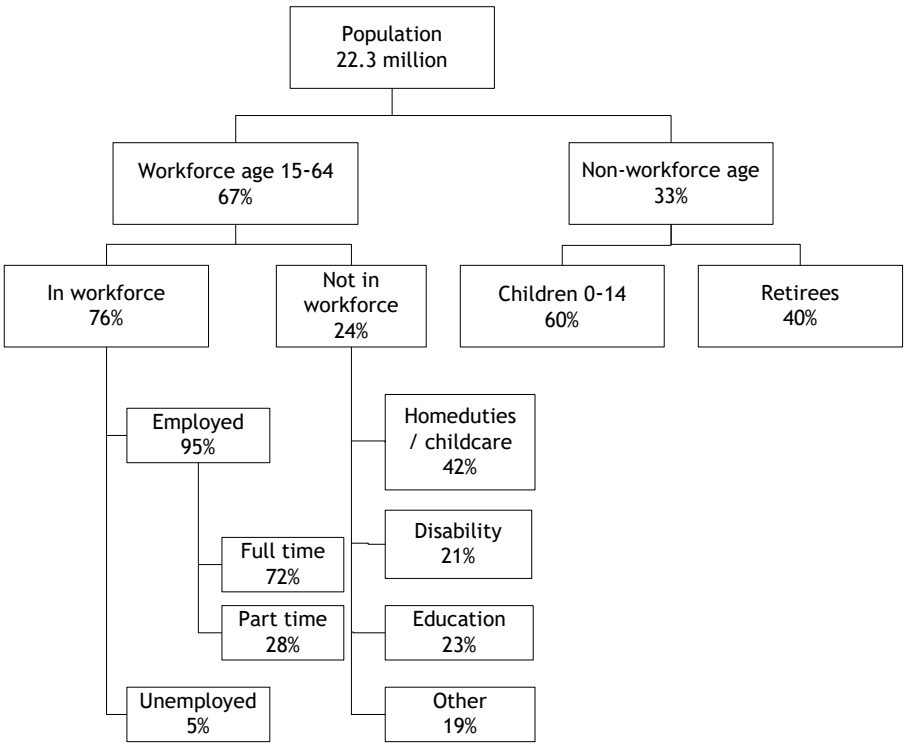
Social Welfare Provision in Australia

To provide a perspective on social programs, Figure 22.1 shows the major social groups in the Australian population in mid-2010. Of the then population of 22.3 million, nearly a quarter received income support (not including family support payments). Three main population groups received support:

- Individuals of non-workforce age (0 to 14 and 65 and older), especially those in low-income households. These make up a third of the population, but not all are poor.
- Individuals of workforce age (16 to 64) who are in the labour market but who are unemployed or working part-time for a low income. These are about 10 per cent of the total population.
- Individuals of workforce age who are outside the labour market, including parents with home care duties, individuals with a disability and students. These groups constitute about 16 per cent of the population, but again not all are poor.

There are also separate welfare programs for most sub-groups with needs.

In Australia, cash benefits to individuals are the basis of the welfare system. The Commonwealth and state governments also provide various benefits-in-kind, notably health care and housing services. In addition, employers are required to provide various welfare support programs for employees.[3]



Sources: ABS (2010) *Australian Social Trends*, Cat. No. 4102.0 and author estimates drawing on other sources.

Figure 22.1 Population groups in Australia in mid-2010

[3] For an informative review of the US welfare system, see Martin (2011).

Table 22.1 Commonwealth welfare payments by major category, 2010–11

Payments to/for:	*$bn*	*$bn*	*No. recipients (2010)*	*Comments*
Aged persons	44.0			
Age pensions		31.9	2 158 303	
Other payments to the aged		12.1		
Veterans and dependants	7.0		256 826	
People with disabilities	20.3		792 581	
Families with children	30.4			
Family tax benefits A and B		17.8	2 151 900	
Other family benefits		12.7	1 537 560	Includes parenting and baby bonus payments and childcare benefits
Unemployed and sick persons	7.0		642 422	Includes full-time students aged 16 to 24 and job seekers under 21
Other welfare programs	1.3			
Aboriginal advancement	1.5			
General administration	3.5			
Total	115.0			

Sources: Treasurer, *2010–11 Budget Paper No. 1, Statement 6*; Department of Families Housing, Community Services and Indigenous Affairs (2011) Statistical Paper No. 9, *Income Support Customers: A Statistical Overview, 2010*.

Table 22.1 shows the major Australian government cash transfers in 2010–11 along with the number of recipients in 2010. The largest programs, in terms of dollars and people, are pensions for aged persons and persons with disabilities; assistance to families; and unemployment and youth allowances. A common feature of most income recipients is lack of employment income (excepting families receiving family tax benefits Part A). Pensions to persons outside the workforce are generally higher than allowances to persons inside the workforce.

Australian assistance programs have three main features. First, most welfare payments are funded from consolidated revenue rather than from a mandated insurance fund. Unemployment and retirement are treated as welfare issues requiring social assistance rather than as social insurance. However, there are mandatory employer requirements for contributions to superannuation and to workers compensation for injury. Second, benefits are generally targeted to those in need rather than provided universally. Most benefits are based on categories of people deemed to be in need and means tested. However, the means tests for some family benefits are generous. Thirdly, benefits are paid at flat rates; they are not related to the recipient's prior income or contributions. Box 22.1 overleaf highlights these and other features of Australian assistance programs.

As summarised by Whiteford (2010), these system design features have some striking results. Overall Australia spends a small proportion of GDP on cash transfers compared with other OECD countries. But it provides the highest share of spending on income-tested programs and it is the most redistributive welfare system. In 2005, in OECD countries the average ratio of transfers received by the poorest quintile compared with the richest was 2:1. In Australia the ratio was 12.4:1. In absolute terms, Australian redistributed more to the poorest 20 per cent of the population than any other OECD country than Denmark.

Targeting welfare payments ensures that those with greatest need get most support. The main disadvantage of targeting is the high implicit marginal tax rates that occur when benefits are withdrawn as income is earned. This can create adverse incentive effects. Also, targeting requires subjective judgements of need and intrusive administration.

Box 22.1 Features of Australian social assistance programs

- Australian social assistance is a single nationwide system financed and run by the Commonwealth government.
- Most payments are funded from consolidated revenue.
- Most benefits are targeted to those in need rather than provided universally. Most programs are means tested for income and assets. For example, family payments and pensions are means tested.
- Many benefits are based on categories of people. The support granted varies according to the person's current personal circumstances such as age, marital status, parental status and employment.
- Most programs provide cash benefits, which are paid to individuals who satisfy the stipulated eligibility criteria.
- Benefit recipients may also receive concession cards that may be quite valuable, for example for medicines or public transport.
- Benefits are paid at flat rates; they are not related to the recipient's prior income. This means that unemployment benefit rates are low relative to many other countries.
- In addition to direct payments, minor assistance is provided indirectly through the personal income tax system via tax offsets for persons with dependants or mature age workers.

The major in-kind transfers in Australia are health care services, education and housing. Although the health services provided through Medicare are sometimes described as a national health insurance system, the free hospital and medical services are funded through consolidated revenue and provided to everyone independently of contributions (see Chapter 24). For housing assistance, the Commonwealth provides rent assistance payments and, with the states, subsidises public and community housing. The states also subsidise some other services, usually public transport and sometimes utilities.

In addition, governments at central and state level require the private sector to provide various benefits. Examples include:

- National compulsory occupational superannuation, whereby in 2017/18 all employers must contribute a minimum of 9.5 per cent of an employee's earnings to a registered superannuation fund.
- Compensation arrangements for work injuries and deaths which provide no-fault earnings-related benefits. These are financed by compulsory risk-related premiums paid by employers to commercial insurers or state government funds.
- Compensation for road accident injuries and deaths financed by compulsory levies on vehicle owners paid either to commercial or state insurers.

Cash Transfers

Categorical transfer
A transfer in cash or kind that depends on the category of the recipient

Non-categorical transfer
A transfer that is independent of the category of the recipient

Most OECD countries have some form of safety net system that provides cash transfers to low-income individuals, generally funded from consolidated revenue and independent of taxes paid or of contributions to an insurance fund. Such a system aims to ensure that individuals receive at least a minimum income. The benefits would be available to all adults, subject to means testing. The benefits may also depend on household composition. Such benefits are typically between one-third and one-half of the median income level in the country. Other cash transfers, for example family benefits and child allowances, are targeted at specific groups and are not universal.

Cash transfers may be categorical or non-categorical. A transfer is **categorical** if it depends on the recipient belonging to a specific social category, for example mothers, single parents, disabled persons and so on. A transfer is **non-categorical** if it depends only on the income of the recipient and not on his or her social category. Most benefit transfers involve some categorisation, even if it is simply age for aged pensions, whereas taxation is often non-categorical. We start below by discussing non-categorical transfers and then turn to categorical transfers.

Negative income tax

Negative income tax system
A transfer system in which individuals may pay tax or receive grants from government

The most general non-categorical transfer method would be a non-categorical **negative income tax** (NIT) system. With a NIT system, individuals may receive income (grants) from government or pay taxes. NIT is non-categorical when the grants and taxes are based on the gross private income of individuals regardless of their household status or of any other social category to which they may belong. However, even with an NIT system, transfers can be conditional on a person's social category.

Figure 22.2 illustrates how NIT can work. In both panels gross private income is measured along the horizontal axis and net income on the vertical axis. Net income equals gross private income plus or minus government cash transfers. With no transfer, net income equals gross private income. This case is represented by the 45° line, *OXB*. In panel (a) government provides a minimum grant of *G* to each individual. If Amy's private income is less than *L*, she receives the difference between *G* and her private income. But note that her private income is effectively taxed at 100 per cent. When her private income exceeds *L*, she pays tax. The line *GXA* represents her net income after cash transfers. Panel (b) shows a variation. Here, government provides a minimum income shown by *G'*. However, as Amy earns some income, she can keep part of it. In this panel the line *G'XA* represents net income. This form of NIT provides some encouragement to grant recipients to earn income.

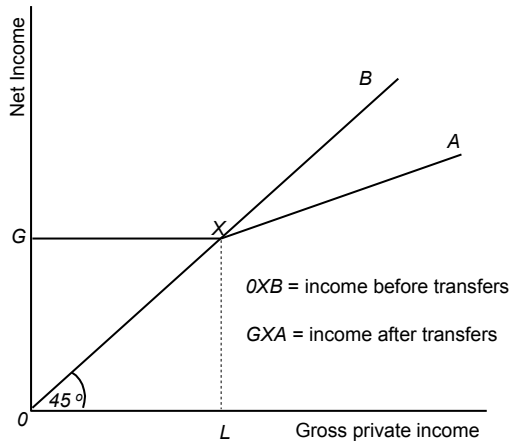
Effective marginal tax rate
The percentage of an extra dollar that an individual loses due to tax and loss of grants

Of special importance is the **effective marginal tax rate** (EMTR). The EMTR is the percentage of an additional dollar of private income that an individual loses due to the combination of income tax and the withdrawal of benefits.

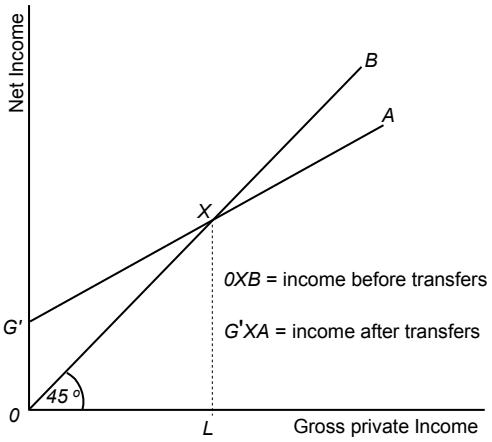
$$EMTR = [1 - (\Delta NY/\Delta PY)] \times 100 \qquad (22.1)$$

where *NY* is net income, *PY* is private income and Δ is the change in the variables. If *PY* increases by a dollar, but *NY* does not change, EMTR is 100 per cent. This is the case along GX in panel (a). There is effectively a 100 per cent tax on any earnings below *L*.

On the other hand, if the slopes of the *NY* and *PY* curves are equal, EMTR is zero. Generally, the flatter the net income schedule, the higher is EMTR. In panel (a), EMTR varies with income. In panel (b), EMTR is constant. In practice, EMTR is rarely constant. Indeed, it is often higher at low levels of income because of the combination of withdrawal of benefits with income tax payments.



(a) NIT with a fixed minimum grant



(b) NIT with retention of some private income at low income levels

Figure 22.2 Negative income tax: two structures

Table 22.2 EMTRs in a hypothetical negative income tax scheme

Private income (PI) ($)	*Government grant = 7000 - 0.5 (PI - 7000)[a] ($)*	*Government tax on PI = 0.2 ($25 000 - $10 000) + 0.33 ($50 000 - $25 000) ($)*	*Net income = PI + grant - tax ($)*	*Nominal tax rate (%)*	*Effective marginal tax rate (%)*
0-7000	7000	0	7 000	0	100
10 000	5500	0	15 500	0	50
20 000	500	2000	18 500	20	70
30 000	0	4650	25 350	33	33
40 000	0	7983	32 017	33	33

(a) In this example the government grant falls as private income (*PI*) exceeds $7000.

Table 22.2 provides a hypothetical example of an NIT with a non-linear EMTR. Government provides a minimum grant of $7000 to each adult but reduces the grant by 50 cents for each dollar earned above $7000. There is no tax on income below $10 000 but government taxes incomes between $10 000 and $25 000 at a MTR of 20 per cent and higher incomes at a MTR of 33 per cent. However, EMTR equals the tax paid plus the grant lost as a proportion of an additional dollar of income. At private incomes below $7000, EMTR is 100 per cent. At an income of $10 000, it is 50 per cent. At an income of $20 000 it rises to 70 per cent. EMTR then falls at higher income levels.

A major advantage of NIT is the potential to integrate welfare benefits with taxes. This should improve the transparency and consistency of the welfare system. Eligibility requirements are simplified. In a non-categorical NIT system, benefits do not depend on employment status, pensionable age or any other social category. Administration costs are low with large savings in the costs of running multiple benefit programs.

However, a non-categorical NIT has two main disadvantages. First, it does not allow for any needs of recipients other than low income or for differences in household composition. Consequently, non-categorical tax-transfer systems may not target aid to individuals with greatest needs. An NIT system can allow for individual needs. But, if this is done, the NIT system becomes categorical and more complex.

Second, a full non-categorical NIT system requires a high MTR. Dawkins *et al*. (1998) estimated that, with even a low basic grant, a revenue-neutral system in Australia would require a linear MTR of at least 45 per cent. This would have high leisure incentive effects and DWL and would be unlikely to be optimal. The EMTR could be reduced by withholding credits for people who are not working or by introducing categorical benefits. It could also be lowered for low-income individuals at the expense of increasing it for others. However, these modifications would introduce categorisation and complexity without eliminating high EMTRs for many taxpayers.

Earned income tax credits

Earned income tax credit
A tax credit per dollar of earned income grants

With the NIT system discussed above, when Amy earns income, she loses part of her grant. In contrast, with an **earned income tax credit** (EITC), a worker on a low income receives an income supplement in the form of a tax credit per dollar of earnings. When the credit exceeds taxes owed, the worker receives a payment. Box 22.2 outlines the EITC system employed in the United States.

The basic form of an EITC, with no supplementary grant for non-income earners, is shown in Figure 22.3a. Note that in this figure the *x*-axis shows *earned* income instead of total private income. The *OVWXA* line represents income after transfers. In the *OV* segment, the

Box 22.2 The earned income tax credit system in the United States

The earned income tax credit is a large anti-poverty program in the United States. The program has three ranges: the subsidy, flat and phase-out range.

In the subsidy range, each dollar of income is combined with a tax credit. In 2011, for a single earner with two children, the government paid 40 cents per dollar earned up to a total subsidy of $5285 on earnings of $12 570.

In the flat range (from $12 570 to $16 400) extra earnings do not affect the amount of credit received.

In the phase-out range from $16 400 to $40 295 the credit is reduced by 21 cents per dollar earned.

The EITC is categorical: it varies with the number of children and family income.

To receive credits, it is necessary to file tax returns and provide any extra information required, for example about qualifying children.

For an update to 2015 and further analysis, see Gruber (2016, pp. 666-673).

Figure 22.3 Earned income tax credit schemes

worker receives a subsidy. This line rises more steeply than the *OXB* line. From *V* to *W*, the tax credit is constant, so *VW* is parallel to *OXB*. From *W* to *A*, there is a net income tax.

A major aim of the EITC, in addition to poverty alleviation, is to encourage employment. Findings on this are mixed. Studies in California indicate that the EITC increased employment among some groups, for example the low skilled and single women (Hotz *et al*., 2002; Eissa and Liebman, 1996). More recently, Gruber (2016, p.536) concluded that overall the program has been "very successful in terms of increasing the labour supply of low-income workers". However, EMTRs are high in the phase-out stage of the tax credit as each additional dollar earned is both taxed and results in a reduction in the EITC. Thus, EITCs could increase part-time employment, but not full-time employment. Dickert *et al*. (1995) found that the EITC discourages employment among married females.

Some other issues may be noted. First, consistent with our analysis of tax incidence in Chapter 26 below, workers may not get all the benefits of the EITC. Unless the demand for labour is perfectly elastic, the EITC will result partly in a fall in wages. Leigh (2010b) estimated that a 10 per cent rise in the EITC in the United States would lead to a 5 per cent fall in the wages of high school drop-outs and a 2 per cent fall in the wage of those with only a high school diploma.

Second, the EITC is presumably intended to assist workers with a low hourly wage rate rather than those who choose to work only a few hours at relatively high hourly rates. In the UK EITC scheme, there is a minimum hours of work requirement. But it is hard to monitor whether low incomes reflect low wage rates or low hours and the UK scheme has been criticised for its administrative complexity.[4] In effect, EITCs are sometimes attached to low incomes rather than to low wage rates.

Third, EITCs do not assist non-working individuals, who are a high proportion of the low-income population. A comprehensive poverty relief program would require combining EITCs with a basic grant for all adults (see Figure 22.3) or a parallel income support scheme for non-workers. A combined EITC/grant scheme would be more generous for workers than an NIT scheme, because low income workers would obtain extra benefits from the EITC. Accordingly, EMTRs would be higher as both grants and subsidies are withdrawn. On the other hand, running a separate welfare scheme for workers and non-workers may create anomalies and administrative complications.

Low income tax offsets

Low income tax offset
An increase in the threshold at which low earners are taxed

Another component of the redistribution system in Australia is the **low income tax offset** (LITO). The LITO is an amount subtracted from an individual's taxable income. In 2016/17, the LITO provided a tax offset of $445 to all tax payers with a taxable income between $20,542 and $37 000. Adult tax payers eligible for the full LITO pay no tax until their income exceeds $20,542. Taxpayers with a taxable income between $20,542 and $37 000 pay a MTR of 19 per cent which is offset partly by the LITO. The LITO was withdrawn at 1.5 cents in the dollar until it eroded entirely as $66,667. This means that persons with a taxable income of $37 000 to $63 750 face an EMTR which is 1.5 per cent higher than the nominal MTR.

LITO has two main aims. One is to increase the progressivity of the tax system by reducing average tax rates for low-income individuals. The other is to maintain the tax revenue from medium and high-income individuals by retaining the low threshold for them. It succeeds in doing this without increasing MTRs for high income individuals. However, there is a small cost. As at 2016/17, this was the 1.5 per cent increase in EMTRs for anyone with an income between $37 000 and $66 667 as the tax offset benefit tapers off and is withdrawn.

Categorical benefit systems

In a categorical welfare system benefits vary by social group as well as by income. The aim is to target income as accurately as possible to those most in need. If this is achieved, each welfare dollar would provide maximum welfare benefit and not be expended on low priorities or, worse, on households not in need. Accurate targeting minimises the total revenue required for income transfers and lowers MTRs.

However, a categorical welfare system can have significant disadvantages unless the categories are clear and simple. A complex system with numerous categories has high administrative costs. The impacts may not be transparent or clearly understood. And the scheme may not satisfy the principle of horizontal equity—that poor persons who are similar in important respects to other poor persons are treated equally. Importantly, a highly targeted system requires high EMTRs as benefits are withdrawn from those who no longer need them.

Traditionally Australia has been a leader in targeting welfare. Mitchell *et al*. (1994) reported that there were over 30 separate tax transfer programs. Each program had its own rules for eligibility, differing means tests and varying rules for contingencies such as changing personal circumstances. There were also many differences within programs. For example, there were four payment categories for unemployed single persons aged over 18

[4] EITCs are inexpensive to administer in the United States where most workers file their tax returns. Administration is more complex with pay-as-you-go tax systems.

without dependants, with further differentiation for persons between 18 and 20 years, depending on whether they live at home or not and for the over 21 age group by age and length of time on Newstart. Recognising the administrative and other costs, the Australian government has reduced the number of benefit categories in recent years.

Concluding points

Most social assistance systems are based on cash payments and involve significant categorisation of benefits. Categorical payments can target those most in need and meet vertical and horizontal equity objectives with lower revenue requirements than non-categorical systems. However, to provide adequate security to all in need, most social assistance systems require substantial revenue and involve high MTRs. Targeted categorical benefits reduce the total tax revenue required but create high EMTRs as benefits are withdrawn and tax rises with income. These EMTRs may well affect labour supply decisions.

Applying January 2007 income tax and social security rules to 2002–03 personal incomes Kalb (2007) estimated that nearly 30 per cent of Australian taxpayers faced an EMTR of over 40 per cent, 16 per cent faced an EMTR of over 50 per cent and more than 5 per cent faced an EMTR of over 70 per cent. Harding *et al*. (2009) reported slightly lower EMTRs for 2006–07. They estimated that 18.4 per cent of working Australians faced an EMTR over 40 per cent and that 9.5 per cent faced an EMTR over 50 per cent. They also reported a wide range of estimated EMTRs for 10 European countries. In 2003 over 20 per cent of workers faced EMTRs greater than 50 per cent in Belgium, Finland and Germany, whereas less than 2 per cent of workers faced an EMTR over 50 per cent in Austria, Greece and Spain.

The other major problem with cash transfers is that taxpayers often prefer to provide transfers-in-kind, especially services such as housing that may have external benefits, rather than income transfer that in their view may be misspent.

Transfers-in-Kind

In some countries, transfers-in-kind represent a major alternative to untied cash transfers. In the United States, the purchase of health care services, through Medicaid, constitutes 50 per cent of all welfare benefits; purchases of food, housing and energy account for another 20 per cent of benefits (Stiglitz, 2000a). Indeed, the food program is larger than the major US welfare program (Temporary Assistance for Needy Families, TANF) which is provided on a temporary and provisional basis to low income families (see Rosen and Gayer, 2014; Gruber, 2016). Note, also, that in the United States retirement income payments are provided by social insurance rather than under a welfare system, and unemployment insurance is funded via payroll taxes in state-based systems.

There are three main arguments for in-kind transfers: merit goods (paternalism), positive externalities (interdependent preferences) and commodity egalitarianism (see Currie and Gahvari, 2008). In-kind transfers may increase the consumption of goods regarded as merit goods and so raise the recipient's welfare more than would an equivalent untied cash transfer. In-kind transfers may also provide positive externality benefits to taxpayers. This could occur because recipients increase expenditure on goods such as food or housing rather than on alcohol or drugs that are associated with anti-social behaviour. Third, in-kind transfers are consistent with the notion of commodity egalitarianism. This is the idea that some basic commodities should be made available to everyone. There is also a political factor. Public officials may favour transfers-in-kind because they require more administration and control than do cash transfers.

We now consider the implications of two scenarios: one in which a tied transfer changes Amy's consumption basket and one in which Amy's consumption is unchanged. Figure 30.4

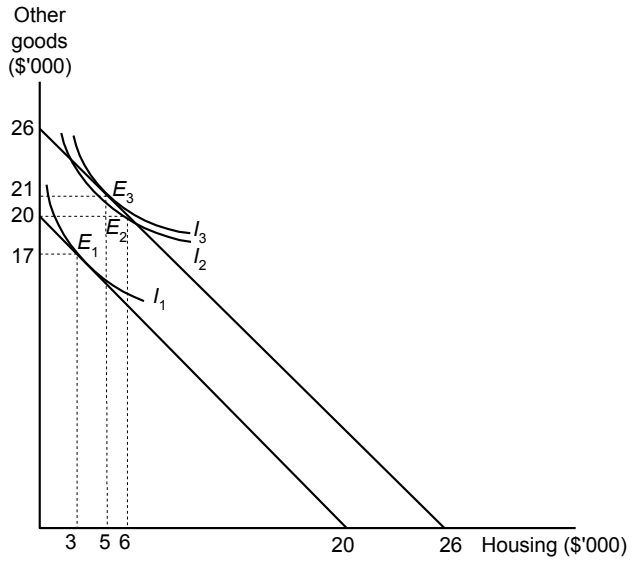


Figure 22.4 Welfare implications of transfers-in-kind

depicts Amy's income constraints and her preferences between housing and other goods embodied in indifference curves. Amy starts with an income of $20 000 and chooses to spend $3000 on housing and $17 000 on other goods. Now suppose that government provides Amy with housing valued at $6000 per annum and that Amy spends her own income of $20 000 on other goods. This is depicted by point E_2, which sits on indifference curve I_2. However, if government gives Amy $6000 in cash, her new untied budget constraint is $26 000. Amy now consumes $5000 of housing and $21 000 of other goods, which is point E_3 on a higher indifference curve I_3. In general, if Amy makes rational decisions about her own welfare, she cannot be better off with a tied transfer than with the income equivalent. On the other hand, if an in-kind transfer changes her consumption basket compared with what she would have chosen with an untied cash transfer, she obtains less welfare.

Now suppose that the transfer-in-kind does not change Amy's expenditure pattern. If government provides Amy with $3000 to be spent on housing and she is already spending $3000 or more on housing, the transfer-in kind becomes a *de facto* income grant, effectively increasing Amy's income to $23 000.

In **summary**, if transfers-in-kind do not change the consumption patterns of recipients, they serve no purpose compared with cash transfers. Also, they usually involve significant administration costs. On the other hand, if consumption patterns are changed, the welfare of recipients is potentially reduced. However, tied transfers-in-kind may be justified by merit good considerations or positive externalities.

Social Insurance

Social insurance funds are generally schemes that provide benefits to individuals based on (1) previous contributions to the insurance fund and (2) the occurrence of a specified contingency such as unemployment, sickness or retirement. Note that the beneficiary may be different from the contributor (as with some health insurance schemes). These schemes may be

managed by government or by the private sector subject to the government's rules and regulations.

Because benefits are related generally to past contributions, but not to current income, they are not limited to low-income workers. Social insurance funds are generally expected to be self-funded, with contributions funding future withdrawals. However, public insurance funds that in principle guarantee a defined benefit or pension are often in actuarial deficit and withdrawals may have to be funded from current contributions rather than past ones.

Most OECD countries have social insurance schemes covering unemployment, disability and retirement. An unemployment benefit program replaces part of the earnings lost due to involuntary loss of a job. The benefit is usually between 30 and 80 per cent of lost earnings. Periods of eligibility for employment benefits vary but can exceed a year. A disability program provides income to workers who become handicapped during their working age years, either on or off the job. Receipt of these benefits is usually of long duration. Retirement programs provide pensions on retirement from the workforce. These pensions are as high as 70 per cent of lost earnings in some countries.

The rationale for public insurance programs (or for mandated private insurance) is based on the familiar two main grounds: the market failures and inequity of market provision of insurance. As we saw in Chapter 4, insurance markets suffer from adverse selection and moral hazard problems. If insurers cannot identify who is likely to be unemployed or sick, they have to offer average premiums. These premiums are unattractive to those with low risk of unemployment or sickness, so the insurance market is incomplete. On the other hand, if insurers can identify individuals who are most at risk, they will charge them high premiums that are inconsistent with the welfare purpose of insurance. Moreover, private firms will not provide insurance for large-scale unemployment that may be created in part by government fiscal actions.

Social insurance differs from private insurance in two fundamental respects. First, social insurance usually has a redistributive component. It does not necessarily provide an actuarially fair return to each individual—that is, a fair matching of contributions and benefits. Lower paid workers may get defined minimum payments that exceed the value of their contributions. Also, government may vary the terms of the benefits to cover unanticipated needs. In effect, although there is a conceptual distinction between government acting in response to failures in insurance markets and provision of social assistance to alleviate poverty, the distinction between social insurance and assistance is often fuzzy. Most public insurance schemes, including those in the United States and the UK, include significant social net components and are not solely actuarial schemes.

Second, unlike private insurance, a public insurance program may be funded on a pay-as-you-go basis rather than be fully funded by contributions in advance. This is also redistributive, in this case from future generations to present ones. For further discussion of financing alternatives see Chapter 23.

The different approaches to social welfare (social insurance and social assistance) in different countries are mirrored in the literature. Much economic literature (especially US literature) analyses unemployment benefits and pensions, as well as health care, primarily in terms of the social insurance that is needed to offset failures in insurance markets rather than as welfare payments.[5] On the other hand, in this book we view welfare, like most Australians, primarily as social assistance rather than as social insurance. If the social insurance provisions are inadequate, safety-net expenditure programs to assist the poor are then also required.

---

[5] Gruber (2016, pp.4-5) notes that in 2010, before the *Affordable Care Act* in the US, there were 49 million persons (18.5% of the non-elderly) who did not hold health insurance. He observed that this did not imply a problem in the market. "It just implies that those without insurance don't value it enough to buy it at existing prices". He acknowledges that there may be negative externalities that would justify some government intervention. But this is not described as a problem of poverty or as a social welfare issue.

Incentive Effects of Social Welfare Programs

Virtually all welfare programs have incentive effects of some kind that distort behaviour and create a deadweight loss (DWL). Most welfare benefits are means tested, which creates an incentive for the potential recipient to choose leisure instead of earning income. In the US literature (for example Rosen and Gayer, 2014), this incentive is typically described as a moral hazard. **Moral hazard** occurs when the provision of a benefit against an outcome, for example unemployment or retirement, increases the likelihood of the outcome occurring.

Moral hazard
The risk, when a payment depends on certain conditions, that an individual will change their behaviour to make the conditions more likely to occur

The two main sources of incentives are the EMTR and the replacement rate. The EMTR was defined in Equation 22.1 and discussed above under Cash Transfers, where we also noted some high EMTRs in Australia and elsewhere. The **net replacement rate** (NRR) compares the net income from social assistance with after-tax income that it replaces:

$$\text{NRR} = \text{social security payment after tax/after-tax income replaced} \qquad (22.2)$$

Net replacement rate
The ratio of a social security payment after tax to the after-tax income replaced

If Amy can earn $1000 a week gross and $800 after tax and can receive a pension of $600 a week after tax, the NRR would be 75 per cent.

Similar calculations apply to marginal work choices. Suppose that Amy is willing to work for $15 an hour. If she can earn $20 an hour after tax but social assistance would provide her with $12 an hour after tax, the NRR is 60 per cent. Amy's additional consumption for an hour's work is only $8, which is not enough to induce her to work. As the NRR increases, the incentive to work falls.

Figure 22.5 illustrates the effect of social assistance on labour supply. The *OCB* line shows the initial budget constraint, that is, how the consumption of goods may rise with hours of work assuming a zero or constant MTR. The indifference (*I*) curves show the hours of leisure and the consumption of market goods between which Amy is indifferent. A loss of leisure has to be compensated (increasingly) by extra goods. Given the initial budget constraint, Amy maximises her welfare by choosing to work H_1 hours.

When social assistance (*OA*) becomes available, Amy's budget constraint changes to *ACB*. Note that the slope of the budget line between *A* and *C* is flatter than the original budget line. This indicates that the EMTR has risen and that Amy obtains fewer goods for each extra hour of work. This is because she loses some assistance income as she earns private income. In

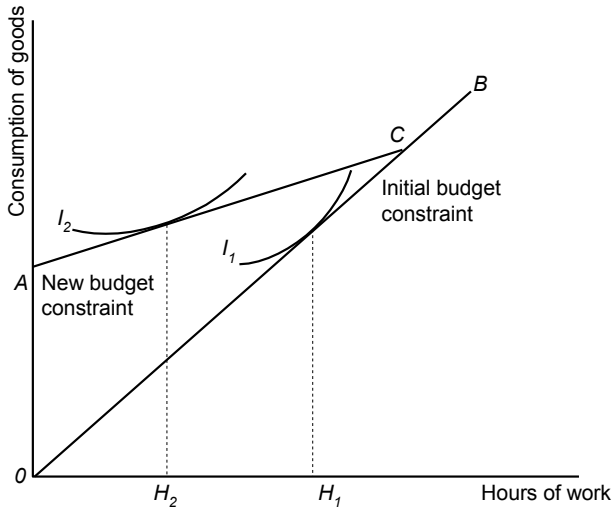


Figure 22.5 Effect of social assistance on labour supply

other words, the price of leisure has fallen. Amy now chooses H_2 hours of work. Of course, with different preferences, Amy might now choose to do no work. Note also that the income effect works here in the same direction as the substitution effect. The social assistance increases Amy's real income. Because leisure is a normal good, Amy demands more of it.

Replacement rates for unemployment. As an example of NRRs, Table 22.3 shows NRRs for unemployment for four family types in relation to the average wage in 2005 in 17 OECD countries. In 10 of these countries NRRs in all categories exceed 60 per cent. In other words, working increases disposable income by less than 40 per cent. This is a major disincentive to work. NRRs are generally lower in Australia than in all other countries because most benefits are flat-rate safety-net benefits not related to income, whereas benefits in most other countries are based on social security contributions and relate ultimately to income. On the other hand, unemployment benefits are of unlimited duration in Australia (subject to continuing search for employment) whereas the duration is limited in some other countries.

Estimating incentive effects

So much for theory, but how are incentive effects estimated? The key problem is to identify cause and effect and to distinguish this from correlations. There are three basic empirical methods known as observational, experimental and quasi-experimental. We discuss each of these below.

Observational studies
Studies based on data from observed economic behaviour

Observational studies. Observational studies are based on observed data of actual economic behaviour. Much of this data is obtained by surveys. Observed data are natural data and not data obtained from experiments. The data sets may be time series, cross-sectional or a combination of time series and cross-sectional data. Econometric methods (the application of

Table 22.3 Net replacement rates for unemployment for four family types, 2005, in selected countries[a]

	Single	*Married couple*[b]	*Couple-two children*	*Lone parent-two children*
Australia	33	29	65	54
Canada	63	66	78	77
Denmark	63	64	75	78
France	67	66	67	67
Germany	60	60	73	73
Ireland	31	49	59	58
Italy	63	69	70	71
Japan	54	53	53	54
Korea	48	48	48	49
Netherlands	65	65	70	70
New Zealand	38	33	46	63
Norway	64	65	70	77
Spain	62	63	75	76
Sweden	62	62	69	67
Switzerland	70	71	85	85
United Kingdom	41	41	60	60
United States	62	62	56	58

(a) Income after tax and including unemployment benefits, family and housing benefits in the first month of benefit receipt, relative to the average wage.

(b) Married couple with only one earner.

Source: OECD (2007) *Benefits and Wages, OECD Indicators*.

statistics to economic data), especially regression analysis, are employed to identify relationships, especially between the dependent and independent variables. These methods are generally guided by economic reasoning or theory.

Time series analysis typically draws on aggregate or average data over time. The variable to be explained could be labour force participation or unemployment. The explanatory variables would include the policy variable, such as the replacement rate, as well as other key explanators such as the change in GDP. In principle such an approach should provide estimates of the effect of a policy change. However, there are significant difficulties. The use of aggregate data necessitates use of average benefit levels or NRRs, although there are often many NRRs. There are also issues of cause and effect between the 'dependent' and the 'independent' variables (does unemployment drive GDP or the reverse?). Also, it is often difficult to isolate the impact of policy changes from other factors that may cause a change in the dependent variable.

Many empirical studies are based on cross-sectional data. Cross-sectional studies are usually based on individual data, which greatly increases the number of observations. However, care must be taken to identify the relationships between an individual's behaviour and their benefits and their personal characteristics. If the benefits are correlated with other individual characteristics, the precise effects need to be sorted out. In contribution-based schemes, variations in benefit entitlements are often correlated with previous income, which may be correlated in turn with unobserved characteristics that affect labour supply decisions. Moreover, cross-sectional analysis models relative behaviour. It may tell us that people with high benefits act differently from those with low benefits. But this is still consistent with there being no overall effect when decisions are interdependent. The replacement rate may determine who is employed rather than the overall rate of employment.

Combined time series and cross-sectional data may draw on aggregate comparisons as between countries or between states within a country or on longitudinal data on individuals typically collected as panel data. In cross-country or cross-state studies the analyst explores the effects of national or state differences in benefits and institutional arrangements on labour supply or other behaviour. However, differences in other economic or social factors may also influence behaviour and be difficult to model precisely. With panel data, the aim is to infer how changes in individual circumstances or benefits influence individual decisions. Given the need for extensive longitudinal data sets, there have been few such studies although their use has increased greatly over the last 10 or so years.[6] As with other approaches, modelling all the critical factors and their inter-relationships is not easy.

Experimental studies
Studies based on a sample of individuals randomly assigned to a treatment and a control group

Experimental studies. A general problem with using observational data is their lack of randomness. This puts a lot of weight on the econometric analysis to sort out cause and effect. This problem is avoided in experimental studies where a sample of individuals is randomly assigned to a treatment and to a control group. Because on average the individuals in the two groups may be viewed as similar, any differences in outcomes can be attributed solely to the difference in treatment. Hotz *et al*. (2002) analysed the effect of a policy experiment in California in the 1990s where one-third of families randomly selected received about 15 per cent higher non-work (unemployment) benefits than other families. They found that there was about a 10 per cent increase in employment in families receiving the lower level of benefits, indicating an elasticity of employment relative to benefits of about –0.67. Zabel *et al*. (2010) reported on a Canadian randomised trial in which the treatment group of unemployed workers were given a subsidy that roughly doubled their pre-tax salary for three years if they found

[6] In Australia by far the largest household-based panel data set is generated by the Household Income and Labour Dynamics in Australian (HILDA) Survey conducted at Melbourne University. Starting in 2001, HILDA collects information annually about labour market and family dynamics from a panel of nearly 20 000 individuals and nearly 8000 households.

long-term full employment within 12 months. Four years later, the employment rate for the treatment group was approximately 25 per cent higher than for the non-treatment group.

While it is generally agreed that randomised experiments represent the 'gold standard' method of isolating and identifying cause and effect, there are several drawbacks. These include the need for a large sample size, problems associated with drop-outs or attrition that may not be random, making inferences from behaviour in experiments to longer-term behaviour, the costs of the experiments and the ethics of providing different levels of benefits or services to different groups in the community even if this is done randomly.

Quasi-experimental studies. These studies draw on observational data but identify situations in which individuals happen, due to external or natural circumstances, to be randomly assigned to different groups. These situations are sometimes described as natural experiments. There are several different quasi-experimental methods.

Quasi-experimental studies
Studies that draw on observational data where external or natural events randomly assign individuals to different groups

One method is **difference-in-difference analysis**. Suppose there are two jurisdictions *X* and *Y* and that a policy change (say a higher unemployment benefit) is applied in *X* and not in *Y*. The difference-in-difference method examines how, if at all, the *difference* in the unemployment rates has changed due to the policy change. This recognises that other factors may cause systemically different levels of unemployment in the two jurisdictions. But it does leave open the possibility that changes concurrent with the policy change could have affected the differences.

Instrumental variables analysis is another quasi-experimental study method. An oft-cited issue is estimating the impact of class size on child performance. The estimation problem arises because more education-motivated parents may send their children to smaller classes. Thus the analyst needs to find a variable which influences participation in smaller classes but which itself will not affect the educational outcome. Hoxby (2000) found that birth dates (which determine school entry) vary randomly over time so that class sizes in the United States vary randomly from year to year. She was then able to use these random annual differences in kindergarten class sizes within the same schools to test whether differences in class size affected educational outcomes.

Regression-discontinuity analysis is a third form of quasi-experiment study. This method estimates the impact of a welfare program by comparing outcomes for program participants who are just below the threshold level for participation with outcomes for non-participants who are just above the threshold but can otherwise be assumed to be similar to the program participants. Because both sets of individuals or households are close to the threshold, it is assumed that they are comparable individuals or households in most important respects.

Incentive effects: employment and household formation

The heavily studied issues are the effects of social welfare programs on labour supply and household formation. In a wide-ranging survey of the incentive effects of the US welfare system, Moffit (1992) concluded that there is:

> unequivocal evidence of effects on labor supply, participation in the welfare system, and some aspects of family structure. Mostly these effects arise for female heads of family, the major recipient group under the current system. The econometric studies show that labor supply is reduced by the AFDC and Food Stamp programs, that higher potential benefits induce greater participation in these programs, and that the programs affect family structure though usually weakly.

These conclusions are consistent with the findings of labour supply studies (Dandie and Mercante, 2007), which suggest that the compensated labour supply for males has a mean wage elasticity of about 0.3 and for females a mean elasticity of about 0.6 (although there is much variance). These estimates imply that, ignoring income effects, for every 10 per cent fall in net hourly wage, male hours worked fall by 3 per cent and female hours by 6 per cent.

US studies quoted by Rosen and Gayer (2014) suggest that for every $100 offered to a male-headed family, earning falls by $25–$30.[7] Meyer and Mok (2007) found that a 36 per cent increase in unemployment benefits for high income earners in New York state led to a large increase in unemployment insurance claims.

A labour supply issue that has attracted much empirical work is the relationship between unemployment duration and level of unemployment benefits. Røed and Zhang (2003) reported that estimated point elasticities range from –0.2 to –0.9, with a mean of about –0.6. Men are typically at the higher end of the scale and women at the lower end. An elasticity of –0.6 would mean that a 10 per cent reduction in benefits would reduce unemployment duration by 6 per cent. A 6 per cent reduction in duration implies that someone who would be unemployed for 26 weeks would now be unemployed for about 24.5 weeks. This suggests that unemployment duration is not very sensitive to the level of unemployment benefits.

The possible impacts of social assistance on long-term welfare dependency or family structures are also issues of concern. If welfare benefits depend on household family income, families may be better off if one parent leaves the family because this increases its eligibility for assistance. Many welfare recipients in the United States are unmarried mothers and the possible implication is that welfare assistance has encouraged births outside marriage. However, the incidence of out-of-marriage births does not seem to be much higher in the states that offer higher benefits. Rosen and Gayer (2014) found that the evidence on the effects of welfare on family structure is inconclusive.

Policy Implications

The core public policy problem is how to produce an equitable welfare program that meets social welfare objectives without significant economic distortions and DWL and at least cost to taxpayers. Highly targeted programs are efficient at alleviating poverty but have high EMTRs as benefits are phased out with increased income. A slow phase-out means that some higher income individuals with less need of support receive benefits. Also, lower EMTRs for one group mean higher EMTRs for others.

In general, it is impossible to redistribute income without creating some DWL. It is especially difficult to provide everyone with an adequate minimum income, still less with a generous minimum income, without adversely affecting labour supply.

In practice, three problems are common to most social welfare systems: complexity and high administration cost, inequity in the treatment of some people and DWLs due to high NRRs or EMTRs. Governments also experience information problems with income testing, which often results in intrusive welfare administration. A major finding of the Henry Tax Review (2010, Part One, Chapter 9) was that the transfer system in Australia is overly complex and can treat people of similar means differently. There is also significant evidence of the DWLs of welfare programs in Australia and elsewhere, some of which is cited in the next chapter.

Reform options

Most discussion of reforms of social welfare focuses on simplification, transparency, equity and lower MTRs. But views vary as to the best strategies for achieving these outcomes. We discuss below a unified welfare/tax reform strategy, policy reforms with separate welfare and tax systems, and some policies to deal with moral hazard that could be integrated with either main strategy. Box 22.3 summarises the major conclusions of the Henry Tax Review.

[7] For more discussion of labour supply elasticities see Chapter 27. For discussion of some other incentive effects see Chapter 23.

Box 22.3 Major welfare findings of the Henry Tax Review

The Henry Tax Review (2010) found that Australia has the most progressive transfer payment system in the OECD reflecting effective targeting of support to low income households. However, the Review found that the system is overly complex and can treat people of similar means differently and that clear work incentives should be built into the levels of income support payments.

The Review concluded that it is not practical to fully integrate the tax and transfer systems. However, within the transfer system payments should be streamlined into three main types: pensions for people who are not expected to work including retired persons and persons with disabilities; participation (unemployment) payments for people of working age who are expected to support themselves; and payments for full-time students. Pensioners would receive the highest basic income support and students the lowest as they can usually do some part-time work.

In addition, there would be income supplements for parents (Parent Supplement) and for children (Family Assistance) and assistance for the direct costs of renting housing.

The Review also found that there should be increased support for child care to encourage workforce participation.

On the other hand, the use of concessions or payments that are linked to the purchase or supply of particular goods or services and the related use of concession cards provided by all levels of government should be reviewed. These concessions are often regressive and the benefits create strong incentives to achieve eligibility for the concessions.

There should also be a comprehensive means test base that incorporates a deemed return on assets with all other income to replace the diverse set of separate income and asset tests.

Systemic reform. This generally means unifying the welfare and tax systems, with both taxes and transfers based on individual income but with some allowance for household composition. This attempts to reduce the problems that arise when tax is based on individual income and welfare payments on household income. However, basing welfare payments on individual income makes welfare payments less targeted when low-income persons have a high-income partner. This is expensive and inequitable. On the other hand, basing tax on household income can create high MTRs for a secondary income earner in a household. This may be both inequitable to the second earner and a significant deterrent to labour supply.

Most studies conclude that such systemic reform is not feasible. Dawkins *et al*. (1998) examined the creation of a unified tax and transfer system, using a negative income tax or a tax-credit system. This would involve rolling some or, in extreme form, all social assistance benefits into the tax system. All individuals could receive grants (or tax credits) that would vary with household composition and other important needs and tax would be levied on an individual's gross income only above a threshold. There could be an income tax surcharge to recoup tax credits as income rises. The study was limited in that it did not incorporate behavioural responses. However, it concluded that the tax rates associated with a radical negative income tax system would be unacceptably high. Studies by Dawkins *et al*. (2003) and Scutella (2004) reach similar conclusions. Scutella examined two flat tax systems combined with a basic untaxed grant that varied with household circumstances that reflected the current social security system. She found that it was hard to avoid MTRs over 50 per cent and a level of income redistribution that was likely to be unacceptable.

Reforming the separate tax and transfer systems. Ingles (1998) and Keating and Lambert (1998) proposed that each system should be rationalised internally into single structures and the relationships between them rationalised. The large number of transfer payment schemes in Australia would be reduced. Also overlapping separate means tests for various cash payments would be replaced. Arguably, most of the means tests could be consolidated by having two separate tests of entitlement—one test based on a family's income and another on its eligibility characteristics. The assumption behind such proposals is that the administrative and

efficiency benefits from simplification of the tax and transfer systems would offset any loss in equity from a reduction in categorisation.

Kalb (2007) reviewed five possible reform policies. They were (1) reducing the lowest income tax rate, (2) increasing the lowest income tax threshold, (3) increasing the LITO, (4) reducing withdrawal rates on social security payments and (5) introducing an EITC. Given that these policies may be introduced separately or in combination, there are a large number of options. Thus, conclusions cannot be simply drawn. Moreover, a change that reduces EMTRs for some individuals often increases it for others. For example, reducing withdrawal rates on social security payments for one group usually has the effect of increasing the rates for higher income individuals unless the payment is made universal. Subject to these caveats, Kalb found that that an extended LITO or an EITC system would be relatively equitable and beneficial for labour supply. Changing the tax threshold for everyone, or changing the bottom tax rates, are more expensive and less targeted because they benefit high as well as low income individuals.

Detailed policy design. Gruber (2016) describes various approaches to reduce moral hazard and particularly impacts on labour supply. The first is to employ categorical benefits whereby individuals who are least able to work are placed into a category with the highest level of benefit and those with greater ability to work are placed in a category with the lowest level of benefit. Gruber cites blind people and single mothers as examples of people who could be in the former category. In fact, the Australian welfare system, like some others, does distinguish between disability (an inability to do any work) and other types of unemployment. But as Gruber concedes, there are considerable practical problems in categorising people by their capacity to work or earn income more generally.

A second policy to reduce moral hazard is to place 'ordeals' in the way of obtaining benefits, for example by requiring people to undertake training when unemployed or to join queues to obtain food relief. These ordeals are expected to discourage those who are more able to work from seeking income benefits. Where these ordeals may provide benefits to the recipients, such as training, they may be acceptable. Where they make life for needy people even more miserable they are unlikely to be acceptable.

The third policy is to provide incentives to people to work. Key examples here are the earned income tax credit, increased child support, increased training opportunities and reducing the rate at which other benefits decline with earned income. However, any system that provides improved conditions for targeted low-income earners increases EMTRs as the benefits are withdrawn. Alternatively, the benefits must be provided more generally, which is less progressive and more expensive.

Thus, we conclude where we started—it is hard to produce an equitable and generous welfare program without significant economic distortions or high expense, or both.

Summary

- The objectives of social welfare programs include: support for basic living standards, general economic security, reduction in inequality, social integration, minimising economic distortions and good governance.
- There are three main categories of social welfare programs: social assistance through cash transfers, transfers-in-kind and social insurance.
- Cash transfers provide the core welfare support in most countries and are usually preferred by recipients. Transfers-in-kind may be preferred by taxpayers and may provide positive externalities.
- Social insurance aims to protect individuals against adverse changes in circumstances.
- Many benefits are categorical. That is, they are made to individuals in specific household or other categories. A non-categorical payment depends only on the income of the recipient.
- Targeted categorical payments ensure that needs are met at least cost. However, targeted systems are often complex, expensive to administer, create anomalies and involve high effective marginal tax rates as grants are withdrawn.
- A negative income tax (NIT) scheme would integrate income transfers with income taxes. This could simplify the transfer system and reduce administration costs. However, there are complex issues in combining transfer that are based on households with taxes that are based on individuals and an NIT scheme would likely involve high marginal tax rates.
- High replacement rates and effective marginal tax rates can have significant effects on behaviour, notably on labour supply and even on household formation.
- The three main ways to estimate these effects are via observational studies (using econometric methods such as regression analysis), experimental studies and various quasi-experimental methods.
- The core policy challenge is to provide generous and equitable assistance without creating undue incentives that lead to lower labour supply and deadweight losses. Most policy reform attempts to simplify transfers and to reduce effective marginal tax rates and opportunities for "moral hazard" behaviour.

Questions

1. Outline the difference between social insurance and social assistance welfare systems. What is the guiding rationale behind each system? Which system do you think is preferable?
2. Why is a non-categorical negative income tax likely to be more expensive than a categorical system of assistance to needy households?
3. Do in-kind transfers sometimes affect incentives to work? Why?
4. If a company supplies an employee with a company car and pays the annual operating costs to a total value of $6000 per annum, is the employee $6000 better off than without the car?
5. Suppose that government provides $500 each month to everyone with no income. Clare can work for $16 per hour. The benefit falls by $1 for every $4 that she earns. What does Clare receive in total if she works 20 hours? How many hours does she have to work to reduce her benefits to zero? How much does she then earn? Using a diagram, show the income–leisure trade-offs that Clare faces with and without the government grant. What kind of indifference curves would ensure that Clare participates in the labour market?
6. Should income tax rates on individuals be adjusted to allow for their family or other responsibilities? If so, on what basis should these responsibilities be determined?
7. What are the relative merits of social (public) insurance compared with mandated private insurance?
8. How do welfare programs affect labour supply decisions? What is the evidence that welfare programs affect labour supply?
9. What methods are available for estimating labour supply effects of transfer programs? What kinds of problems arise?
10. Compare the merits of an earned income tax credit and a low income tax offset for assisting low-income earners.
11. It has been observed that married men earn higher wages. What, if anything, can be inferred from this correlation?
12. What methods are available for reducing moral hazard? Which ones would you recommend and why?

Further Reading

Barr, N. (2004) *The Economics of the Welfare State*, 4th ed., Chapters 4, 5, 10, 11, Oxford University Press, Oxford.

Creedy, J. (2010) 'Tax-and-transfer tensions: Designing direct tax structures', *Australian Economic Review*, 43, 103–113.

Currie J. and Gahvari, F. (2008) 'Transfers in cash and in-kind: theory meets the data', *Journal of Economic Literature*, 46, 333–383.

Gruber, J. (2016) *Public Finance and Public Policy* 5th ed., Chapters 3, 12 and 17, Worth Publishers, New York.

Harding, A., Vu, Q.N., Payne, A. and Percival, R. (2009) 'Trends in effective marginal tax rates in Australia from 1996-97 to 2006-07', *Economic Record*, 271, 449–461.

Kalb, G. (2007) 'Interaction of the tax and social security systems in Australia: the effect on effective marginal tax rates', *Australian Economic Review*, 40, 186–193.

Kaplow, L. (2008) *The Theory of Taxation and Public Economics*, Chapter 7, Princeton University Press, NJ.

Levitt, S.D, and List, J.A. (2008*) Field Experiments in Economics: the Past, the Present, and the Future*, Working Paper 14356, National Bureau of Economic Research, Cambridge, MA.

Martin. M.C. (2011) 'Welfare reform in the U.S.: a policy overview analysis', *Poverty and Public Policy*, 3, 1–38.

Rosen, H.S. and Gayer, T. (2014) *Public Finance* 10th ed., Chapters 2, 11 and 13, McGraw-Hill, New York.

Whiteford, P. (2010) 'The Australian tax-transfer system: architecture and outcomes', *Economic Record*, 86, 528–544.

Chapter

23

Welfare Programs for the Workforce, Families and the Elderly

Accidents will happen in the best-regulated families.

Mr Micawber, in Charles Dickens, *David Copperfield*

Welfare for the Workforce ♦ Welfare for Families ♦ Welfare for the Elderly

In this chapter we discuss the major welfare programs for the three social groups that receive most support in Australia and most other OECD economies: persons of workforce age with little or no employment, families and the elderly (see Table 22.1). Although the three groups are quite different, similar issues arise in each case. These are the identification of needs, how to support those in need while minimising adverse incentives and inefficient outcomes, and the roles of government and the private sector.

Welfare for the Workforce

In this discussion of workforce welfare, we consider five programs: wage-rate subsidies for employing low-income earners, employment promotion policies, unemployment benefits, injury compensation and disability pensions.

Wage-rate subsidies for low-income workers

In previous chapters we have discussed three strategies to support low income workers: minimum wage regulation, earned income tax credits and the low-income tax offset. Wage-rate subsidies are another strategy. A **wage rate subsidy** involves payments to an employer for all or, more usually, part of the wage paid to low-wage employees. The aim is to increase both employment and wages.

This is usually achieved to some extent.[1] Figure 23.1 overleaf shows a simple scenario with a fixed (unit) dollar employment subsidy (s) per hour of labour employed. This raises the demand for labour from schedule D_1 to D_2. Given the upward sloping labour supply curve (S), labour employment rises from L_1 to L_2 and the take-home wage from w_1 to w_2. On the other hand, the cost to the employer falls from w_1 to $(w_2 - s)$. Thus, labour receives only part of the employment subsidy. The outcome depends on the relative elasticities of labour demand and supply. Employment and wages rise by more when labour demand is elastic. Employment rises when labour supply is elastic. Wages rise when labour supply in inelastic.

[1] A wage rate subsidy increases take-home wages and employment providing the substitution effect of the wage on labour supply is greater than the income effect (the labour supply curve is not vertical or backward bending).

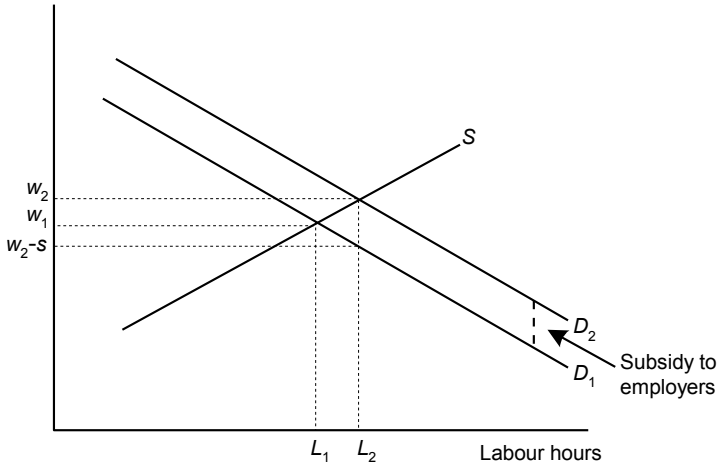


Figure 23.1 Effects of an employer wage-rate subsidy on wages and labour supply

As the discussion of tax incidence in Chapter 26 shows, the outcome is independent of whether the public subsidy goes to the employer or employee.[2]

However, wage-rate subsidies are not without complications. To be effective, subsidies must be targeted at marginal employment. Otherwise, the government simply pays part of the wage bill that a firm would incur in any case. Alternatively, to obtain the subsidy, employers may replace higher paid and more efficient workers by lower paid and less efficient workers.

OECD experience. Many OECD nations have adopted wage-rate subsidies to try to improve labour outcomes. In Canada, Germany, France and Spain employers have received wage subsidies for over a decade employing several hundred thousand young employees. In Australia, wage subsidies are available to encourage employment among vulnerable groups. Up to $10,000 is available to businesses hiring new employees between 15-25, or over 50 years of age and Indigenous Australians. Evaluations have found mixed results.

Studies in Ireland by Denny and Harmon (2000), the UK by Van Reenan (2001), and Sweden by Carling and Richardson (2001) found that wage subsidies increased youth employment at low cost. In Australia, Stromback and Dockery (2000) found that the temporary wage subsidy program in the mid-1990s had a high net job impact at a low cost per job created (see Table 23.1). An estimated 60 per cent of participants in wage subsidy programs were in unassisted employment three months after leaving the program compared with less than 20 per cent in the control group.

Other studies have found that wage subsidies result in subsidised workers displacing other workers and in deadweight losses. Martin and Grubb (2001) reported that studies in Belgium, Ireland and the Netherlands found that for every 100 jobs subsidised only 10 were net gains in employment. A DWL occurs when an efficient worker is displaced by a less efficient worker.

---

[2] Take simple labour demand and supply functions, $Q_d = a - bw$ and $Q_s = -c + dw$, where Q_d and Q_s are labour hours demanded and supplied, w is the wage rate and a, b, c and d are parameters. Assuming equilibrium, $w = (a + c)/(b + d)$. If employees receives a wage rate subsidy (s), the labour supply equation becomes $Q_s = -c + d(w + s)$. Solving out, the wage rate paid to the worker *before* the subsidy is received equals $(a + c - ds)/(b + d)$. If the employer receives the wage subsidy, the demand equation becomes $Q_d = a - b(w - s)$. Solving out, the *gross* wage paid by the employer to the employee equals $(a + c - bs)/(b + d)$. It can readily be shown that the *pre-subsidy* wage to the worker plus the subsidy is equal to the *gross wage* paid by the employer.

Table 23.1 Estimated net impacts of Australian labour market programs in mid-1990s (%)

		Net impact[a]				
Program	*Net unit cost ($)*	*1995/1*	*1995/2*	*1996/1*	*1996/2*	*1997*
Wage subsidy	1263	11.8	-0.3	38.3	42.8	44.3
Job creation	7105–10 000	8.5	13.6	15.3	13.7	19.0
Training	970–1170	3.7	1.3	2.9	3.6	1.4
Job assistance / search	625	8.7	1.4	4.5	7.7	8.0

(a) Percentage increase in unassisted employment 3 months after exit from program cf. control group.

Source: Stromback and Dockery (2000).

These studies found that the wage subsidies produced small net employment gains and that they determined mainly who was employed rather than how many people were employed.

Employment support policies

Unemployment may be frictional, structural or cyclical. Frictional unemployment occurs when workers are between jobs. Structural unemployment occurs when workers lack the skills or desire to work at market rates. Cyclical unemployment occurs when aggregate demand is low. We are concerned here with frictional and structural unemployment and policies to reduce these forms of unemployment. We examine training programs, job search assistance and job creation programs. [3]

Training programs include general traineeship and apprenticeship schemes for school leavers lacking workplace skills and special programs for persons with learning problems.

OECD (2000) cites considerable evidence that apprenticeship systems reduce unemployment. NCVER (2001) reported that 93 per cent of those who completed a New Apprenticeship in 2000 were in an unsubsidised job three months later. However, only a half of all those enrolled in apprenticeships complete the course and the drop-out rates for general traineeships are still higher (NCVER, 2010). In Australia, mid-2017, about 269,000 individuals hold apprenticeships or traineeships, a decrease of 4.7% from the previous year, usually involving on-the-job training part-time for two to four years, funded principally by the Commonwealth government, while apprentices also receive relatively low wages.

Most OECD countries also run special training programs for young persons who lack basic literacy and numeracy skills. Unfortunately, evaluations of these programs are not very positive. Heckman *et al*. (1999) found little evidence that such programs increased youth employment, especially for more disadvantaged youth. Martin and Grubb (2001) found that labour market training courses have little benefit for young people who leave school with inadequate reading, writing and numeracy. In Australia, Stromback and Dockery (2000) found that the special training programs in the mid-1990s had only a small impact on employment (see Table 23.1). Given the importance of employment for disadvantaged youth for themselves and for society, it must be hoped that effective training programs can be found.

Job assistance programs include help with job searches, job interviews, counselling, training in job search techniques and so on. They are designed to assist unemployed job seekers who are poorly informed about the availability and nature of jobs or their suitability for them.

Martin and Grubb (2001) report that job assistance schemes in OECD countries generally provide a small increase in employment at low cost. On the other hand, Van den Berg and Van der Klaauw (2001) found no significant impact of such programs in the Netherlands.

[3] We briefly discussed industry assistance policies to promote employment in Chapter 21.

In Australia job placement programs include job seeker assessment (by Centrelink) and job matching, job search training and intensive assistance, which are outsourced under the Job Network program. Government contracts over 200 groups who employ some 15 000 persons to provide services to unemployed persons across Australia. Evaluations of job search programs in the 1990s indicate that they produced a small net employment impact at low cost (see Table 23.1). The Productivity Commission (2002) reported that the Job Network produces similar employment gains to earlier job search programs but at lower cost.

Job creation. Spending on job creation in the public sector has also been significant in many OECD countries. However, these jobs usually have low productivity and often do not create marketable skills. Martin and Grubb (2001) found that such job creation programs had 'little success in helping unemployed people get permanent jobs in the open labour market'.

In Australia, job creation was a major labour market program in the mid-1990s and appeared to generate net employment (Table 23.1). Stromback and Dockery (2000) reported that 30 per cent of participants were in unassisted employment three months after exit from the job creation program compared with 15 per cent of the control group. Also, the positive net impact was 15 per cent after a year. Piggott and Chapman (1995) found that displacement effects were lower than with wage subsidies. However, the net cost (gross cost less savings in income support payments) of $7000–$10 000 per job was high. The Commonwealth government has now stopped supporting job creation programs except for some indigenous Australians schemes, although Work for the Dole has elements of temporary job creation.

In summary, workforce training programs, such as apprenticeship schemes, appear to have promoted employment in many countries. Special purpose training programs for disadvantaged groups have provided much more limited benefits. There appear to be fair returns to job-search assistance programs. Job creation may be helpful in some cases, but it is usually high cost and may not produce long-term productive skills.

Unemployment benefits

In many OECD countries unemployment benefits, for example the United States and the UK, are funded through a compulsory national insurance fund. In Australia, unemployment benefits (called Newstart Allowances) are funded from consolidated revenue.[4]

A Newstart Allowance is paid to unemployed or partly employed people over the age of 21 and under the pension eligibility age, who are Australian residents, who satisfy income and asset tests and enter into an Employment Pathway Plan.[5] The agreement requires recipients of the allowance to undertake certain activities to increase their opportunities for employment. This may include part-time study or training or Work for the Dole (WFD). Under WFD, people on benefits for six months or longer must provide services for a community organisation to increase their skills and job prospects.

Unemployment benefits vary with household circumstances. At the start of 2018 the maximum benefit was $486.50 per fortnight for a person in a partnership and $582.80 per fortnight for a single person with dependent children. The benefits are indexed to the Consumer Price Index and adjusted twice a year. Recipients of Newstart may also receive about $130 per fortnight in housing rent assistance (depending on family circumstances). As of the beginning of 2018, individuals with an income of less than $104 per fortnight were

[4] In many OECD countries, unemployment benefits are provided for a limited period, for example in the United States for six months. In Australia, benefits are provided without a time limit. Also, as we saw in Chapter 22 (Table 22.3, page 391), replacement rates are lower in Australia than in most OECD countries.

[5] Unemployed young people aged 16 to 20 and full-time students aged 16 to 24 are eligible for the Youth Allowance.

eligible for the full Newstart Allowance. In 2015 this represented a net replacement rate of 28% of the average wage, but significantly higher for unskilled workers. The allowance falls by 50 cents in the dollar for other income between $104 and $254 per fortnight and by 60 cents in the dollar for income above $254 per fortnight. Thus, recipients of Newstart pay an EMTR of 50 to 60 per cent on most of their income. Similar EMTRs apply to earnings of recipients of the Youth Allowance.

Labour supply effects. Unemployment benefits may affect labour supply in various ways. They may increase the probability of unemployment by making it less costly. They may also increase the duration of unemployment. On the other hand, they may increase workforce participation as the benefits are available only to persons in the workforce. Fourth, they may affect the work responses of partners of unemployed workers. Also, high EMTRs as benefits are withdrawn may encourage black market activity.

Several studies have found that the short-term labour supply elasticities with respect to welfare benefits are considerably higher than the long-term labour supply elasticities with respect to after-tax wages reported in Chapters 27 and 30. In an analysis of 20 OECD countries, Nickell (1998) found that the elasticity of unemployment with respect to the replacement rate was close to one. In a review of US and other empirical studies, Krueger and Meyer (2002) concluded that the elasticity of unemployment duration with respect to unemployment benefits was about 0.5 and that the elasticity of unemployment from the combined effects of incidence of unemployment and duration was about 1.0. Krueger and Meyer also report that higher unemployment benefits were associated with less work by wives of unemployed men. Gruber (2016, p. 415) concludes that higher unemployment benefits clearly "have a significant effect on unemployment durations".

Injury compensation

In most OECD countries, payments to injured workers are a separate category of benefit with high replacement rates. In Australia, the state governments require employers to take out insurance, known as workers' compensation, against workplace injury. Requirements vary across the states but usually involve a public–private arrangement. For example, in New South Wales, the government determines the annual premium payable by each firm as a percentage of its annual wage turnover inclusive of pension contributions. Premiums vary for several hundred types of business, from mining and farming businesses to all kinds of service businesses. It then contracts a private insurer to collect the premium from each firm and to forward this to a government agency (NSW Workcover). When claims are made, NSW Workcover reimburses the claimant via the private insurer according to various rules governing payouts. Generally, payouts replace a high proportion of lost wages, with replacement rates higher than for unemployment benefits. Proponents of this system argue that it is cost-effective because it achieves economies of scale and operates generally on a no-fault basis which minimises legal disputes, although firms are liable for damages if they ignore workplace safety regulations. On the other hand, regulation of premiums regardless of the safety record of the insured business provides little incentive to firms to minimise injuries. Also, the lack of competition between insurers, who are simply outsourced administrators for regulated premiums, is not cost-efficient.

Krueger and Myer (2002) note four potential effects of workers' compensation insurance. Firms may under-invest in worker safety. The likelihood of claims for injury may increase. Also, the duration of time out from work due to injury may increase. On the other hand, the availability of insurance may make some jobs more attractive and increase labour supply. Most work has been done on the effects of insurance on claims for, and duration of, injury. Reviewing these studies, Krueger and Myer report that combining the effects for injury claims and duration produces a total elasticity of lost work time with respect to insurance

benefits of between 0.5 and 1.0. This is slightly lower than the elasticity for unemployment benefits. However, because workers' compensation replacement rates in many countries are close to 100 per cent of earnings net of tax, the loss of working time associated with workers' compensation is large.

Disability pensions

Persons of workforce age who are deemed unable to work due to chronic incapacity are generally eligible for disability pensions. In Australia eligibility conditions are that the person must be 16 years or over but under pension eligibility age, have a physical, intellectual or psychiatric impairment rating assessed at 20 or more points, be unable to work for at least 15 hours a week for at least the next two years due to this impairment and meet residence requirements. Like unemployment benefits, disability pensions vary with household type. However, disability pensions are higher than unemployment benefits and the income and asset tests are less stringent. In addition, persons eligible for a disability pension may receive rent assistance and mobility allowances. Carers of persons with a disability also receive allowances.

In Australia, the number of people receiving a disability pension rose from 135 000 in 1979 to nearly 760,000 at the start of 2018. This is about 5.8 per cent of the workforce population, compared to an unemployment rate of 5.5 per cent. The trend elsewhere is similar. In many OECD countries the proportion of the workforce claiming disability payments trebled between 1970 and 2000 (Haveman and Wolfe, 2000). From a recipient perspective, disability pensions have several advantages. They are usually more generous than unemployment benefits. There is often no activity requirement. Disability benefits are conditional on the level of disability and not on looking for work. Once established, a disability pension is rarely taken away.

The relative generosity of disability benefits creates a significant incentive (moral hazard in US literature) to individuals to transfer from the unemployed category to the disabled category. Haveman and Wolfe (*ibid*.) reported on various attempts to estimate the relationship between the number of claimants of disability pensions and the size and accessibility of disability benefits. They found much variability but concluded that the median elasticity of disability claimants to disability benefits is about 0.6. Gruber (2016, p.420) suggests that the figure is closer to 0.3. These results suggest that a 10 per cent increase in disability benefits increases the disability roll by between 3 and 6 per cent.

Disability is a complex policy area. It is hard to design policies that simultaneously provide an adequate safety net for severely disabled persons and encourage work by those capable of some work. Identifying and monitoring needs are subjective tasks. They are made harder when persons with disabilities have a strong incentive to maximise their perceived disability. The substantial growth in recipients of disability benefits has reflected mainly a widening of the definition of disability (Krueger and Meyer, 2002). Indeed, Autor and Duggan (2006) argue that disability pensions in the United States have become a *de facto* unemployability insurance program rather than a medical insurance program. In response to these trends, many OECD governments are now considering stronger conditions on disability and part-time work requirements for persons with milder disabilities (Carcillo and Grubb, 2006). This is also current Australian government policy.

Welfare for Families

Most welfare for families revolves around children. In this section we examine welfare and market failure reasons for supporting families and various policy options. We also discuss Australian family policies along with some international policies and outcomes.

Welfare objectives

Households with children have greater needs than those without children but less ability to meet these needs as parenting reduces workforce participation. Ideally, we would have measures of need. But a family's need for assistance is neither readily observable nor a simple function of income. Need depends on the number and age of persons in the family, the employment and maternity status of the parents and the functional or dysfunctional nature of the family itself. The need for assistance rises with single-parent families. Need is also greater for younger children, especially for pre-school children who require more home care. And need rises greatly with the incidence of childbirth, which imposes substantial extra burdens on families, especially on mothers.

A common approach to estimating these needs is to estimate the expenditure (and hence income) required for a minimum acceptable level of welfare (see Chapter 20). Expenditure requirements may be estimated for each child by age group inclusive of child care costs. Single-parent income earners with limited opportunity to supply household services as well as earn income are often viewed as especially disadvantaged and, accordingly, generally receive the highest level of assistance.

Mothers are particularly disadvantaged by loss of employment prospects and may not be compensated by their partner. Accordingly, many family programs provide support directly to mothers at the time of childbirth and after. Again, measurement of equivalent need is complex when families, and indirectly mothers, enjoy different levels of income. Government must decide whether to condition support for mothers on family or individual income tests.

On the other hand, having children is largely a matter of choice. If children enhance the utility of a family, the principle of equivalent welfare across families implies that child subsidies should be less than the sum of the extra household expenditure plus income forgone. The proportion of incremental family costs that should be met by other taxpayers, including those who chose not to have children, is a normative judgement.

Finally, it may also be contended that assistance should focus on child need rather than family need. Unfortunately, there are huge differences in the emotional and practical support that children receive in the first 12 months of their life which may have enormous implications for their futures. This implies that social assistance should be tied as closely as possible to the child and / or to family support services.

Market failures may provide additional reasons for supporting families**.** Externalities are one possible market failure. Both families *per se* and high-quality early childhood education may provide positive social externalities. Intact families may contribute to social cohesion. On the other hand, children from non-intact families may perform less well on average in school and be more prone to anti-social activities. Various studies (for example, Heckman *et al*., 2010) report evidence of positive externalities from early quality childhood care. However, the policy conclusions are ambivalent. It may be inferred both that intact families should be encouraged and that children of single parents should receive extra support.

The fertility rate may also be viewed as an externality. As people live longer, the ratio of workers to dependants falls and the tax burden of supporting the elderly rises. This burden may be reduced by increasing birth rates encouraged by policies that reduce the cost of having children. But after 20 years, the standard of living is likely to be higher than today. It is not clear that current fertility rates should be subsidised to raise living standards in 20 years' time.

Conclusions. Welfare is the primary motivation for supporting families. In this case, the amount of support is a normative decision. Market failures provide some extra justification for family policies. However, as we will see below, family policies can distort economic choices, primarily labour supply and sometimes household formation. As usual, the public policy aim is to provide the desired amount of welfare benefits with minimum distortions.

Family policies and instruments

Government can assist families directly by income grants, assistance tied to services or tax relief. It may also assist families indirectly by subsidising child care services or by mandating private employer assistance.

Untied cash payments are a major form of assistance to families. They are transparent and can be targeted to family needs. Anomalies may occur when welfare payments are based on household structure and tax is based on individual income. Also, the separation of welfare payments from tax raises administrative costs

Alternatively, assistance may be tied to the use of specified services, for example registered child care services. As we saw in Chapter 22, tied grants are effectively income grants if they do not alter the composition of household expenditure that would occur with an untied grant. On the other hand, tied grants may require a family to use more of some service than they would with equivalent untied assistance. Accordingly, recipient benefits are usually maximised by income payments rather than by transfers-in-kind. Tied assistance should therefore be justified by an externality or merit good argument, such as the benefits to the children.

Government can also provide support in tax relief, via either tax deductions or rebates. For example, expenses for child care could be deducted from taxable income (a tax deduction) or from income tax liability (a tax rebate). Tax deductions are less equitable because the value of a deduction increases with the MTR and hence with taxable income. A tax rebate can be a constant amount or may vary inversely with the income or other circumstances of the taxpayer. If the rebate exceeds tax owing, the taxpayer may receive a grant.

Government may also subsidise agencies that supply services to families, notably child care services. Child care providers often argue that such a subsidy enables them to provide the quality services that young children should receive. In a competitive market, the distribution of the benefits of the subsidy would depend on the relative elasticities of demand and supply for child care services. However, in the absence of competition suppliers may use the subsidy to increase management and staff remuneration, or inputs per service provided, without necessarily providing a higher quality service.

Finally, government may mandate employer assistance to families. For example, government may require employers to provide maternity (or parental) leave. However, as we saw in Chapter 21, the costs of such regulations are usually borne in part by the employee and have a DWL. If the cost of employment rises, the demand for potential mothers as employees is likely to fall and the employment and wages of potential mothers are also likely to fall. It may also be held that assistance to families is a general taxpayer responsibility and not the responsibility of company shareholders (in so far as they would bear the costs).

Family policies in Australia

The main Australian Government family programs relate to:

- Children—family tax benefit schemes Part A and B.
- Parenting—separate programs for partnered and single parents.
- Childbirth—paid parental leave or baby bonus
- Child care—various subsidies for child care services.
- Various other assistance—for example, housing rental assistance.

Table 23.2 provides a summary of the benefits. In most cases there are several benefit rates depending on household income and other household characteristics.[6] The largest family assistance program, the 'family tax benefit' program, provides untaxed cash grants to families based on the number and age of the children. Part A is intended to assist families with the cost

[6] The Henry Tax Review (2010) recommended a major rationalisation and simplification of these benefits.

of raising children. Part B provides supplementary grants to single-income families and to families where the second parent stays at home as a child carer. Withdrawal rates for family benefits are quite low at 20 to 30 per cent and families on quite high incomes receive some benefits from Part A. These schemes are administered by the Department of Human Services through Centrelink, which pays most benefits fortnightly. Parenting payments provide extra income support to single and partnered parents with low incomes who have primary care of young dependent children. It is the only family assistance payment that is taxed. Also, withdrawal rates for parenting payments are 30 or 40 per cent.

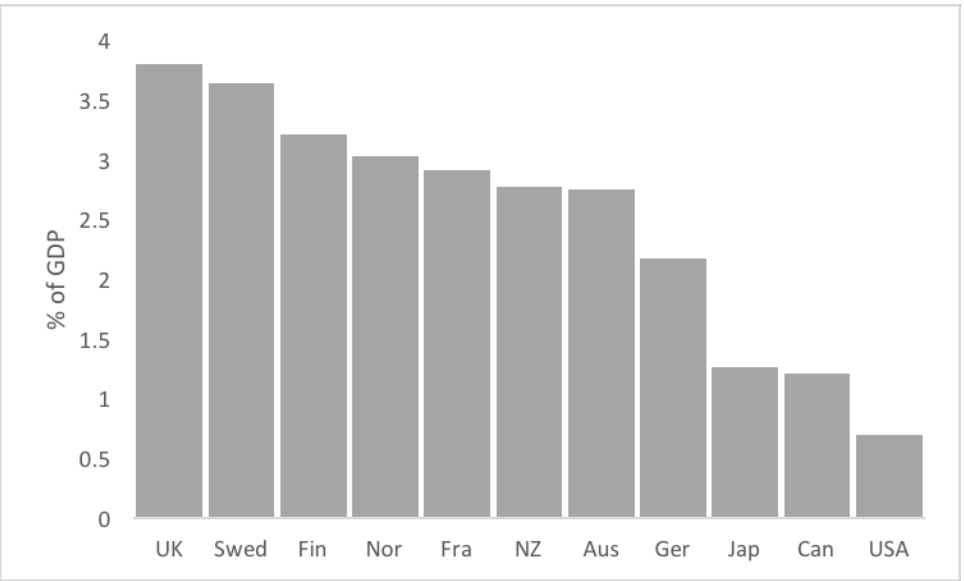
Expenditure on family assistance is about one-third of all social welfare expenditure (see Table 22.1) and about 2.7 per cent of GDP. Figure 23.2 overleaf shows family (cash benefits) assistance as a proportion of GDP for 11 OECD counties. Note however that some counties, including the United States, offer tax credits for family expenditures. Thus, full international comparisons of the generosity of family payments are complex. Box 23.1 (page 409) describes some differences with respect to childbirth assistance payments.

Equity issues. The Australian system is a mixture of highly targeted and less targeted measures. Parenting payments, rent allowances and the dependent spouse rebate are targeted at low-income households. On the other hand, family tax benefits, the childbirth allowances and child care benefits provide substantial benefits to middle-income households. Whether this complex system produces overall vertical and horizontal equity between families or more generally between all households is essentially a normative issue.

Table 23.2 Australian Commonwealth Government family assistance as at 30 June 2017[a]

Program	*Summary description*
Family tax benefit Part A	Part A is paid per child <16 or dependent full-time student aged 16–19. Benefit depends on child age and household income. Full benefits for family income <$52706. For a child under 13, this is $5504.20 p.a. Complex withdrawal rates of 20% and then 30% as family income rises depending on number of children.
Family tax benefit Part B	Extra assistance to single-parent families and families when one parent stays at home. Part B is paid per dependent child under 18. Full rate of $4412.65 for child <6 and $3190.10 ages 6-18. Single parents are eligible for the maximum rate up until adjusted taxable income of $100,000, after which they are ineligible. Benefit paid to second earner with <$5,548 p.a. but are now family income tested also. Withdrawal rate 20%.
Parenting payments	Benefit paid to Australia residents for dependent child <6 if partnered and <8 if single. Single parents receive up to $19567 p.a. if their income is less than $54,602.6. Partnered parents receive up to $12,649 p.a. Withdrawal rules are complex, but 40% for lone parents and 30% for couple families.
Paid parental leave	Primary carer (usually the mother) receives 18 weeks of government funded pay at the rate of the National Minimum Wage $695 a week before tax providing he or she has a taxable income under $150 000 in previous year.
Baby bonus	$540 upfront payment, then a maximum of 1,618.89 per week for 13 weeks for the first child and a maximum of 540.54 per week for 13 weeks for subsequent children. Subject to being eligible for Family Tax Benefit Part A
Child care benefit	Payments for approved or registered child care services. If parent is working at least 30 hours a fortnight, up to 50 hours of approved care allowed per week for non-school age child or $215 per week.
Child care rebate	Payments of 50% of out-of-pocket costs up to $7613 per annum for approved child care services less any money received under the child care benefit. There is no means test.
Rental assistance	Means-tested rent assistance that depends on number of adults and children in household as well as the level of rent paid. The maximum for a couple with 3 or more children is $176.54
Dependent spouse rebate	Taxpayers with low-income spouse (no children) obtain maximum rebate on tax of $2355 when spouse income <$9702 p.a. Withdrawal rate 25% as income of spouse rises.

(a) It should be noted that the details of assistance programs are frequently changing.



Source: OECD Social Expenditure database.

Figure 23.2 Public expenditure on family cash benefits as share of GDP in 2013

Among many equity issues, two perennials are (1) whether the system of taxes and transfers treats single- and double-income families equitably and (2) whether the lower income earner in a family is treated equitably. In Australia, with rising MTRs, a single-income family pays more tax than a double-income family with the same gross income. For example, in 2014-15 a person earning $120,000 a year paid nearly $10,000 more tax than did two people each earning $60,000 a year. On the other hand, a two-adult family with one worker at home gains significant untaxed household services. In addition, single-income families receive extra benefits from Family Tax Part B and parenting allowances. Because household types vary greatly with respect to income, size and so on, it is dangerous to generalise that single- or double-income families are invariably better treated by the tax-transfer system.

Another effect of government policy is that EMTRs are often higher for second incomes than for primary incomes because the second income attracts tax and results in withdrawals of benefits as total family income or individual income rises. Apps and Rees (2010) estimated that where the primary earner earns $50 000, then including the low-income tax offset, the Medicare levy and family tax benefits, second incomes between $5527 and $14 000 and between $14 001 and $19 423 attract EMTRs of 41.5 per cent and 56.5 per cent respectively. They argue that these are unfair (and inefficient) rates of tax on second-income earners.

Efficiency issues. Family payments may change behaviour especially when incentives, replacement rates or EMTRs are high. Gans and Leigh (2006) estimated that the $3000 baby bonus introduced on 1 July 2004 caused over 1000 births to be delayed so that mothers could obtain the bonus! More generally, the major concern is the impact of family support grants on labour supply.

Net replacement rates (NRRs) are the sum of the various support grants as a proportion of income. They depend on employment allowances as well as various family payments and rent assistance. Whiteford and Angenent (2002) estimated that, including the Newstart Allowance, family and parenting payments and rent assistance, the NRR for single parents was about two-thirds of the minimum wage rate and a half of average weekly earnings. For unemployed

Box 23.1 Childbirth assistance

The Australian government provides (subject to a generous mean test) childbirth payment to mothers in the form of paid parental leave. The payment rate is related to the National Minimum Wage at $695 per week for a max of 18 weeks. Also, employers pay varying leave entitlements. Australia does not have a separate social insurance scheme.

By contrast, many OECD governments fund maternity leave through a social insurance fund to which employers, employees and government contribute.

The amount paid to mothers varies from half of weekly insurable earnings in some countries to 100 per cent of the basic wage in Austria. Leave entitlements vary from two to thirteen months. In the UK publicly funded maternity leave pays 90 per cent of earnings with no ceiling for the first 6 weeks and mothers can take another 33 weeks of 90% of earnings up to a ceiling of GBP139.58 per week. Some countries (for instance Germany and Sweden) combine employer and social security funded schemes with separate payments by employers and government.

couples with three children, the NRR rose to around 80 per cent of the minimum wage and 67 per cent of the average wage.

Such high NRRs are likely to reduce labour force participation. EMTRs are also high for many Australian families. Individuals can receive income support from several sources without immediately losing income support. However, when the income-free threshold is reached, the interaction of income tax and withdrawal rates often produces a high EMTR. Harding *et al*. (2009) estimated that, in 2006–07, 20 per cent of sole parents and 14 per cent of couples with children experienced EMTRs of over 50 per cent. Given that the labour supply elasticity for second-income earners averages about 0.75, an EMTR of 50 per cent causes a decline in labour supply of about 35 per cent.

On the other hand, some forms of family assistance encourage labour force participation. Child care support reduces the cost of working and so tends to increase labour supply (see below). Also, maternity leave arrangements may increase female participation in the workforce. However, this labour supply effect may be partly offset by reduced demand for female workers.

Empirical evidence. There has been much international research into the impacts of family assistance. We cite examples relating to child care, fertility rates and marriage before drawing some general conclusions.

Kalb (2009) provides a wide-ranging survey of the studies of child care costs and labour supply measured both as participation rates and hours worked. Overall the studies find a small inverse correlation between child care costs and labour supply in OECD countries or conversely a very small positive effect of child care subsidies on labour supply. Many of the studies, including Australian studies (Doiron and Kalb, 2005), estimate elasticities for both workforce participation and hours worked of less than –0.1 with respect to child care costs. However, a more recent study by Gong *et al*. (2011) estimated that, for an average mother with young children, the elasticity of employment with respect to the gross child care price was –0.29 and the elasticity of hours worked was –0.65.

Turning to maternity assistance and fertility, in a study of 22 industrialised countries, Gauthier and Hatzius (1997) found that maternity leave had no effect on fertility. Fertility rates are high in the United States, which does not have a universal paid maternity scheme, and low in Sweden which has a generous maternity scheme. On the other hand, in France mothers receive generous subsidies for child care and early childhood education and the birth rate has risen from 1.6 to 1.9 per woman.

With regards to marriage, in the United States couples with dissimilar incomes pay less tax in marriage but couples with similar incomes pay more tax. Also, a two-adult family receives lower benefits per adult than a single-adult family. On balance, there are more losers than

winners from marriage. Whether this affects marriage numbers is controversial. Ehrlich and Kim (2007) found that the social security system in the United States significantly reduces incentives for individual workers to form families or to invest in children and that this has contributed substantially to decreasing marriage and increasing divorce. On the other hand, Gruber (2011) concludes that the welfare system has had little effect on single motherhood.

In conclusion, like other welfare programs family assistance involves normative judgement and trade-offs between generosity (equity) and efficiency (labour supply). Analysis of expenditure needs can illustrate the concept of equivalent household need. However, concepts of equivalence also involve value judgements. In any case, almost any program generates deadweight losses. A high tax-free family benefit creates a high NRR for low income earners. And if assistance is withdrawn sharply to medium-income households, high withdrawal rates create a high EMTR. High NRRs and EMTRs reduce labour supply. Gruber (2016, pp. 577-567) shows that no tax system can meet three generally desired goals: progressivity (average tax rates rising with family income), horizontal equity (families with equal income paying equal taxes) and marriage neutrality. However, by describing the implications of the programs, economic analysis may help to make them more equitable and efficient.

Welfare for the Elderly

The Australian Treasurer (2010) forecast that, between 2015 and 2054-55, the proportion of people over 65 will rise from 15 per cent to 22.6 per cent. Structural changes to indexation and age of eligibility in the 2014-15 budget are expected to stabilise age pension payments as a percentage of GDP. But the amount spent per person is going to increase for $2,000 today to $3,200 in 2054-55. On the other hand, the report forecast that the ratio of workers to retirees will fall from 4.5 to 2.7. Feldstein and Liebman (2002) estimated that social security payments for the aged in the United States were likely to rise from 4 per cent of GDP today to 8 per cent in 2050. In other OECD countries, such as Germany and Italy, social security retirement programs could rise to twice this amount unless programs are modified.

Clearly, most societies want to protect those in old age who have been unable to provide adequately for their retirement. On the other hand, those who can save for retirement should be encouraged to do so and the elderly who wish to work should not be discouraged from doing so. Another policy goal is intergenerational equity—a fair sharing of the burdens of old age between current and future generations.

In the following sections we discuss whether markets provide adequate retirement incomes, options for retirement pensions and the major pension policies, namely public safety-net pensions and compulsory employment-related pensions.

Markets and savings

As we saw in Chapter 5, competitive markets with rational well-informed participants provide an efficient amount of saving. Markets allow individuals to smooth out their consumption over their lifetime, saving while on relatively high incomes to augment their lower income in later periods.

Individuals can save in many ways. Housing often forms over half of an individual's net assets at the time of retirement. Pension contributions are another major form of saving. Typically, an employer contributes part of an employee's salary to a pension fund, which the employee accesses on retirement. The pensions may be defined or contributed pensions. With a defined pension, a person's pension usually depends upon their earnings and length of time with the firm, regardless of contributions made to the pension fund. With a contributory pension, benefits depend on the amount of the contributions plus investment returns net of expenses. Often an employee can take a lump sum payment in lieu of a pension.

Markets can also allow individuals to convert their financial assets, including the capital value of accumulated pension funds, into annual income by way of annuities. An annuity provides a fixed income from an agreed age until an individual dies, regardless of the length of their life. In some markets, annuities can be purchased against equity in a house. Thus, markets provide a means whereby individuals can spread out their wealth, or some fraction of it, into lifetime consumption.

However, there is a constraint on these market processes—that is government's double taxation of savings (once when income is earned and then the return on savings) that is likely to discourage savings.

Moreover, individuals may not save if government is going to provide a retirement pension. To offset the double taxation and to encourage savings (and avoid calls on public pensions), many governments provide large tax concessions to pension schemes.

Market failures, equity and role of government. Diamond (1977) identified four reasons for government intervention in retirement provision: inadequate saving for retirement, market failure in consumption smoothing and risk management, administrative efficiency in large-scale retirement provision, and redistribution.

Inadequate saving occurs when individuals can save sufficient for retirement but fail to do so because of poor decision making. Bernheim and Rangel (2008) report that many people under-save compared with expected utility-maximising consumption smoothing behaviour.[7] Individuals may accumulate insufficient wealth because of poor returns on their portfolio. In some countries, though not in Australia, employers choose the pension fund for their employees. Indeed, some large US firms invest over half of the pension funds in themselves. Sometimes such firms go bankrupt. Enron was the seventh largest company in the United States before it went bankrupt in 2001. Pension funds went with it.

Second, markets do not always provide consumption smoothing or manage risk well. Longevity, inflation and returns on investment are difficult to predict. Life annuities with periodic payments indexed for inflation provide some insurance against these risks. However, the market for such insurance is thin and it is difficult to insure against market risk. Annuities are costly and returns on them are generally below market rates of interest because of high transaction costs and adverse selection. Unless insurers can predict longevity, they must offer average premiums. But individuals with short life expectancy have less incentive to take out annuities at average premiums. So, premiums for annuities rise. Adverse selection narrows the market for annuities and raises the cost. Mandatory social insurance resolves this problem by forcing everyone into some form of insurance.

Third, if there are economies of scale in pension management, a national public pension system would minimise administration costs per pension. However, large programs in the public and private sector have x-inefficiency (behavioural) costs that may offset potential technical efficiency gains.

Fourth, and certainly not least, whatever the efficiencies or otherwise of markets, many people will have insufficient means to provide adequately for their retirement.

There are therefore several reasons for government involvement in retirement provision. This may mean tax concessions to encourage saving and to offset its own taxes on savings, policies to encourage annuity markets and the provision of public pensions for low-income retirees. However, provision of public pensions may discourage both saving and labour supply. Also, any shortfall in individual savings places pressure on government to raise taxes or to cut other programs. Government has good reasons therefore to encourage people to provide for their own retirement where they can do so.

---

[7] This may be viewed as a merit good argument rather than as market failure. Pigou (1920) argued similarly that many people underestimate the value of future consumption, which he described as myopia.

Retirement policies

Most retirement provision schemes are based on one or more of three pillars:

1. A public age pension based on a means-tested, safety net system funded from consolidated revenue designed to ensure that basic needs are met.
2. Compulsory contributory, employment-based payments to designated retirement funds.
3. Voluntary savings, with or without tax concessions.

Table 23.3 summarises these options. As noted, retirement systems vary in different countries. The Australian system is described in Box 23.2.

Some countries, like Australia, have tax-funded, safety net public pensions (the first pillar) targeted at low-income persons. Payments are generally made on a pay-as-you-go (PAYG) basis out of consolidated revenue. The benefits usually depend on a person's current status and are independent of employment history or contributions to a retirement fund.

Many countries have second-pillar systems—compulsory employment-based payments for retirement provision either to a publicly-owned pension fund or to private funds. Pensions may be related to contributions or earnings (if they are defined benefits) but they may also include basic or minimum pensions. Some funds operate on a PAYG basis, other are fully funded as liabilities occur. Traditionally, public funds usually operated on an unfunded PAYG basis and provided defined benefits. This is still broadly the system in the United States. Other countries have moved towards greater funding of liabilities as they occur and to benefits related to contributions and investment returns.

Alternatively, and increasingly, government may require employers to make payments to private pension funds. Government may also encourage this by tax concessions at three levels: (1) on contributions to retirement funds, (2) on fund earnings and (3) on retirement incomes taken out of the funds. Private pension funds have always been funded more fully than public funds. But, as in the public sector, there has been a move away from defined benefit and towards contribution schemes.

Table 23.3 Taxonomy of main retirement funding options

Basic system	*Nature of funding*	*Nature of benefits*	*Comments*
First pillar			
1 Safety net pensions	PAYG from consolidated revenue	Usually targeted defined benefits; no employment history required	Not common Over 70 per cent of Australian retirees receive the age pension
Second pillar schemes: employment-related contributions			
2a. Compulsory payments to government (usually a national insurance fund)	PAYG. Current workers pay for current retirees. Usually small reserves	Generally defined benefit schemes	Most common pension system, especially in Europe
2b. Compulsory payments to a public authority	Fully or partially funded; payments made as liability is incurred	Mainly defined benefit schemes	Examples: India, Malaysia, Singapore
2c. Mandated payments to private pension funds	Fully or partially funded as liability is incurred	Benefits usually related to contributions and net investment returns	Examples: Australia, Chile, Switzerland, UK
Third pillar			
3. Voluntary savings	Funded by various private savings with or without tax concessions	Benefits usually related to contributions and net investment returns	Many countries provide tax concessions to long-term savings, e.g. Australia, US, UK

Box 23.2 Retirement income provision in Australia

Retirement income provision in Australia comprises elements of all three pillars, including a means-tested public pension financed from consolidated revenue, mandatory superannuation contributions by employers and voluntary saving with tax-preferred concessions.

The public Age Pension provides an income of 27.7 per cent of average male earnings to single people and about 20 per cent of earnings to each of a married couple. The pension is means tested on assets and income. Despite the means tests, three-quarters of retired persons of eligible age (over 65 for men and 63 for women increasing to 65 by 2014) receive some age pension and about 60 per cent are paid the full rate. Retirees who rely fully on the aged pension pay no income tax. This implies a net of tax replacement rate of 37 per cent of average weekly earnings or about 70 per cent of the minimum wage.

The Superannuation Guarantee legislation requires employers to make superannuation contributions of at least 9.5 per cent of gross wages on behalf of their employees to a complying superannuation fund. The maximum amount that employers can provide has varied greatly. In 2017-18, employer contributions are capped at $25 000 per annum. Also, retirement funds are now capped at $1.6 million (with exceptions for pre-existing larger funds). The contributions must remain in the fund until statutory age for access is reached (currently 55 rising to 60 by 2025), when employees are entitled to all accrued benefits. Superannuation can be taken as lump sums or as pensions.

Concessionary taxation applies to all stages of the pension process. Employer contributions are tax deductible to the employer but taxed in the hands of the superannuation fund at a rate of only 15 per cent. Fund earnings are also taxed at 15 per cent prior to retirement. Finally, when the fund starts to pay out a pension, the relevant fund earnings are not taxed and payments to pensioners are tax free.

In addition, in 2017-18 the Australian government contributed 50 cents for each dollar of an individual's after-tax superannuation contribution for persons earning less than $36,818 up to a maximum co-contribution of $1000 with smaller co-contributions for incomes up to $51,813, subject to at least 10% of your income coming from eligible income sources.

To make the superannuation system simpler and more equitable but also to retain the incentives to save, the Henry Tax Review (2010) recommended the following changes to the employer-based pension system.

- Employer contributions should be treated and taxed as employee income at marginal tax rates but with an offset making an effective progressive tax rate of zero, 15 and 25 per cent. Contributions in the fund would not be taxed.
- The government's co-contribution for low income earners and the superannuation spouse contribution offset would be abolished.
- The rate of tax on superannuation fund earnings should be halved to 7.5 per cent.
- The restriction on workers aged 75 and over from making contributions should be removed.
- The government should also consider offering annuity products.

However, the Review did not recommend any increase in the then 9 per cent superannuation contribution. As Guest (2010) notes, there is dispute whether this is sufficient to ensure an adequate standard of living in retirement. But Rothman (2007) contends that there will be no savings gap for cohorts who have contributed 9 per cent over most of their working like.

The third pillar is voluntary saving. This may also be supported by tax concessions to super funds to encourage savings. Typically, these savings would complement a mandated employer-based pension scheme.

We consider below some key issues including financing of public pensions, the role of public versus private schemes, the design and effects of tax concessions, and impacts on labour supply and savings.

Funding for public pensions

The key funding issue is whether to employ pay-as-you-go (PAYG) or to fund liabilities when they occur. This issue arises principally with defined contribution (DC) pensions including public pensions rather than with contributed pensions where the retiree takes the risk of poor pension fund performance. There are three main funding questions. Is PAYG or

full funding fairer? Is PAYG sustainable? And, what are the efficiency implications, if any, of the funding alternatives?

Before responding to these questions, we should note that there is no fiscal difference between a PAYG scheme run with balanced budgets and a fully funded insurance scheme combined with budget deficits. The two strategies have the same effect on public liabilities and net public worth. To differentiate PAYG and fully funded systems, we must assume that fully funded pensions are not funded by increasing debt.

With PAYG, taxes on existing workers are paying for the pensions. With rising income per capita and pensions linked to current earnings (for example to average weekly earnings as in Australia), each retiree receives more consumption goods than he or she pays for. Current taxpayers also gain because they in turn make lower contributions towards their pensions than they would make under a fully funded system. Thus, there is constant inter-generational redistribution. If per capita income rises, this outcome is consistent with vertical equity.

However, the sustainability of a PAYG system depends on the growth in the workforce. If the proportion of aged in the population rises with declining birth rates and rising life expectancy, PAYG systems may not be sustainable without increases in tax rates. This can be demonstrated as follows. In a PAYG system the benefits received by retirees must equal payments by current workers. Thus,

$$N_r \times B = N_w \times W \times t \tag{23.1}$$

where N_r is the number of retirees, B is the benefit per retiree, N_w is the number of workers, W is the wage and t is the tax rate to pay for the pension. This can be rearranged as:

$$t = (N_r/N_w) \times (B/W) \tag{23.2}$$

The term (N_r / N_w) is the ratio of retirees to workers (the dependency ratio). (B/W) is the ratio of average benefits to average wages (the replacement ratio). The tax rate, which is a function of these two ratios, can remain stable so long as the aim is simply to maintain the real level of pensions and the real wage rises fast enough to offset any rise in the dependency ratio. However, tax rates must rise if pensions rise with average earnings and the dependency ratio rises.

Turning to efficiency issues, if real pension values are always rising, at any point in time a PAYG system will result in lower tax rates on existing workers and so distort labour supply by less than would a fully funded system. On the other hand, Feldstein and Liebman (2002) note that full funding of pensions could increase the present value of consumption over time if it increases savings and investment and the marginal product of capital exceeds the social time preference rate.

Mandated employment-based pensions

Under a market-based contribution insurance system everyone receives a return on their contributions and the earnings of these contributions. Benefits are independent of family structures and earnings tests. In a mandated private insurance system, like the Australian one, there is no redistribution as a function of individual needs other than longevity. This means that a complementary public pension system is required for people who cannot finance their own retirement.

Under most state-funded retirement systems, however funded, there is some redistribution. Ignoring variations with longevity, high earners receive less, and low earners receive more, than they contribute. Even the national insurance systems in the UK and the United States combine some redistribution with contribution-related pensions.

The chosen channel (public or private) for retirement funds can have significant impacts on fund management and the operations of asset markets. Traditionally managers of public funds were required to invest prudently, mainly in bonds and little or nothing in equities or property.

Managers of private funds can make a wider range of investments and offer clients a wider range of risk–return options. They also compete more vigorously than public funds, especially those funded through taxation. Although returns are more volatile and market risks are higher with private funds, given competition and adequate choice private funds can replicate the security of public funds if the market wishes.

On the other hand, management costs and fees tend to be higher in private funds than in a publicly administered scheme either because of aggressive fund managers or diseconomies of small size in some cases. This has led the Australian Government to mandate that all private funds must offer a default 'My-Super' pension based on safe and simple investments with a low management fee.

Taxation concessions

Table 23.4 summarises how the Australian taxation system treats the major avenues of saving quite differently. As seen in Box 23.2, savings in retirement funds are treated favourably at each main point: on entry, on earnings and on exit. Owner-occupied housing is also treated very favourably with no tax on imputed rents (earnings) or on capital gains (on exit). On the other hand, investment in property and in equities is fully income taxed on entry and on earnings and attracts capital gains tax. Thus, the tax system encourages allocation of savings to superannuation funds. The concessions to earnings in super funds also reduce the effective tax rates on super fund investments in financial assets, equities and property. Arguably, these concessions reduce the distortions that otherwise occur in investment decisions.

Most countries provide substantial tax concessions on contributions to retirement funds (such as tax-free contributions up to a maximum amount) as well as on the earnings of these funds, but tax income when exiting the funds (when income is consumed). Systems where tax is wholly exempt in the first two stages (contributions and earnings) but retirement income is taxed are known as EET systems. The Australian system is the reverse: a TTE system. The EET approach is more consistent with the notion that consumption should be taxed and savings should not be taxed. It is also more equitable to tax income at MTRs (when it is consumed) rather than to levy flat tax rates on contributions and earnings.

Household saving

Do the tax concessions for superannuation affect the total amount of savings? In general, the answer depends on the relative strength of the income and substitution effects. A higher after-tax return on savings reduces the need for saving (the income effect) but increases the value of savings relative to current consumption (the substitution effect). Also, tax concessions for retirement funds may have little effect on total savings as savers may substitute savings in lower taxed retirement funds for other savings.

Turning to forced retirement saving, this may not increase total saving as people can reduce their other savings. This may occur even if the forced saving exceeds a person's voluntary saving! Suppose that Clare has an income of $50 000 a year and would save $5000 by investing in shares or repaying the house mortgage, but that the government mandates her to contribute $7000 to a pension fund. Clare could reduce her personal investments by $5000

Table 23.4 Australian taxation and forms of saving

Form of saving	*Taxed on entry*	*Taxed on earnings*	*Taxed on exit*
Retirement fund	Partially income taxed	Partially income taxed	Not taxed
Own housing	Fully income taxed	Not taxed	Not taxed
Term deposits / bonds	Fully income taxed	Fully income taxed	Not taxed
Investment in property	Fully income taxed	Fully income taxed	Capital gain taxed
Equities	Fully income taxed	Fully income taxed	Capital gain taxed

and increase her total savings by only $2000. But, more than this, in a competitive capital market Clare could borrow $2000 to finance the consumption forgone and so end up with no net change in savings. It follows that forced saving could be fully offset by a decline in private saving and social security saving would have no effect on aggregate saving in the community.[8]

In practice, it appears that forced saving does increase total saving, but not by the full amount because of falls in private saving. Feldstein and Liebman (2002) review the extensive time series and cross-sectional studies in the United States. In the time series studies, per capita consumption is typically estimated as a function of disposable income, per capita wealth and per capita social security wealth. Drawing on Feldstein (1996), they estimate that a dollar increase in social security wealth reduces private saving (and increases consumption) by 56 cents. In cross-sectional studies, the financial wealth of individuals at the time of retirement is typically estimated as a function of lifetime income, demographic variables and the present discounted value of future social security benefits. The results are variable. Feldstein and Liebman (*ibid*.) conclude that each dollar of social security benefit is likely to reduce private savings by between zero and 50 cents, with the most likely estimate lying near the middle of the range. On the other hand, Gruber (2011) concludes that studies of the effects of tax concessions for employer-provided pensions on savings in the United States are inconclusive.

In Australia, the ratio of household saving of GDP fell significantly both in the 1980s when the superannuation guarantee was introduced and in the 1990s when the government mandated that employer contributions to employee superannuation rise from 6 to 9 per cent of gross wages. However, conclusions about cause and effect must be drawn cautiously. Other factors, for example an increasing proportion of elderly in society, may have contributed to the changes in the national household saving ratio.

There is less ambiguity about the impact of the age pension on savings. The Australian Age Pension is currently about $18 000 per person and a retiree loses 50 cents in the dollar for each dollar of private income over about $4000 up to about $40 000 at which point there is no public age pension. Thus, retirees have little incentive to save $100 000 or $200 000 (outside the untaxed family home) as returns on investments will attract an implicit 50 per cent tax rate.

Labour supply

Retirement benefits can affect labour supply in various ways. The benefits may reduce both the need for income (the income effect) and the net income from work (the substitution effect). This is especially the case for the public age pension. As just noted, publicly supported retirees are discouraged from earning up to $40 000 per annum because there is an implicit tax of 50 per cent on each dollar earned. Put another way, the price of leisure has fallen by 50 per cent. Naturally this increases the demand for leisure. Retirement benefits may also affect the labour supply of spouses.

Various time series studies have examined the effects of changes in social security payments on labour supply over time. Cross-sectional studies have examined how variations in social security payments within, or across, countries affect labour supply. Studies that draw on longitudinal data draw on both time series and cross-sectional variations in benefits.

There can be no doubt that increasing wealth has reduced the labour force participation of the aged. In most European countries labour force participation of males aged 65 and over has fallen from around 60 per cent to under 10 per cent (Feldstein and Liebman, 2002).

[8] This is known as Ricardian equivalence. See Barro (1978) and Chapter 24.

However, the effect of retirement benefits on labour supply is debated. The cross-country studies in Gruber and Wise (1999) suggested that social security systems do affect labour supply. The level of implicit tax rates on continued work above retirement age explained an estimated 80 per cent of the variation in unused labour capacity of 55 to 65-year persons. Also, a high proportion of people retire at ages around 62 to 65 as they are eligible for social security benefits. On the other hand, Krueger and Meyer (2002) concluded from their review of US studies that social security wealth has only a modest impact on labour supply in the United States.

Conclusions

The major policy objectives of retirement policies are (1) to ensure that individuals who can provide for their retirement do so and (2) to protect those who cannot. Government has a financial interest in (1) and a social welfare interest in (2). Objective (1) presumes that retirement assistance is not an entitlement due to previous tax contributions, which is how it is sometimes viewed. In addition, the provision of retirement savings may be viewed as part of the objective of adequate aggregate saving for the economy. However, this begs the question as to whether aggregate domestic savings matter in an open economy where foreign savings are readily available.

To achieve objective (1), most governments allow substantial tax concessions to pensions. The Australian government allows major tax concessions to both contributions to pensions and to earnings of pension funds and pensions are themselves tax-free. To achieve objective (2) the government provides means-tested public pensions to about 80 per cent of retirees on a pay-as-you-go (PAYG) basis from consolidated revenue.

PAYG funding introduces issues of sustainability and intergenerational equity. In many countries, including Australia, the cost of public pensions as a proportion of GDP is forecast to rise significantly and the share of workers available to fund it to fall significantly. Therefore, future generations will pay some of the costs of current workers. This may be regarded as equitable because future generations are likely to be better off than the present one. Guest and McDonald (2002) estimated that, with realistic assumptions about growth in labour productivity and demographic changes, living standards per capita are likely to be 75 per cent higher in 2042 than they were in 2002. It is, of course, a value judgment as to how much current pension liabilities can fairly be passed on to future generations, who may turn out to be less affluent than this forecast. The move towards contributed pensions and away from defined benefits will significantly reduce the costs for future generations.

Turning to efficiency issues, the main concerns are the impacts of retirement policies on labour supply, private savings and the allocation of capital. Both theory and most evidence suggest that retirement benefits significantly affect labour supply. A large proportion of workers retire fully once pensions become available. The impact of retirement policies on savings is less clear. Tax concessions for saving in retirement funds probably increase savings and may reduce the inter-temporal distortion of consumption caused by income taxation. On the other hand, forced retirement savings reduce private savings significantly so that the net savings effect may be small. Public age pensions certainly reduce savings. Finally, the tax concessions to super may distort saving avenues but they probably have little final effect on capital allocation.

Summary

- Welfare programs of any kind raise three general issues: the measurement of needs; how to provide generous support to households with needs without distorting behaviour; and the roles of the government and the private sector.
- Welfare programs for persons of workforce age include wage-rate subsidies, training and job assistance programs, unemployment benefits, injury compensation and disability pensions.
- Wage-rate subsidies have a modest effect on employment and wage rates, with some benefits accruing to employers and existing workers rather than new workers.
- Training and job assistance programs can improve employment, but job creation specifically for unemployed workers appears to have limited benefits.
- Unemployment benefits, injury compensation and disability benefits create disincentives: the levels of unemployment, injury and disability are responsive to some extent to the levels of benefit provided.
- Family needs are a function of income, the number of children, mothers working outside the home and single-headed families. Quantification of these needs is generally based on expenditure requirements but is ultimately a normative decision.
- In Australia family assistance payments exceed 2.5 per cent of GDP. Some assistance is heavily means tested and some much less so. Benefit withdrawal rates vary from 20 cents to 60 cents in the dollar. The higher withdrawal rates have significant labour supply effects.
- Provisions for retirement include public age pensions funded from consolidated revenue, compulsory employment-based payments to accredited pension funds and voluntary saving. Tax concessions are an important complementary policy.
- Government has a major welfare role in retirement provision primarily to protect low-income households as well as households that make inadequate personal saving for retirement.
- Theory and most evidence suggest that all forms of retirement schemes reduce labour supply.
- Tax concessions and mandated retirement schemes tend to increase household saving, but not necessarily by a large amount. Public pensions reduce household savings.

Questions

1. Suppose that the government is considering whether to provide income support to low-income workers by minimum wage regulation, earned income tax credits or employment subsidies. Which method would you recommend and why?
2. What is the evidence that unemployment benefits reduce labour supply?
3. Given that disability is often a matter of degree, how can government distinguish between those people who have a disability and those who do not? And how can government establish adequate benefits for people with minor disability without destroying the incentive to work?
4. Should family assistance be means tested on the family income or the income of the second earner?
5. Do children provide utility to parents? If so, should people who do not have children be required to subsidise families with children?
6. Who benefits from governments subsidies to suppliers of child care services? Explain the reasons for your answer.
7. What does full funding of a pension mean? Should public pensions be funded by a pay-as-you-go system or fully funded?
8. Tax concessions may be applied to contributions to pension funds, to earnings of funds or income from funds. Are some tax concessions more equitable or efficient than others?
9. In Australia employers are responsible for paying the 9 per cent Superannuation Guarantee for employees. Does this mean that the employer bears the cost of the superannuation payments? If not, how are the payments borne?
10. The Superannuation Guarantee requires that employees save 9 per cent of their gross wage in a superannuation account. Does this mean that employees will save a minimum of 9 per cent of their wage each year?
11. What are the major impacts of alternative retirement schemes on (i) labour supply and (ii) household savings?

Further Reading

Apps, P. and Rees, R. (2010) 'Australian tax reform and the targeting fallacy', *Australian Economic Review*, 43, 153–175.

Australian Treasurer (2010) *Australia to 2050: Future Challenges, the 2010 Intergenerational Report*, Canberra.

Barr, N. (2004*) The Economics of the Welfare State*, Chapters 8–10, Oxford University Press, Oxford.

Feldstein, M. and Liebman, J.B. (2002) 'Social security', pp. 2246–2323, *Handbook of Public Economics*, Vol. 4, A.J. Auerbach and M. Feldstein (eds), Elsevier Science, North Holland.

Gruber, J. (2016) *Public Finance and Public Policy*, Chapters 14, 17 and 22, Worth Publishers, New York.

Guest, R. (2010) 'Policy options to increase retirement saving in Australia', *Australian Economic Review*, 43, 293–301.

Kalb, G. (2009) 'Children, labour supply and child care: challenge for empirical analysis', *Australian Economic Review*, 42, 276–299.

Katz, L.F. and Meyer, B.D. (1990) 'The impact of the potential duration of unemployment benefits on the duration of unemployment', *Journal of Public Economics*, 41, 45–72.

Krueger, A.N. and Meyer, B.D. (2002) 'Labour supply effects of social insurance', pp. 2327–2391, *Handbook of Public Economics*, Vol. 4, A.J. Auerbach and M. Feldstein (eds), Elsevier Science, North Holland.

Rosen, H. and T. Gayer (2014) Public Finance, 10th ed., Chapter 13, McGraw-Hill, New York.

Whiteford, P. and Angenent, G. (2002) *The Australian System of Social Protection—an Overview*, Occasional Paper No. 6, Department of Family and Community Services, Canberra.

Whiteford, P. (2010) 'The Australian tax transfer system: architecture and outcomes', *Economic Record*, 86, 528–544.

Chapter

24

Health

Who shall live and who shall die, who shall fulfil his days and who shall die before his time …

Yom Kippur (Day of Atonement) prayer book

Health Care Services and Health ♦ Market Failures, Equity and the Role of Government ♦ Optimal Provision of Health Care Services ♦ Financing Health Care Services ♦ Delivery of Health Care Services ♦ Health Care in Australia

Our enjoyment of all goods depends on our health. Without good health our enjoyment of all other goods is diminished. In extremity, life itself is lost. Not surprisingly, there is intense popular concern about the supply of health care services. On the other hand, there is concern about the rising expense of health care services driven by rising demand, increasing longevity and technological advances that can deliver ever more services. This raises such fundamental policy questions as: What is the optimal provision of health care services? How can these services be provided efficiently and fairly? And, what is the role of government in the provision of health care services?

Governments have long been involved in the provision of public health. Ancient Rome was famous for the clean water and sanitation systems provided by the government. Health care is today, more than ever, one of the main functions of government. This can be justified partly by market failures. But, for most people, equity considerations provide the most compelling reasons for government involvement in financing and regulating health care services. Many governments also deliver many health care services, but private delivery of health services has an important complementary role in many countries.

In this chapter we discuss the nature of health care services, the role of government in provision of health services, the optimal provision, financing and delivery of health care services, and the provision of health care services in Australia.

Health Care Services and Health

Health and health care services are not the same. Good health is the outcome that people want. Health care services along with genetic endowments and lifestyle are inputs into the production of health outcomes (see Figure 24.1).[1]

[1] In public administration terminology, health care services are often described as outputs, which draw on factors of production as inputs, to provide health outcomes (see Chapter 16).

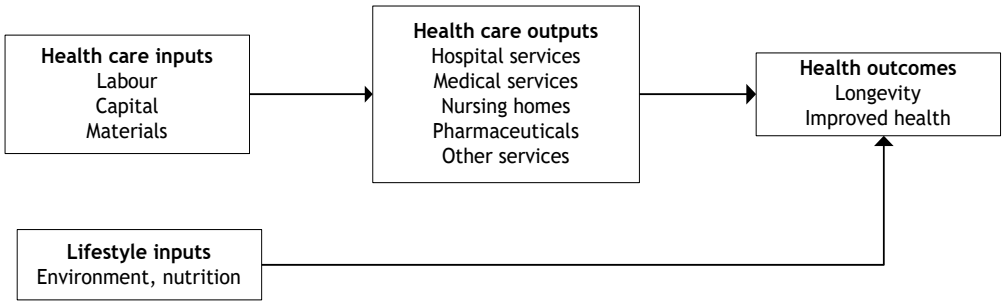


Figure 24.1 Health resources, services and outcomes

The distinction between health inputs and outcomes is critical. It means that valuing health outcomes is a crucial issue in allocating resources to health care services. But here we start by describing the quantum of resources allocated to health care services and the nature and funding of these services.

Health care expenditure and funding in OECD countries

Table 24.1 shows expenditure and funding for health care services in 12 OECD countries in 1998 and 2008 and data for Japan for 1998 and 2004. The United States and France had the highest health expenditure in relation to GDP. The UK, Australia, Norway and Japan had below-average expenditures. For the 29 OECD economies in 2008, the median health care expenditure was 9.1 per cent of GDP.

In most OECD countries expenditure on health care rose by over one percentage point of GDP in the 10 years to 2008. This continued a long-run trend of rising health expenditures as a proportion of GDP.

Table 24.1 Health care expenditure and funding for selected OECD countries

	1998		*2008*		
Country	*Health care to GDP (%)*	*Per person ($A)*	*Health care to GDP (%)*	*Per person ($A)*	*Government expenditure as % of total*
United States	13.4	5549	16.0	11156	46.5
France	10.1	3030	11.2	5470	77.8
Switzerland	10.1	3905	10.7	6848	59.1
Germany	10.2	3249	10.5	5351	76.8
Canada	9.0	3026	10.4	6037	70.0
Netherlands	8.1	2691	9.9	6013	75.3
New Zealand	7.8	1901	9.9	4136	80.4
Sweden	8.2	2596	9.4	5136	81.9
Italy	7.7	2401	9.1	4248	77.2
United Kingdom	6.7	2042	8.7	4631	82.6
Australia	7.8	2535	8.7	5021	68.5
Norway	9.3	3323	8.5	7404	84.2
Japan	7.3	2289	8.7[a]	5021[a]	81.5[a]
OECD median	7.8	2345	9.1	4801	74.2

(a) 2004 data.

Source: Australian Institute of Health and Welfare (2010).

The two major factors driving this growth in expenditure are the increasing proportion of aged persons and technological developments. Older persons have greater demands for health care services. Also, expensive interventions can increasingly keep people alive for a few extra years, sometimes in a highly weakened state. Cancer treatment for the not-old is also often very expensive. This raises in acute form the value of an additional life year and the optimal allocation of scarce resources.

In 2008, government expenditure accounted for 74 per cent of total health care expenditure in the median OECD country. Government expenditure accounted for over 80 per cent of health care expenditure in Norway, the UK, Sweden and New Zealand. The proportion fell to 59 per cent in Switzerland and to 47 per cent in the United States (pre-Obamacare). As can be observed casually from Table 24.1, health care expenditure as a percentage of GDP tends to fall as the proportion of health care expenditure funded by government rises. There are two competing explanations of this phenomenon. One is that government provision of health care services is more efficient, reduces the costs of health care services and makes health care affordable. The other is that public provision of health care services crowds out private provision and results in under-allocation of resources to health care.

Health care expenditure and funding in Australia

In 2008, Australian expenditure on health care was 8.7 per cent of GDP. As in other OECD countries, expenditure on health has risen substantially in real terms. In 2008 dollars, expenditure per person rose threefold from about $1700 in the mid-1980s to $5021 in 2008. Currently health care services account for about a fifth of all government spending in Australia.

Table 24.2 shows health care expenditures in Australia by major areas and funding sources in 2008–09. Expenditures on hospitals, medical services and medications accounted for 39

Table 24.2 Health expenditure in Australia by area and by source of funds in 2008-09

	Expenditure		*Government funding*			*Private funding*			
Area of expenditure	*($bn)*	*(%)[a]*	*Central (%)[b]*	*Other (%)[b]*	*Total (%)[b]*	*Insurance (%)[b]*	*Individuals (%)[b]*	*Other (%)*	*Total (%)[b]*
Recurrent spending									
Public hospitals	33.4	31.1	38.2	54.0	92.2	1.8	2.9	3.1	7.8
Private hospitals	8.4	7.8	34.6	4.1	38.7	51.3	4.6	5.4	61.3
Patient transport	2.4	2.2	11.2	65.1	76.3	6.2	13.9	3.7	23.7
Medical services	19.8	18.5	78.1	0.0	78.1	4.6	12.1	5.2	21.9
Dental services	6.7	6.2	13.5	9.3	22.8	15.4	61.5	0.3	77.2
Other health practitioners	3.4	3.2	34.6	0.0	34.6	14.6	41.8	9.1	65.4
Community health	5.6	5.2	13.1	83.0	96.1	0.0	2.8	1.2	4.0
Public health	2.3	2.1	51.5	42.8	94.3	0.0	0.9	4.9	5.8
Medications	15.2	14.2	51.5	0.0	51.5	0.3	47.7	0.5	48.5
Aids and appliances	3.3	3.1	15.6	0.0	15.6	11.1	71.5	1.5	84.4
Administration	3.0	2.8	54.8	13.1	77.9	31.0	0.0	1.1	42.1
Research	3.7	3.5	75.0	16.9	91.9	0.0	0.5	7.7	8.2
Total recurrent	107.1	100.0	45.0	25.4	70.3	8.2	18.1	3.3	29.7
Capital expenditure	5.7								
Total health care expenditure	112.8								

(a) Percentage of total recurrent expenditure.

(b) Percentage of category of expenditure.

Source: Australian Institute of Health and Welfare (2010).

per cent, 19 per cent and 13 per cent respectively of recurrent health care expenditures (71 per cent in total). Several other areas of expenses make up the rest of the expenditures. Hospitals also account for most of the capital expenditure.

Government funded 70 per cent of total health care expenditure including capital expenditure. The Commonwealth government funded 43 per cent and the state and territory governments funded 26 per cent. Government collectively funded a high proportion of the expenditures in public hospitals, medical services, pharmaceuticals, community health and public health. The Commonwealth is also responsible (through the Therapeutic Goods Administration) for determining what medicines can be used in Australia. Governments collectively deal with national health issues such as HIV/AIDS, drug abuse and obesity.

Private funding sources paid for the balance of 30 per cent of all health care expenses. Private sources paid for over half the expenses in private hospitals, other health practitioners, aids and appliances, and dental services. They also paid for over 40 per cent of the costs of medications. Households paid for over one-fifth of all health care expenses. Although over 40 per cent of households hold private health insurance, private insurance funds contributed only 7 per cent of all payments for health care expenses.

Market Failures, Equity and the Role of Government

Various forms of market failure occur in the health sector. Some health care services provide substantial positive (non-excludable) externalities which effectively make them a public good; some consumer products such as alcohol and tobacco have negative personal health and social (externality) impacts; and the health care market is characterised generally by asymmetric information and by limited competition in low density population areas.

A prime example of the public good nature of health services is public health immunisation programs. It provides non-excludable benefits. Where disease is contagious, health care services supplied to one person reduce the risk of illness to others. In private markets individuals would underspend on such services. Medical research and knowledge is also a non-rival public good. Use of the research by one agent does not reduce its value to anyone else. Thus, support for basic medical research is a public good.

Consumption activities such as smoking, drinking and fast driving may cause health risks or traumatic injuries to others. Pregnant women who consume a large amount of alcohol are liable to bear babies with brain damage. In addition, heavy drug takers, smokers and drinkers may do self-harm and require substantial public health services with the costs borne by other taxpayers. Government has an interest in mitigating all such behaviours on externality (social and financial) and merit good grounds.

Turning to asymmetric information, medical practitioners are generally better informed about the service needs of patients than are the patients themselves and patients rely heavily on their advice. This may create supplier-induced demand (SID). SID occurs when a health care supplier induces a patient to choose a service that he or she would not choose if they were better informed. Thus, suppliers of health care service can to some extent create a demand for their services, especially when patients do not pay the full cost of the service. Unlike in an informed market, an increase in the number of health care suppliers may increase the supply of services without reducing their price. Cromwell and Mitchell (1986) found that a 10 per cent increase in the ratio of doctors to population in the United States resulted in a 9 per cent increase in doctors' fees (though possibly this reflected high income areas).

Of course, information asymmetry is not unique to health care or necessarily a warrant for government action. When we take our car to a garage for repairs we may have little idea of the repairs required. However, we do have an incentive to find a cost-effective garage. In a non-subsidised health care market, consumers would have more incentive to find the best and

most cost-effective service. However, the gap in knowledge between suppliers and consumers appears greater and more critical in health care than in most other markets.

Information asymmetry also affects the market for health insurance. In this case, consumers are often better informed about their own health than are insurers. If an insurance company cannot distinguish between individual risks and offers each person the same (average) price for insurance, high-risk individuals will accept insurance and low-risk individuals will not. This raises the price of insurance and leads to more low-risk individuals forgoing health insurance. This phenomenon, known as adverse selection, causes the market to under-supply health insurance to low-risk clients.[2]

Fourth, there is limited competition in the supply of private health care services in low density areas. This reflects the role of fixed costs and economies of scale or scope in the supply of hospital services. In many regions, only one hospital is economically viable. In many towns, only a few doctors could earn a professional income and they may believe, not unreasonably, that the community is served better by cooperation than by competition between them. There may be scale economies and limited competition in the provision of specialist services like radiology and pathology. Limited competition reduces the incentive to produce services at minimum cost and may reduce the range of services available.

Social welfare in health care. Access to health care is widely perceived to be an individual right. In the words of the constitution of the World Health Organization (1948):

> The enjoyment of the highest attainable standard of health is one of the fundamental rights of every human being without distinction of race, religion, political belief, economic or social condition.

Such a belief may translate into the view that government should provide universal basic health care services to all in need. At the very least, it implies that government should provide such services to those who cannot afford to pay for them. Social welfare lies at the heart of the case for public funding of health care services.

The idea that all individuals should be assured of a basic service in a specific area, such as health or housing, is sometimes called specific or commodity egalitarianism (Tobin, 1970). The objective that each member of society should be assured of a minimum level of health care extends beyond the idea that each member should receive a minimum level of income.

Role of government. The welfare objective that everyone should have access to basic health care services implies, as a minimum, significant public funding of these services for low-income households. As Cutler (2002a) observed, 'the central question facing governments is how to design a medical care system for the poor'. However, given the high cost of many medical treatments and the uneven distribution of health, many communities would extend the welfare objective to assisting most, if not all, people in serious ill health. Of course, this does not necessarily imply that government should be involved in the physical delivery of health care services.

Analysis of market failures also suggests that government has major roles in funding health care services and in regulating certain goods and services. Markets are likely to undersupply services that provide public health benefits and health insurance. On the other hand, markets may oversupply services where there is supply-induced demand. Also, some markets may not supply health care services efficiently due to a lack of competition. In such cases government may be able to use its monopsony (buying) power to drive down prices for the benefit of

[2] Alternatively, health insurance firms may develop strategies for excluding high-risk individuals. In so far as these 'cream skimming' policies are successful, the firms will not offer insurance to high-risk consumers.

consumers of health goods and services. However, the complexity of the market failures is such that policy responses need to be thought through quite carefully.

Optimal Provision of Health Care Services

In most OECD countries it is accepted that basic health care services should be supplied to all those in need of them physically and financially. However, terms like basic health services, need and supply can be interpreted in many ways. Basic health services and need are relative concepts that change with income, technology and costs. Supply of a service is a function of price as well as the availability at a convenient time and place. Inevitably, views on appropriate service levels vary.

However, resources are limited in health as elsewhere and hard choices must be made regarding the allocation of scarce resources.[3] Should funds be moved from schools to health services? How should a public health authority allocate resources between heart bypass operations, hip replacements, nursing services for patients disabled by strokes and so on? Should health care providers give priority to services for young, basically healthy, people with a long life ahead over services for elderly people in pain but with limited life expectancy?

Estimates of the costs and benefits of health care services provide important, if not decisive, inputs to answers to these questions.[4] Like other services, health care services should be provided when the marginal benefit of a service exceeds the marginal cost. Conversely, services should generally not be provided when the marginal cost exceeds the marginal benefit, unless equity considerations are deemed to override any such net cost. However, as will be seen, equity is often built in to estimates of the benefits of health care services.

Generally, the average and marginal costs of health care services can be estimated with reasonable accuracy, for example with activity-based costing as discussed in Chapter 16. The major benefits of health care services are the sum of the personal benefits of improved health outcomes, including work and lifestyle benefits, and reductions in future health care costs. Of course, to put values on health care services, we must first have reasonable estimates of the relationship between health care services and health outcomes. These data would normally be provided by clinicians or epidemiologists. Here we focus on the valuation of health outcomes.

Valuing health states

Following standard economic principles, we seek to estimate what individuals are willing to pay for improved health states. However, we face a major problem in trying to estimate willingness-to-pay (WTP) values for improved health because of the lack of prices and other market data.

Traditionally analysts often employed the human capital approach. Under this approach, health improvements were valued by increased earnings. However, this is wrong for several reasons. Fundamentally it is not a measure of what individuals are willing to pay for improved health. It ignores any benefits to people outside the workforce. It also ignores any relief of pain and suffering and gains from improved lifestyle of people within the workforce.

To-day, health states are valued in two main ways. We can estimate directly what people are willing to pay for health or we can adopt an indirect two-step process. Using direct

[3] A feature of health spending is that it is highly variable across persons. As Cutler (2002a) points out, in the United States one per cent of the population consumes 30 per cent of all medical care resources and the top 10 per cent consumes about 70 per cent of resources.

[4] Because benefits are hard to measure, many health care services are justified on cost-effectiveness grounds – that a given outcome can be achieved at lower cost by one program than by another. However, cost-effectiveness analysis provides limited guidance to resource allocation decisions.

methods, we can draw on revealed or stated preference methods of valuation described in Chapter 11, where we also described how these methods could be applied to estimate the value of life. Adopting indirect methods, we first establish the relative disutility of health states (such as a broken leg or angina) in terms of a quality of life index and then apply a monetary metric to this index to produce monetary equivalents for each health state. We describe the direct methods first as they reflect more closely economic valuation principles. However, the indirect two-step process provides a practical alternative approach.

Direct methods of valuing health states. Revealed preference (RP) studies infer values from observations of individual behaviour. A common form of RP study in the health sector is hedonic wage–risk analysis that infers the value of life from the relationship between wages and risks of fatal injuries in different occupations (see Box 11.2, page 193). There are fewer studies of the relationship between wages and risks of non-fatal injuries. However, all wage–risk studies are constrained by weak perceptions of risk by economic agents (Jones-Lee and Loomes, 2004). There have also been hedonic studies of product prices which infer the value of life or other health outcomes from trade-offs of price and risk implicit in, for example, motor vehicle prices, the purchase of home smoke detectors, the use of bicycle helmets and house price responses to hazardous site risks. For example, Andersson (2005) analysed the price premiums that Swedish consumers were willing to pay for safer motor vehicles and estimated that the value of a statistical life (VSL) was between US$1.0 million and US$1.5 million in 1998 prices, which was significantly lower than the value that had been inferred from several other American and Swedish studies.[5]

Stated preference (SP) methods derive estimates of WTP values from individual responses to survey questions. SP studies can address the relevant context directly, are flexible and can present rich information sets. Using the contingent valuation approach, following various contextual questions, individuals are asked simply to state their values for hypothetical goods. For example, Ho *et al*. (2005) asked individuals what they were willing to pay for pain relief from permanently disabling occupational injuries. However, individuals find it hard to provide direct WTP responses to unfamiliar or complex options. On the other hand, the provision of monetary cues, such as a set of dollar amounts to choose between, tends to bias the results. And sometimes respondents object to saying how much they would pay for services that they consider should be free.

Consequently, many researchers have turned to choice modelling, especially discrete choice experiments (see Chapter 11). In these studies, respondents are asked to choose between alternatives including health outcomes that are characterised by various attributes, including a monetary attribute. The analyst models the probability of respondents choosing one option compared with another, or compared with all other options, in terms of the attributes of the options, including the dollar cost associated with the option. Valuations of outcomes are inferred from the monetary trade-offs implicit in the choices. Johnson *et al*. (1999) and Tsuge *et al*. (2005) are examples of discrete choice experiment studies. Johnson *et al*. (*ibid*.) estimated what individuals are willing to pay for improved respiratory and cardiovascular health, including small changes in conditions. The study aimed to obtain estimates of the dollar amounts that individuals would be willing to pay to avoid various specified injuries. Tsuge *et al*. (2005) used a choice model to value reductions in mortality risks due to accident, cancer and heart disease and estimated a VSL of US$2.9 million. Carson and Louviere (2010) and Louviere and Fiebig (2010) describe various other discrete choice studies.

[5] The estimated value of life is generally termed "the value of statistical life" because the studies estimate the willingness-to-pay values that will result in the saving of one life but no particular life.

Indirect quality of life index methods. Health states are often measured by a quality of life (QoL) index where 1 equates to perfect health and 0 to a state equivalent to death. Thus a migraine headache may have a QoL score of say 0.85, severe angina a score of 0.60, inoperable lung cancer a score of 0.35 and so on.[6] These scores may be obtained from surveys of medical experts or from the general public. There are three main elicitation methods. One is simply to ask respondents to rate each health status on a scale of say 100 to 0 (or 1 to 0), where the top figure represents perfect health. Another method involves a time trade-off. For example, individuals may be asked how many extra years of life would compensate them for loss of an arm. Suppose that someone considers that life without one arm for 10 years is equivalent to a perfectly healthy life for 8 years. Then, if perfect health is 1.0, the valuation of a health state without one arm is 0.8 (the product of the life years and the health state index are the same in both cases—8 years). The third approach is to ask people to choose between a certain poor health state and a prospect of perfect health with a probability of p and immediate death with a probability of $1 - p$. If a person is indifferent between the poor health state and the risky prospect when p is say 0.7, the poor health state would have an index value of 0.7. Such methods have been widely used to develop indices of health states (see Mathers *et al*., 1999).[7] The normal presumption is that the index is arithmetic. Each one-point change in the index has equal value.

To estimate the value of a health state economists then draw on the concept of the value of a healthy life year (VLY). VLY is usually derived indirectly from the estimated value of life (VSL) rather than from a separate revealed or stated preference study. Most often VLY is taken to be the constant annual sum which, taken over a remaining life span, has a discounted value equal to the estimated VSL. For example, if the VSL for healthy persons with a life expectancy of 40 years is $4.0 million, then applying a private time preference discount rate of 3 per cent, the value of a healthy life year would be about $173 000.

The values of various health states can then be derived as the simple product of the estimated QoL index value and VLY. If VLY equals $173 000, the value of a year with a QoL equal to 0.8 is $138 400. Conversely, avoiding the impairment for a year would be worth $34 600 (0.2 × $173 000). This approach can be used to value acute (short-term) as well as chronic health conditions. If someone has an illness for seven days with a QoL measure of 0.8, the cost of the illness would be $663 (0.2 × $173 000 × 7/365). Abelson (2003) provides a summary of estimated costs of various acute and chronic morbidities for Australia.

Summary. The use of WTP values from stated preference methods or quality of life indices informs the debate about the allocation of resources for health care services. Of course, willingness to pay is a function of income, which means that low-income households may not be willing to pay average WTP values. To ensure that health care services are provided equitably and not biased towards more affluent households or areas, health economic evaluations commonly adopt average WTP values. These can be compared with costs to aim for efficient and equitable service outcomes.

Financing Health Care Services

Most OECD countries fund a wide range of health care services for all citizens. In most cases the services are provided universally without a means test, subject to rationing and queuing, at low or no cost to the user. The United States is a major exception, where public insurance has been traditionally means or age tested and is far from universal.

---

[6] See Abelson (2003) for a summary of QoLs for morbidities and values attached to them.
[7] In Mathers *et al*. (1999), perfect health is 0 and a health state equivalent to death is 1.

Most publicly funded services are financed by general taxation or social insurance. Despite the difference in funding source, both are often described as public insurance. Australia, Canada, Sweden and the UK finance health services mainly from taxation. France, Germany and the Netherlands have national social insurance schemes. Where applicable, social insurance is generally compulsory. Unlike income tax, social insurance is usually levied on labour earnings rather than on total income. Unlike private insurance, social insurance is not based on risk factors and the services provided are usually independent of contributions. However, social insurance is sometimes combined with private insurance, as in Japan and the Netherlands where some form of health insurance is compulsory but private insurance is an option. In such cases the private health funds are substitutes for public funds.

Most private payments for health care services are payments by users made directly to service suppliers or payments by private insurance funds. In most countries, including Australia as we have seen, direct private payments are much greater than private insurance payments (Wagstaff and Doorslaer, 2000). Direct payments for health services include payments that the government or private health funds require users to contribute towards the service. These payments are usually co-payments—fixed amounts that a patient pays for a service before receiving reimbursement. Under co-insurance a patient pays for a percentage of the cost. It is also possible to combine co-payment with a share of the cost in excess of the co-payment. Co-payments are low in Canada and the United Kingdom. They are higher in New Zealand and some are high in Australia. Co-payments also vary with health care service. They tend to be a small proportion of the cost of hospital services, a slightly higher proportion for medical services and a still higher proportion for medications, dental care and nursing homes. For medications, co-payments are typically 30–40 per cent of costs in OECD countries.

Health care financing and equity. To determine whether the method of funding achieves equity objectives, we need to define equity criteria and to consider specific health systems in practice. In terms of equity criteria, the ability-to-pay (and the related vertical equity) principle suggests that higher-income individuals should pay more for health services than low-income individuals. But the implications for service delivery are not clear. For example, it could be further argued that low-income households should receive service priority from publicly funded health services because higher-income households can obtain private services. In practice, service delivery is rarely means tested. Most health agencies determine service priorities on a health-needs basis. Those with greater need have greater service priority in terms of timing and resources. This meets the horizontal equity principle of equal service for equal health need.

In a comprehensive review of equity in the provision of health services, Wagstaff and Doorslaer (*ibid*.) found, as might be expected, that the most progressive way to finance health care services is from consolidated revenue. Social insurance is progressive when it is universally compulsory, but regressive when high-income households can opt out. Private insurance is regressive when most of the population relies on it; that is, the cost of health services rises less than proportionately with income. It does not mean that a high-income person pays less for a service than does a low-income person.[8] However, in the United States the financing system has been not only relatively, but also absolutely, unfair. Before the 'Obamacare' changes, employed persons, mainly upper and middle-income persons in large companies, received an annual $200 billion private insurance subsidy via corporate tax deductions, whereas 47 million mainly lower-income Americans had no private or public health insurance (Gruber, 2011, 3rd ed., Chapter 15). Even post-Obamacare, 43 million in the US were uninsured (Gruber, 2016, 5th ed. page 447)

[8] Chapter 25 defines the terms progressive, proportional and regressive. A financing system may be regressive even though high-income individuals pay more for services than do low-income individuals.

Wagstaff and Doorslaer (*ibid*.) also found that equity in service delivery is related to the financing system. Countries with a high level of public finance, such as in Canada, are more pro-poor in service delivery than are countries with a lower level of public funds, such as the United States. On the other hand, although government funds nearly 70 per cent of health care expenditure in Australia, the location of public as well as private health care facilities is biased towards high-income areas. Also, as we see below, many of the public subsidies for health care in Australia principally support middle- and upper-income households.

Health care financing and efficiency. The financing of health care can also affect the quantity, quality and costs of services provided. The principal effect is through user charges for services (or the lack of them). Without user charges there is no price restraint on demand. Nor is there any consumer monitoring of costs. Thus, when government or a private health insurer pays all or most of the cost of a health care service, there is likely to be both excess demand for the service and excess unit cost.

The lack of a price system is a major reason for the chronic inflation of expenditure of health care services. In particular, what matters is the marginal payment for service. Cost sharing has no impact on expenditure if it does not influence choice of service. Where someone pays a fixed cost co-payment for a medical service or medication that is below the price, she has little incentive to limit their marginal consumption or costs.

The lack of pricing often gives rise to the risk of moral hazard. Moral hazard occurs when insurance (public or private) causes someone to consume more health care services than they would if they were paying for the full cost of the service. This may occur because insurance reduces a person's care to avoid accidents or ill health or, more often, because he or she consumes more health services for a given level of health than he or she otherwise would. The more generous is the insurance, the greater is the moral hazard. In the mid-1980s the US Rand Corporation ran an experiment in which insured individuals were randomly assigned to different levels of co-payments. Higher co-payments led to significantly lower use of medical care with generally no adverse outcomes, though there were some reported exceptions. The estimated elasticity of use of medical care with respect to price was a modest −0.2, but taken over the whole US market indicated a deadweight loss of between $125 and $400 billion a year (Gruber, 2008, 2nd ed.).

Without a price mechanism and incentives that lead individuals to choose an appropriate product mix and to contain costs, government and private insurers must find other ways to constrain the use and costs of health services. Governments often adopt a supply-side strategy. For example, Australian governments limit the funds available to hospitals and restrict the number of doctors. However, expenditure is a product of quantity of services and unit costs. Health care suppliers may respond to financial constraints by rationing the quantity of services or reducing their quality rather than by reducing unit costs. Hospitals may ration services by increasing waiting times or by reducing the quality of their services, for example by restricting time in hospital. Doctors may increase waiting times or reduce service times or, in some systems, increase the prices of non-controlled services. These responses are inefficient because they result in a less preferred quantity and mix of services at higher prices than would occur in a funding system based on price for service.

In a competitive health insurance market, private health funds have incentives to constrain costs on behalf of their members or they would lose market share and profits. In the United States over 90 per cent of private insurers have adopted some form of 'managed care' strategies (Cutler, 2002a). These insurers provide their members with a comprehensive package of services available generally from a defined network of providers for a fixed payment. Preferred provider organisations selectively contract with a network of hospitals, doctors and others to provide services at a discounted price schedule. If members visit providers who have agreed to accept a lower price, they incur lower co-payments. Given the

transformation of unmanaged fee-for-service insurance organisations into managed care practices, it may be inferred that they have been quite successful at controlling costs. On the other hand, the level of health care expenditure in the United States is well above any other OECD country.

Where private health insurance is publicly protected, regulated and subsidised, as in Australia, insurance organisations have less incentives to control costs (see below). Moreover, subsidies for private insurance may lead to excess insurance and hence to excess use of medical services. Gruber (2016, 5th ed.) reported that the high level of tax subsidy for private insurance in the United States leads to a rise in health care spending with an elasticity of spending with respect to the fall in price of insurance (the tax subsidy) of −0.7.

Delivery of Health Care Services

Given the quantity of health services to be supplied, however determined, what is the most cost-effective method of service delivery? As we have seen, government often attempts to contain costs by capping expenditures. However, this may reduce the quantity of service and have little impact on unit costs or productivity. Certainly, when there is supplier-induced demand, some controls on suppliers may be necessary. However, it is generally more efficient to control the variable of concern (in this case, outputs) than to control the number of suppliers of these outputs, which reduces competition and variety to service. Here we focus on four complementary strategies for containing unit costs.

The first strategy is separation of service funding and delivery (or equivalently, development of a purchaser–provider framework). The aim is to separate the decision on the amount and type of services to be purchased from the way the services are supplied. In the UK, government provides funds to area health services and to general practitioners, who then purchase the services that they consider are needed most urgently from the most efficient possible source. Although the funds typically come from taxation, the fund holders can purchase services from private agencies. In Australia, purchasing and supply are separated in the general medical practice market, where government and households purchase services supplied by general practitioners. However, in the hospital sector, area health services are often purchaser and provider of health care services.

The second strategy for efficient production is competition in service supply. The purchaser should be able to buy services from the most cost-effective supplier meeting quality standards. Competition reduces costs in the health care sector as in others. Hospitals provide a wide variety of health care and accommodation services. In Chapter 16 we noted significant savings from contracting out hospital cleaning services. Savings may be achieved from contracting for other hospital functions, including hotel functions (food and laundry), nursing, radiology and pathology services and so on. Even when there are natural economies of scale or scope, competition for service provision may reduce supply costs. There is some evidence that competition has controlled costs of hospital services in the United Kingdom.

However, it is important that the health care services can be monitored, especially for quality. The quality of medical services is not readily observable. When quality is not observable, competitive markets may not produce high quality services. Indeed, a profit motive may be inimical to quality if a for-profit health care provider skimps on care because skimping cannot be detected. Of course, monitoring quality of service is also hard in public institutions. Nevertheless, the issue of monitoring means that outsourcing to private agencies is not always the best policy.

Third, suppliers of health care services should be paid as accurately as possible for services provided. Payments can be based on services for diagnosed medical conditions and standardised treatment costs for these services. In Australia payments for hospital services are based on treatments for over 500 kinds of services based on diagnostic-related groups. Such

case-mix payments encourage service providers to maximise the real value of services provided and to minimise the use of inputs. By contrast, block grant funding of institutions provides few incentives to reduce costs. Indeed, institutions are often penalised for not spending their full allocation. Also, crude service measures, such as hospital bed days, provide perverse incentives that encourage longer stays in hospital than necessary.

Fourth, governments need to establish sensible rules about the handling of medical risk. Fear of litigation has escalated the costs of professional indemnity insurance and the quantity of tests and other services provided to patients, especially when the tests are subsidised. Increasingly, medical practitioners are practising defensive medicine, for example doing or ordering tests that are likely to have little informational value. Unfortunately, it is always possible that some action or inaction will turn out to be the wrong one. Trying to eliminate all such error, often for legal reasons, is a highly expensive process.

Health Care in Australia

A major feature of the health system in Australia is universal access to public hospitals and medical services free of charge. There is no patient charge for medical services or accommodation for public patients in public hospitals and anyone can elect to be a public patient regardless of whether they hold private health insurance. In some states, individuals can also choose private accommodation and select doctors of their own choice in public hospitals.

People receiving medical services outside public hospitals can obtain reimbursement equal to 85 per cent of Medicare Schedule Fees (MSF) for the services. This is formally done through Medicare and described as a Commonwealth government health insurance system, but in effect it is funded from tax revenues. Actually, medical practitioners can charge any fee they wish. The patient pays the extra amounts that are not reimbursed by Medicare. However, the Commonwealth pays doctors an incentive fee to bulk bill directly to Medicare at the 100 per cent MSF rate in the case of general practitioners (GPs) and the 85 per cent rate for specialists. About three-quarters of GP services are bulk billed to the government, in which case the consumer pays nothing.

Medications are also subsidised. Ninety per cent of prescription drugs sold outside hospitals are approved for public subsidy under the Pharmaceutical Benefits Scheme (PBS). To obtain approval for listing a drug on the PBS, the manufacturer of the drug has to provide evidence that it provides a clinical benefit not otherwise available at a reasonable cost or that it provides a benefit more cost-effectively than other drugs (which is essentially a form of economic evaluation). The level of subsidy varies with the category of consumer, being greatest for concession cardholders. The Commonwealth reimburses pharmacies for the difference between their estimated costs of dispensing a drug (its price to pharmacy plus a dispensing fee and mark-up) and the patient contribution. Concessional beneficiaries and pensioners, who account for 80 per cent of total PBS payments, currently pay about $5 per item up to a ceiling of about $180 per annum after which prescriptions are free. These figures alter annually.

There are 35 private health insurance providers registered under the *National Health Act*. Registered insurance funds are required to write a Basic Table containing a standard set of benefits prescribed under the *National Health Act*. Depending on the fund's rules and the membership category to which a household signs up, these funds pay benefits on some of the costs of private hospital care, inpatient medical care, dental, optometry and physiotherapy expenses and even some ancillaries like gym membership. Government encourages membership of private health funds by providing a subsidy on health insurance premiums of up to 33 per cent depending on individual ages and individual or family income, levying a tax surcharge on high-income earners who do not have private health insurance, and allowing

insurance firms to charge a premium for people who commence insurance at older ages. These fiscal initiatives raised the proportion of households holding private health insurance from about 30 per cent in 2000 to about 40 per cent today. Nevertheless, payments by private insurance firms amount to only 7 per cent of total health payments.

As in other countries, real health expenditures have grown rapidly over the last 20 years. Admissions to hospitals have increased twice as fast as the population. Actually, bed days in hospitals have not increased because of falling lengths of stay in hospital, due to increased one-day surgery and technological advances. But the cost per bed day has risen because of the cost of new technology. Visits to doctors have increased from under five per person per annum in 1970 to about 12 per annum today. Consumption of medications that are listed on the PBS and whose cost is substantially subsidised by government has risen at an especially high rate. Since the mid-1980s, real pharmaceutical expenditure borne by government has increased by 7 per cent a year.

Some economic issues with Australian health care

Australians enjoy high longevity. Between 1970 and today average life expectancy at birth rose extraordinarily, from 67 years to 79 years for males and from 74 years to 84 years for females. Households receive high quality health care services in most urban areas, although there can be long waiting times for non-urgent elective surgery. Low income is generally not a barrier to access to public services. Private services are available to those who wish to pay for them.

However, low-income households have higher mortality rates, greater incidence of sickness and greater prevalence of disability (AIHW, 2006). Aboriginal people experience especially poor health, with life expectancies 15–20 years below those of other Australians. Inequity in outcomes is reflected in inequity in access, with health care services concentrated in higher-income areas. There are more medical services and hospital beds per capita in inner metropolitan areas than in outer areas and many more doctors and hospital facilities in urban areas than rural ones. In addition, higher-income earners are the major beneficiaries of the private health subsidies.

The geographical imbalance in the provision of health care services reflects a conflict between equity and efficiency. The imbalance within the cities reflects the historic creation of health care assets in the inner areas and the preference of clinicians for inner city locations. Efficiency requires that these assets be fully utilised. Also, economies of scale in the cities and the high costs of attracting medical resources to rural areas mean that health services can be supplied more cost-effectively in urban areas. However, the uneven supply of health services increases the distance to hospitals and specialist clinical services for low-income households in cities, and even more so for households in rural areas.

Four efficiency issues. Turning to efficiency, four sets of issues may be highlighted. These relate to the lack of user charges and the implications for allocation of resources, cost containment, private health insurance and Commonwealth–state responsibilities.

Any system that provides extensive services with very limited user charges (price signals) is likely to result in a misallocation of resources, including excessive consumption (with marginal cost exceeding marginal benefit), a misallocation of resources from high needs to low ones and lack of cost controls. Low or no charges for outpatient services at hospitals, GP services, mental health (clinical psychology) services, diagnostic testing and pharmaceuticals encourage excess consumption of these services and goods. Some people view GPs as free counselling services and GPs have no incentive to discourage use of their services. Services are rationed by queuing and available to those with lowest time costs rather than those with most urgent needs. When the Australian government introduced highly subsidised mental health counselling services in 2006 it allowed patients up to 18 sessions per year and

budgeted about $150 million per annum. By 2011 the budget was over $1.0 billion per annum (due to high demand at very low effective prices) and the government responded by limiting patients to a maximum of 10 sessions in a year. Public subsidies for pharmaceuticals more than tripled from $2.7 billion in 1999 to $8.4 billion in 2009–10, almost certainly indicating excessive prescriptions and consumption with prices to consumers far below real cost. Excessive hospital admissions are less likely as people do not lightly enter hospitals even when free. However, once in hospital there is no price mechanism for the control of service provision.

Turning to the costs of health care services, governments adopt various policies to contain costs. One particularly effective one is the Commonwealth's use of its monopsonistic purchasing power to control the price of pharmaceuticals that it chooses to subsidise under the PBS. Also, most of the states and territories have adopted case-mix funding strategies for hospitals, whereby hospitals are funded for services provided. This is more efficient than the traditional annual block funding which is less accountable and encourages hospitals to spend unused funds at the end of the year. Another Commonwealth policy, to control supplier-induced demand by restricting the training and supply of doctors, is more questionable. Many able young persons who would like to become GPs or medical specialists cannot do so despite the high demand for medical services (admittedly inflated by the Medicare arrangements) and a shortage of doctors in many areas of the country. On the other hand, Australian governments make limited use of the purchaser–provider principle. Area health services are often purchaser and provider of services. Related to this, governments usually fund services from public hospitals rather than from private hospitals and there is little price competition between hospitals. Also, public hospitals tend towards employing in-house services to supply the services that they fund.

With regard to the private health insurance market, there are two key issues. What are the objectives? And, how can these objectives be achieved alongside a free public insurance health system? Clearly a major objective of private insurance is to enable individuals to make their own choices about health care services. Another objective is to reduce the expenditure on public health services and so reduce taxes. The first of these objectives is compromised somewhat by the Australian government's requirement that private health premiums be community rated. This means that insurers must offer premiums at uniform prices to all clients of given ages regardless of health status. This restricts the supply of insurance options and makes private insurance costly and unattractive to more healthy individuals.

On the other hand, to achieve the second objective (cost minimisation to the public sector) the government needs to encourage or force as many people as possible into private insurance. It does this by the various fiscal instruments (subsidies and taxes) mentioned above. However, the subsidy for private health fund insurance is now about $5 billion, which are resources lost to public health services and is a high cost to pay. The subsidy has no impact on the behaviour of most insured households who would hold private insurance even without the subsidy. Also, many holders of private insurance continue to use free public facilities (to which they remain entitled), especially for more severe illnesses. On the other hand, the subsidy has not contained the cost of private health insurance as the private health funds have increased premiums well over the consumer price index since the subsidy was introduced, which suggests that the subsidy may support inefficient production. And overall the subsidy is regressive.

Evidently it is hard to run an efficient private health insurance system alongside a free public health insurance system. This led the National Health and Hospitals Reform Commission (NHHRC, 2009) to propose that individuals be allowed to opt out of the Medicare public health insurance and receive a subsidy to enrol in a private health insurance plan of their choice (apparently without community rating). The subsidy would be risk adjusted to reflect the expected cost of services they would have received if they had

remained in the public health system. This policy (called Medicare Select) would aim to achieve the benefits of competition in both the insurance and the service market. Butler (2010) provides a useful summary of the potential benefits and costs. While this scheme has some attractions, such a radical change is rarely enacted as planned and it is not presently under consideration.

A final area of inefficiency is the overlap between Commonwealth and state responsibilities, which leads to costly duplication of services and cost-shifting inefficiencies. The Commonwealth pays subsidies for nursing homes whereas the states bear most of the marginal costs of running public hospitals. To economise, the Commonwealth restricts the supply of nursing homes. This increases length of stay in public hospitals where services are much more expensive. Within the GP sector, the Commonwealth pays for doctors' services delivered outside hospitals, whereas the states pay for similar services in emergency departments in hospitals and in community health centres. To increase standard GP services and shift costs to the Commonwealth, the states and territories have closed several hospital emergency departments and limited services in community health centres. While some of these moves may have been cost effective, others were not. But the key point is that the major incentive is to shift costs rather than necessarily to minimise them.

In order to better integrate all levels of care and to avoid these cost-shifting problems, the NHHRC (2009) proposed that the Commonwealth pay state governments 100 per cent of the efficient costs of hospital services derived from activity-based costing of case-mix classifications. Subsequent Commonwealth–state discussions fell down on two major issues. One was the source of the additional Commonwealth funding which the Commonwealth proposed would be obtained by withholding 30 per cent of GST revenues from the states. The second issue was hospital management which the Commonwealth proposed to take over via local area health services. Not surprisingly the states did not agree these proposals.

However, the August 2010 National Health Reform Agreement introduced major reforms to the organisation, funding and delivery of health and aged care. The main feature of the Agreement is an activity-based Commonwealth contribution using 'efficient' prices for hospital services determined by the newly formed Independent Hospital Pricing Authority. Under the Agreement, the Commonwealth will fund 50 per cent of the efficient costs including growth in costs, while the states and territories are responsible for the remaining costs including residual costs. This should increase transparency and accountability.

Concluding observations. Designing an equitable and efficient health care system is one of the greatest public policy challenges and it is certainly easier to criticise any given institutional framework than to design the perfect solution. Nevertheless, our discussion of both the general principles of health care and the Australian system suggest some general lessons. It seems likely that most health care systems would be more efficient, without loss of equity, if they distinguished more between the purchasers of services and the suppliers, made more use of prices and co-payments to allocate services and constrain costs, and incorporated more competition into the supply of services.

Summary

- In most societies, social welfare is the primary motivation for government funding of health care services.
- Market failures include the public good (positive externality) features of health care, limited competition in some forms of health care services and asymmetric information.
- Efficient health care requires that services should be supplied at least cost when the marginal benefit of a service exceeds the marginal cost.
- Average and marginal cost of health care services can usually be estimated accurately. Benefits are more difficult to measure. Economists have developed methods for estimating the value of life and of health states using revealed and stated preference methods.
- Health care systems with high public funding usually produce more equitable outcomes than systems with low public funding.
- But, health care systems without user charges or low co-payments lead to over-consumption of health care services. Subsidies for private health insurance also increase health care spending.
- Services are delivered most cost effectively when the roles of service funding and delivery are separated, there is some competition in service delivery and suppliers are paid for services rather than given block revenue grants.
- Australian health outcomes are generally relatively good but low-income households and Indigenous people have significantly worse outcomes and poorer access to services.
- There are four other significant unresolved issues in the Australian health care system: over-consumption of services associated with low or zero user charges, the role of private health insurance alongside a dominant public health system, containing costs in a largely uncompetitive system, and inefficiencies arising from mixed Commonwealth–state financing and management of the diverse services in the health care system.

Questions

1. Does health care expenditure fall in proportion to GDP as the share of public sector in funding health care rises? If so, what are the most likely explanations? How would you find evidence for your conclusions? What welfare conclusions would you draw from the analysis?
2. One possible market failure is the allegation that the suppliers of health care services may create demand. Therefore an increase in the supply of doctors may increase the quantity of services and have no effect on the price of services. It is then argued that, to control health care expenditures, it is necessary to control the quantity of doctors. Is this a valid argument?
3. Why is rationing excess demand for medical services by queuing an inefficient outcome?
4. What are the main strengths and weaknesses of the methods that economists have designed to estimate the benefits of health care services? What are likely to be the main areas in which these valuations can be used?
5. Suppose that the number of patient visits to a doctor per annum is related to price as follows:
   $Q^d = 10 - 0.15P$
   i. If the price of a consultation is \$40, how many visits would be made per annum and how much would be the expenditure?
   ii. Now suppose that government pays \$30 for each visit, so that the cost is \$10 per visit. How many visits would be made per annum and how much would be the expenditure?
   iii. What is the deadweight loss of the subsidy?
6. Suppose that your utility function were $U = \ln(Y)$ and that your regular income is \$80 000 in a year. There is a 2 per cent chance that you will be sick and lose \$40 000 income.
   i. What is your expected utility if you are healthy or sick?
   ii. What would be the actuarially fair premium? What is your expected utility if you buy the insurance policy?
   iii. What is the most you would be willing to pay for this insurance policy?
7. What kinds of inefficiencies arise with private health insurance in general?
8. The Australian government provides a subsidy (currently 30 per cent) for private health insurance. What are the efficiency and equity effects of this subsidy?

9. What are the key principles of efficient production and why are they difficult to achieve in the health care sector?
10. What is defensive medicine? Why does it have potential welfare costs? How could public policy deal with the problems that arise with defensive medicine?
11. What problems arise with the delivery of health care services in a federal government system? How might these problems be overcome, or at least minimised?

Further Reading

Abelson, P. (2003) 'The value of life and health for public policy', *Economic Record*, 79, Special Issue, 2–13.

Butler, J.R.G. (2010) 'New opportunities in health financing and governance', *Australian Economic Review*, 43, 71–76.

Cutler, D.M. (2002a) 'Health care and the public sector', pp. 2143–2243, in *Handbook of Public Economics*, Vol. 4, A.J. Auerbach and M. Feldstein (eds), Elsevier Science, Amsterdam.

Cutler, D.M. (2002b) 'Equality, efficiency, and market fundamentals: the dynamics of international medical-care reform', *Journal of Economic Literature*, XL, 881–906.

Dolan, P. (2000) 'The measurement of health-related quality of life for use in resource allocation decisions in health care', pp. 1724–1760 in *Handbook of Health Economics*, Vol. 1, A.J. Culyer and J.P. Newhouse (eds), Elsevier Science, Amsterdam.

Drummond, M. (2004) 'Economic evaluation in health care: is it really useful or are we kidding ourselves?', *Australian Economic Review*, 27, 3–11.

Folland, S., Goodman, A.C. and Stano, M. (2007) *The Economics of Health and Health Care*, Pearson Prentice Hall, New Jersey.

Gruber, J. (2008) 'Covering the Uninsured in the United States', *Journal of Economic Literature*, 46, 571–606.

Gruber, J. (2016), *Public Finance and Public Policy*, 5th ed., Chapters 15 and 16, MacMillan Education.

Jones, A. (ed.) (2006) *The Elgar Companion to Health Economics*, Edward Elgar Publishing, Cheltenham, UK.

Joumard, I., Christophe, C. and Nicq, C. (2010) *Health Care Systems: Efficiency and Institutions*, Working Papers No. 769, OECD Economics Department, OECD, Paris.

Richardson, J. (2010) 'Is there a better alternative for Australia's health system?', *Economic Papers*, 29, 267–278.

Rosen, H. and Gayer, T. (2016) *Public Finance*, 10th ed., Chapters 9 and 10, McGraw-Hill, New York.

Part

8

Public Finance and Taxation

Chapter

25

Introduction to Public Finance

And it came to pass in those days, that there went out a decree from Caesar Augustus, that all the world should be taxed.

Luke, 2:1

Sources of Public Finance ♦ Basic Elements of a Tax System ♦ Tax Hypothecation ♦ Tax Avoidance and Evasion ♦ Evaluating Tax Systems

In this introduction to public finance we discuss five main topics. The first is the sources of public finance. Of these, easily the most important is taxation. Second, we discuss the basic elements of tax systems, namely tax bases, tax rate structures and tax units. Third, we discuss hypothecation, a process that earmarks certain tax revenues for specific expenditures. However, notwithstanding Benjamin Franklin's gloomy observation that "nothing in this world is certain except death and taxes", many people try to escape taxes in one way or another. Accordingly, we also briefly discuss tax avoidance and evasion. Finally, we discuss the core policy issue in public finance: how to evaluate and choose between alternative taxation systems. Consistent with our earlier discussions, we look for a tax system that is efficient and equitable. In this chapter, we discuss what this may mean. In the next chapters, we take up these and other issues in more detail.

Sources of Public Finance

Government can acquire control over resources in three main ways: by compulsory means, by market transactions of various kinds or in diverse other ways. The main instruments under each of these three headings are shown in Table 25.1.

Compulsory acquisition. This is the basis of public finance. Government can compulsorily acquire access over resources in various ways. For example, government can conscript labour. Many countries conscript labour for military service without paying market wages to the coerced labour. Australia did so in the Second World War and again for the war in Vietnam in the 1960s and early 1970s. Government can also compulsorily acquire physical resources such as minerals and land or their income flows. In Australia, government can acquire property for public purposes. However, under the Constitution, government must pay fair compensation, which limits its ability to finance expenditure by appropriating property.

Taxes are payments that are collected compulsorily from individuals or companies by government for which nothing is provided directly in return. Taxation in all its forms, including

Table 25.1 Sources of public finance

Compulsory acquisition	*Market transactions*	*Other sources of funds*
Labour conscription	Borrowing (market loans)	Sale or lease of assets
Taxation including royalties	User charges (competitive prices)	Sale of inventory
Social security contributions	Some fees for services	Bringing income forward
Printing money	Interest income on government loans	Deferring payment
User charges (single supplier prices)	Dividends from shareholdings	Taxing public enterprises
Fines and licence charges		Revenue from central bank

compulsory social insurance contributions, usually accounts for 80 per cent of all government revenue in OECD economies.

Contributions to national social security funds are generally compulsory levies on earned income. Where these are insurance schemes that provide benefits, such as retrenchment benefits, in relation to contributions made, they are different from taxes which contribute to general revenue. However, in some cases the compulsory nature of these levies combined with tenuous links between contributions and benefits make them similar to taxes.

Government can finance expenditure by printing money or equivalently by borrowing at zero interest from the central bank. This may appear to provide citizens with free public services! Alas, this is a mirage. Printing money to finance public services or income transfers increases nominal incomes and the demand for goods. However, unless there are substantial unemployed resources, this generally leads to inflation (depreciation of the currency) and a fall in the real value of savings. The **inflation tax** is the decline in the real value of savings held in fixed monetary assets, including bank balances and government bonds. Holders of these assets suffer a real capital loss. Given a strong relationship between printing money and inflation, between 1950 and 2007 few governments in developed countries resorted to printing money to finance their activities. However, the extraordinary events of the global financial crisis starting in late 2007 led several major economies to adopt a money printing strategy.[1,2]

Inflation tax
The effect of inflation on the value of an individual's monetary assets

Other *de facto* forms of compulsory acquisition include user charges, fees and fines that government imposes as sole supplier of services where the various charges exceed the costs of service provision.

Market transactions. These are another source of government revenue. One such source is government borrowing in the market. Governments generally pay the rates of interest required by lenders. Debt finance spreads out costs over time and allows citizens to pay for services when they are provided. However, in so far as loan repayments and interest payments on the debt are funded from taxation, or at least backed by taxation, they are not wholly market driven (that is the government can generally borrow at relatively low rates).

Other forms of market income are user payments for publicly supplied goods in competitive markets, interest income on government loans and dividend payments on government equity in private companies. Dividends from publicly owned companies may also be viewed as market transactions. However, when government requires its trading enterprises to remit dividends above a normal return on capital, this is a disguised form of taxation.

[1] In exceptional circumstances of high unemployment and low aggregate demand, printing money can increase output without inflation. These issues are discussed further in Chapter 28.

[2] Some governments also adopted quantitative easing. This was the purchase of private debts with electronically created money. This is sometimes described as printing money although the object in this case was not to finance government expenditure.

Other sources of public revenue. Government can finance expenditure in various other ways. For example, it can draw on its cash reserves, sell financial assets, or lease or sell physical assets. When government finances capital expenditure by selling an asset, it alters the composition of its balance sheet (by exchanging one asset for another one) but it does not change the net worth of the public sector. However, if government sells assets to finance current expenditure it reduces its net assets and its future income.

Government may also improve its cash position by bringing income forward or by deferring payments. In 2000–01 the Australian government required most businesses to pay provisional tax on expected company income each quarter. This brought forward a large amount of company income tax payments. In effect, in one financial year the government collected two sets of company income tax from most companies. Alternatively, government may contract private firms to finance and supply public services, notably infrastructure, on a deferred payment basis. The government is effectively borrowing from the private sector to pay for current expenditure, but the debt may not show up in the government's current accounts. Of course, these 'financing methods' are accounting tricks rather than genuine sources of finance. Under a properly constructed set of accrual accounts, revenues and liabilities are accounted for when they are earned or incurred, not when they happen to be received or paid for. In the long run, taxes and user charges are the basis of government revenue (see Table 2.3 page 25).

Basic Elements of a Tax System

There are three basic components of a tax system. The tax base is the economic activity or item on which a tax is levied. The tax rate is the rate at which tax is levied on the tax base. The tax unit is the entity on which tax is levied.

Tax bases

The simplest tax base is a **head tax**, also called a **poll tax**. It is a lump sum tax levied generally on every adult citizen at the same rate regardless of income or expenditure.[3] A head tax does not affect behaviour (unless it causes emigration) or resource allocation and has no deadweight loss. However, a head tax is widely viewed as inequitable because it bears no relationship to capacity to pay tax. In 1990 the UK government under Margaret Thatcher introduced a head tax (called a community charge) for people over 18 with only a few exemptions to finance local government. After a massive outcry against the tax the government repealed it in the following year.

In practice, the two main tax bases are **income** and **expenditure** (consumption). Each of these tax bases is a rough indicator of capacity to pay. Although conceptually distinct these tax bases are closely related. Income (Y) can be represented either by its two primary sources as income from labour (Y_L) plus income from capital (Y_K) or in terms of its disposition as consumption (C) plus saving (S).

$$Y = Y_L + Y_K = C + S \tag{25.1}$$

A general tax on income is a tax on income from all sources, including capital gains. It taxes income that is allocated to consumption and income allocated to savings. A tax on consumption exempts income allocated to savings.

Wealth is another tax base. A wealth tax is a tax on the value of capital assets. It may be applied to individuals or to corporations. A tax on personal wealth is a tax on the value of assets (such as property and equities) owned by an individual less any debt. This is a tax on an

[3] Historically the head tax was sometimes applied to able-bodied adults or adults of a certain race or group or, in the United States, to people entitled to vote, hence the term poll tax.

individual's stock of net assets (K) as distinct from a tax on the flow of income from the assets (Y_K). Taxes on company assets, such as land taxes, are wealth taxes. Royalties (taxes on value of energy or mineral resources or on gross energy or mineral output) are another important form of wealth taxes. Taxes on the value of assets held by individuals or companies when they are sold, often described as stamp duties, are an indirect tax on wealth.

Direct and indirect taxes

Direct tax
A tax borne by the party that pays the tax

Indirect tax
A tax levied on one party that may be passed on to another party

Sometimes income taxes are described as direct taxes and consumption taxes as indirect taxes. A **direct tax** is a tax that is borne by the person or firm that pays it. An **indirect** tax is levied on one party but may be borne by another. Thus, a tax on a firm's product may be passed on to consumers in the form of higher prices. Also, firms pay and usually bear a proportion of payroll and social security taxes. However, if the supply of labour is inelastic, employees bear a large part of the costs in the form of lower wages (see Chapter 26).

General and partial taxes

More importantly, taxes can be general or partial. A **general tax** is levied uniformly on all components of a tax base with no exemptions or variations. A **partial** or **selective tax** is levied on only part of the tax base.

A general income tax is a tax on all forms of income from labour, capital and land. In practice, personal income tax is rarely completely general because some forms of income are tax-exempt or taxed at a concession rate. For example, imputed rent and bequests are rarely taxed and income from capital gains and retirement savings is often taxed at concession rates.[4] Also, some taxes on income are selective. For example, payroll taxes are levied only on labour income.

The most general form of consumption tax is a total **expenditure tax**. This is a tax on annual income minus savings. It works like an income tax except that, to estimate the expenditure tax base, individuals would deduct savings from their income (or add dissaving). As shown in Chapter 32, this is equivalent to exempting the return from savings (interest, dividends and capital gains) from taxation. This avoids taxing savings twice as occurs with income tax. Despite support from some prominent economists, for example Kaldor and Meade, doubtless for practical reasons no country has adopted a comprehensive expenditure tax. However, some countries, including Australia, provide major tax concessions on income from savings, such as home ownership or superannuation funds, and have in effect a partial expenditure tax.

A **value-added tax** (VAT), which is called a goods and services tax (GST) in Australia, is potentially a general consumption tax. A VAT is a percentage tax on value added at each stage of production. In most applications, especially in competitive markets, VAT is passed on and borne eventually by consumers (see Chapter 32). In the United States, most states have a retail tax on goods and services, which has a similar effect. However, value-added or retail taxes are rarely applied generally. Most governments (including Australia) exempt health, education and some food and other expenditures, which together may account for over 40 per cent of total consumption. New Zealand, with relatively few exemptions, is a notable exception. Also, most countries impose selective excise taxes on specific commodities, such as alcohol, tobacco and fuel. Tariffs on imported goods are another selective tax on consumption.

Wealth taxes may also be general or selective. A general wealth tax is a tax on the assessed market value of all assets. Selective taxes are more common. They include taxes on the value of real estate, estate (death) duties and royalties.

[4] Note that we are using here the comprehensive Haig–Simons definition of income (see Chapter 28).

Finally, two further points. First, some taxes do not fall neatly into one or other of these categories. For example, turnover taxes on gross sale income at various levels of activity may not be passed on and are hard to classify as income or consumption taxes. Taxes on use of durable goods, such as motor vehicles, and taxes on bank transactions may be regarded as taxes on consumption or on wealth.

Second, tax bases do not have fixed values. The value of a tax base varies with the tax rate. The higher the tax rate, the smaller the base is likely to be. For example, a high income tax rate is likely to reduce labour supply and therefore the income tax base. A high excise tax usually reduces purchases of the excised commodity and therefore the expenditure tax base.

Tax bases in OECD countries

Table 25.2 shows use of the major tax bases in OECD countries. Over the last 30 years, the income tax base, including payroll and social security contributions, accounted for a constant 61 per cent of tax revenues. However, personal income taxes fell from 30 per cent to 25 per cent of total tax revenues, whereas social security payments rose from 22 to 25 per cent of revenues. Also, corporate tax revenues increased from 8 per cent to 10 per cent of revenues. As shown later in Chapter 31, corporate profits have risen as a share of GDP over the last 25 years. On the other hand, corporate tax rates have fallen with increased tax competition between industrialised countries.

Among consumption taxes, the contribution of general consumption taxes has risen significantly with increased revenue from VAT in the European Union. On the other hand, there has been a relative decline in special consumption taxes, with large reductions in both specific excise taxes and import duties.

In 2009–10 Australia obtained 61 per cent of tax revenue from the income tax base. However, the composition of revenue is very different. Australia collects a high proportion of tax revenue in personal income tax and company tax. There are no social security levies.[5] Australia collects less revenue from the general (GST) consumption tax but more from excise taxes and a range of other taxes including stamp duties on capital and financial transactions.

Table 25.2 Structure of taxation in OECD countries (per cent of total revenue)

	OECD countries					*Australia*
Taxation revenue	*1975*	*1985*	*1995*	*2000*	*2008*	*2009–10*
Income tax base						
Personal income taxes	30	30	27	25	25	38
Corporate income taxes	8	8	8	10	10	18
Social security contributions	22	22	25	24	25	0
Payroll taxes	1	1	1	1	1	5
Total	61	61	61	60	61	61
Other tax bases						
General consumption taxes	13	16	19	19	20	14
Specific consumption taxes	18	16	13	12	10	13
Property taxes	6	5	6	6	5	5
Other taxes	2	2	2	3	3	7

Sources: OECD (2010) *Revenue Statistics, 1965–2009* (Table C); Table 2.4 in Chapter 2 above.

[5] However, in Australia employers are required to contribute to private retirement funds for their employees (see Chapter 23).

Tax rates

A **tax rate** is the ratio of tax collected during an accounting period to the tax base. Two ratios are particularly important. One is the **average tax rate** (ATR), which is the total tax paid divided by the relevant tax base:

Average tax rate
The total tax paid divided by the value of the relevant tax base

$$\text{ATR} = \text{total tax paid/value of tax base} \tag{25.2}$$

The ATR is important for equity reasons.

The second important ratio is the **marginal tax rate** (MTR), which is the extra tax paid per additional dollar value of the tax base:

Marginal tax rate
The extra tax paid per additional dollar of the tax base

$$\text{MTR} = \Delta \text{ tax paid} / \Delta \text{ value of the tax base} \tag{25.3}$$

where Δ represents the change in the relevant variable. The MTR is important when considering the incentive and efficiency effects of taxation.

The **tax rate schedule (or structure)** is the complete set of tax rates in an economy. Tax rate structures are typically characterised as proportional, progressive or regressive. Each of these terms is defined in relation to the ATR.

- In a **proportional tax system,** the ATR is constant for all values of the tax base. In this case, the MTR is also constant at all values and equals the ATR.
- In **a progressive tax system,** the ATR rises as the tax base rises. In this system, the MTR is usually higher than the ATR though exceptions can be designed.
- In a **regressive tax system,** the ATR falls as the tax base rises. In this case the MTR is below the ATR. However, if the MTR is positive, high-income individuals still pay more tax than lower-income individuals even in a regressive system.

In most countries, the tax system in aggregate is progressive in varying degrees.

Tax units

The **tax unit** is the entity on which tax is levied. Most OECD countries, including Australia, treat individuals as the basic tax unit (OECD, 2001c). A few countries, including France, Germany and Switzerland, treat the family (or at least adult married couples) as the tax unit. Typically, this involves splitting the joint income into two equal parts.

If a country has a constant MTR (a flat tax system) an individual's tax liability does not depend on the choice of tax unit. However, in most countries the MTR rises with income. In this case the choice of tax unit affects marginal and average tax liability and can have significant efficiency and equity effects.

With progressive taxation and similar tax structures for individuals and families, the ATR and the total tax is lower for a couple assumed to have equal incomes than for two separate individuals with different individual incomes but the same total income. Also, taxed as a family unit the MTR of the high-income earner falls and the MTR of the low- income partner rises. With normal substitution effects, the high-income earner works more hours and the low-income earner works fewer hours than with individual taxation. The net effect may be efficient in that, for given combined work hours, the value of output may increase.

However, because the ATR falls with families where income is split evenly, in 1969 the United States introduced a higher tax structure for families. Consequently, as Stiglitz (2000) noted, when individuals with similar incomes get married, marriage increases the total tax liability of the couple and this may discourage marriage in these circumstances.

In terms of equity, there are grounds for treating the household as the tax unit. An individual with a dependent (or several dependants) has more needs than someone without dependants. As seen in Chapter 20, measures of need are generally based on household

equivalent measures that allow for differences in household size and composition. On the other hand, Apps (2010) and other commentators strongly criticise family taxation on the grounds that this increases the MTR of the lower wage earner, discourages that earner (often female and with a relatively high labour supply elasticity) from working, and increases the deadweight loss (DWL) of taxation. It also increases the dependence of the lower income earner on their partner.

In practice, many countries adopt a dual (not entirely consistent) approach. They adopt the individual as the basic income tax unit but provide income tax deductions or credits based on family members. Furthermore, most welfare payments are based on the family unit. In effect, many aspects of tax-transfer systems are based more on the family than on the individual unit. This has happened to a considerable extent in recent years in Australia (see Chapter 23) despite the trend away from traditional and intact families based on marriage towards more informal households. Treating households as tax units raises considerable definitional and practical problems. Given these difficulties, it is unlikely that tax systems based on individual units will convert fully to taxation of family or household units.

Tax exemptions

When taxes are selective, parts of the tax base are taxed at a concession rate or exempt from any tax. For example, in Australia some forms of income (imputed rents and capital gains on owner-occupied housing and gifts and bequests) are not taxed. Some consumption goods, such as health care services and unprocessed food, do not attract a consumption tax.

In addition, many organisations providing goods or services of various kinds are not taxed. Most government-provided services are exempt from tax. Also, not-for-profit charities and religious organisations, mutual associations and private clubs are exempt from income tax, although they may be required to pay taxes on inputs. In Australia, most private schools and sporting clubs, including enterprises with multi-million dollar turnovers, are exempt from income tax. Some religious organisations run large commercial operations. For example, the tax-exempt the Seventh Day Adventist Church wholly owns the Sanitarium Health and Wellbeing Company, a business with a turnover of several hundreds of millions of dollars. Evidently religion provides earthly as well as heavenly benefits. Some of these issues are discussed further in Chapter 32.

Tax Hypothecation

A feature of most public finance is the separation of expenditure and revenue collection. When economists do cost–benefit studies, costs and benefits are linked and considered jointly. When government agencies make expenditure decisions, they need to know that the expenditure is authorised, but they generally have no regard for how this expenditure has been, or will be, financed. As Buchanan (1975) pointed out, when government expenditure and revenue decisions are made separately, the resulting choices are likely to be inefficient.

Tax hypothecation
Links the tax revenue from a tax to a specified kind of expenditure

When government links tax revenue collection and expenditure, the taxes are described as 'hypothecated'. **Hypothecation** means linking the revenues from a tax to a specified type of expenditure. Hypothecated taxes are also known as earmarked or dedicated taxes. Hypothecated taxes should be distinguished from user charges. A hypothecated tax does not give the payer any entitlement to a specific service.

Hypothecated taxes are common. Many countries have hypothecated social security charges. In the United States, the states raise about a quarter of their revenues from hypothecated taxes (Warren, 2004). The Australian government has raised several small hypothecated taxes, including an aircraft noise levy, a gun buy-back levy, a milk levy, a sugar levy and a flood damage levy. In each case the funds were used to compensate a group in society that was believed to have lost an entitlement or suffered exceptional damages. The

Australian government also requires taxpayers to contribute a minimum of 9.5 per cent of wages to a pension fund, which is in effect a hypothecated levy. Most Australian states have hypothecated taxes to fund roads, urban fire brigades, rural bush fire services, and insurance protection.

Hypothecated taxes facilitate raising revenue if the revenue is perceived to be dedicated to a good cause. Further, if there is a tight link between expenditure and the revenue to fund it, it is easier to satisfy the classical (Wicksellian) test for the provision of public goods that the collective willingness to pay for a project or program should exceed or at least equal the cost.

However, there is often only a loose relationship between public expenditure and a hypothecated tax and the idea that hypothecated taxes introduce fiscal discipline is quite misleading. For example, the revenue from the Australian Medicare levy (1.5 per cent of taxable income for most taxpayers), which is nominally hypothecated to health care services, spills back into consolidated revenue and does not necessarily fund health care or affect the quantum or allocation of health expenditure. If the revenue from a hypothecated tax is less than the public expenditure that would occur in any case, the tax is in effect simply part of consolidated revenue. On the other hand, if the revenue from a hypothecated tax exceeds desired public expenditure on the linked good, resource allocation is distorted. Treasuries almost everywhere argue that all revenues should accrue to consolidated revenue where expenditure priorities can be established regardless of the source of the funds.

Tax Avoidance and Evasion

Tax avoidance
Any change in behaviour that legally reduces tax liability

Tax evasion
Not paying tax that is legally due

Individuals and firms commonly seek to minimise tax payments. Some do so legally and others illegally. In the economics literature tax avoidance is legal; tax evasion is not. Viewed broadly, **tax avoidance** is any change in behaviour that legally reduces tax liability. For example, this may mean making untaxed house renovations instead of doing taxed commercial work. Sometimes, tax avoidance is defined more narrowly in an accounting sense as 'arranging one's business affairs so that tax is not legally payable' (Black, 2002). **Tax evasion** means not paying taxes that are legally due. It usually involves under-declaration of taxable income or exaggeration of expenses, or not even filing a tax return.

There are many forms of tax avoidance. In the broad use of the term, producing household goods rather than market goods is tax avoidance. In the narrower accounting use of the term, tax avoidance includes shifting income from a high-taxed category to a lower-taxed category and postponing taxes. Examples of income shifting in Australia include salary sacrifice (shifting earned income that would be taxed at a high marginal rate into a contribution to a retirement fund that is taxed at a lower rate), conducting business through a trust entity which can distribute profits to low-income beneficiaries rather than through a corporate entity, and shifting assets from a high-taxed family member to a low-taxed member. Tax can be postponed by not realising capital gains, because unrealised capital gains are rarely taxed. When there are significant differences between marginal tax rates for different individuals or kinds of income, the gains from tax avoidance can be very large.

Tax avoidance has three major efficiency costs. One is the distortion in economic behaviour. To reduce tax many individuals and firms adopt business practices and savings structures that have little or no productive value and that they would not adopt in the absence of the tax. The second major cost is the huge amount of highly skilled professional resources that are devoted to creating and maintaining tax-effective business and savings structures (see some estimates below). These uses of resources have little, if any, economic benefit to the community. Third, tax avoidance means that tax rates must be raised to collect given tax revenue with associated DWL. Tax avoidance also has significant implications for social equity.

Tax evasion is also common. Most often this involves under-reporting of income earned in the local black economy. It may also involve under-reporting of income earned from assets held in international tax-haven jurisdictions. The black economy is the sum of unreported commercial transactions that do not enter GDP. It includes legal and illegal activities, such as the sale of illegal drugs. Bajada (2001) estimated that unreported income averages 15 per cent of GDP in Australia. Overstatement of expenses, notably through paying excessive prices to related firms overseas in lower-tax jurisdictions, is another form of tax evasion. The US Internal Revenue Service estimates that taxpayers voluntarily pay only about 80 per cent of their actual income tax liability (Rosen and Gayer, 2014).

Like tax avoidance, tax evasion distorts the use of economic resources as untaxed activities displace more productive taxed activities. Tax evasion is also inconsistent with the principles of both vertical and horizontal equity (see below). For a detailed analysis of the positive and normative effects of tax avoidance and evasion, see Slemrod and Yitzhaki (2002).

Although tax avoidance and evasion are conceptually distinct, it may be hard to distinguish between them. Firms and individuals often employ expensive tax accountants to advise whether a given practice is tax avoidance or evasion, and even then the tax office may disagree. Court decisions in many countries have upheld the right of citizens to arrange their affairs to keep their tax payments as low as possible. However, Part IVA of the Australian *Income Tax Assessment Act* states that if tax reduction is the dominant motive of a business arrangement this will be construed as illegal and disallowed. Since many common and legally accepted business practices are designed wholly or very largely to minimise tax, this legislation is hard to interpret, but it provides some tax accountants with large mansions.

Evaluating Tax Systems

Revenue is collected for two main purposes: to pay for public goods and services and to redistribute income. Taxes may also be raised to correct externalities or for macroeconomic stabilisation objectives. However, we are not concerned with these latter aims here. If there were no redistributive objective, we could employ head taxes and there would be no deadweight loss. But because of the importance of redistribution, taxes vary across individuals and depend on their business activity. This means in turn that the taxes may distort economic activity. Therefore, the standard public finance question is 'how should taxes be raised to meet government expenditure requirements and distributional objectives in the most efficient way?'

Traditionally public finance economists set criteria that a tax system should satisfy. The more fully the criteria are satisfied the better the tax system. Thus economists have generally considered that a good taxation system should meet four criteria. It should be fair, efficient, administratively simple (cost-effective) and politically responsible.[6] Of course the tax system must also ensure revenue adequacy.

As Kaplow (2008) observed, there is no common denominator between these criteria or any principle of aggregation. In more formal analysis, the best tax system is the one that produces the highest level of social welfare as represented by a social welfare function. The two approaches are not inconsistent. The more formal analysis builds on the traditional analysis and incorporates the two key criteria (equity and efficiency) into an integrated social welfare criterion.

In this section we discuss the four criteria for a good tax system. Applications of these criteria and the development of more formal welfare evaluation criteria are discussed in the following chapters.

---

[6] These four characteristics are similar to those suggested by Smith (1776) and Mill (1948). Keynesian macroeconomists might add that a tax system should be flexible so as to enable short-run macroeconomic management.

Taxation equity

Taxation equity is generally articulated in terms of the **ability-to-pay principle**. In the words of Adam Smith, 'The subjects of every state ought to contribute to the support of government as nearly as possible in proportion to their respective abilities'.

Horizontal equity
Individuals in similar circumstances should be treated equally

Vertical equity
Individuals in dissimilar circumstances should be treated fairly

In current language, the ability-to-pay principle is usually expressed in terms of horizontal and vertical equity. **Horizontal equity** requires that individuals in similar circumstances should be treated equally. This may be interpreted as saying that individuals with a similar productive capacity should pay the same amount of tax. **Vertical equity** is fairness in the treatment of individuals in different circumstances. Thus, vertical equity is generally interpreted as requiring individuals with greater economic capacity to pay more tax than individuals with lower capacity.

Although these equity principles appear clear, they require normative interpretation. How should capacity to pay tax be measured? When are individuals in similar circumstances? What circumstances matter? Individual abilities are not directly observable. As a practical matter, ability to pay must be based on an observable unit of measure. A common measure is an individual's money income. However, as an indicator of capacity to pay this has several weaknesses. Suppose that two individuals, *A* and *B*, each earn $40 000 per annum, but that *A* owns her house, while *B* rents her house. *A* would have a higher capacity to pay tax than *B* has. Or suppose that *A* is single and healthy while *B* has two dependent children and significant health care expenditures. Their nominal income would be a poor measure of their real income or welfare (and therefore of their capacity to pay) because they have different expenditure needs. Nor does income necessarily reflect an individual's economic capacity. Individuals *A* and *B* may have equal ability to earn an income of $40 000, but *A* may go surfing every morning and earn only $20 000. Evidently her income would not reflect her economic capacity. These examples illustrate the difficulty of determining economic capacity and of identifying individuals in similar positions.

Similar problems arise with vertical equity. This principle implies that the ATR should rise with economic capacity. But this again requires a definition of economic capacity. Also, views may differ as to how much extra tax payment represents vertical equity. A possible principle would be that each individual should make an equal sacrifice for a publicly financed good. Suppose that *A* has an income of $80 000 and *B* an income of $40 000 and that *A* receives half the marginal utility from a dollar that *B* does. Assuming fixed labour supply, equality of sacrifice requires that *A* pay twice the marginal tax that *B* does. A tax of $100 on *A* would impose the same burden on her as a tax of $50 does on *B*. However, this is only one possible ethical approach. Some people may consider that *A* should pay more than double what *B* pays because she has a higher absolute utility than *B* has. Indeed, with a constant labour supply, the total utility sacrifice associated with taxation is minimised if tax is levied only on higher-income individuals. Views on how tax should vary with economic capacity or on how vertical equity should be achieved are ultimately normative issues that cannot be determined solely by economic science.[7]

The **benefit principle** is another equity principle. This principle states that individuals should pay for services in proportion to the benefits received. When a public agency provides services to a specific group of citizens, say property owners or a sporting club, it would seem fair that, unless members of this group were disadvantaged in some way, the beneficiaries should pay for those services and not be subsidised by other taxpayers. In such cases the benefit principle would override the ability-to-pay principle. However, the benefit principle cannot be applied to paying for pure public goods where individual benefits differ and are

[7] However, economic analysis can show the deadweight losses associated with changes in behaviour, for example changes in labour supply.

difficult to determine. Nor does the principle apply to the range of income transfers where the main objective is social welfare and ability to pay is the prime equity principle.

A related and important equity issue in public finance is **intergenerational equity**. When government invests in social or economic infrastructure it is providing services to future generations as well as to present taxpayers. Consistent with the benefit principle, it would be fair to borrow to finance part of these expenditures and to levy taxes on future generations to repay the loan. On the other hand, if pension and health schemes are unfunded pay-as-you-go schemes the current generation is effectively expecting that the next generation will fund their retirement and their health outlays.

Evidently, there are several equity principles for allocating the burden of taxation. Moreover, as we saw in Chapter 7, the **principle of fair deserts** suggests that people should be entitled to fair compensation for working to earn income. Much of the public policy debate is couched in terms of equity–efficiency trade-offs. But equity is not a well-defined single concept. Often trade-offs between various concepts of equity also need to be resolved.

Taxation efficiency

Ideally, taxes would not distort behaviour in competitive markets with no market failures. They would not divert labour or capital to less productive activities or consumption to less preferred goods. However, without acceptable lump sum taxes, some distortions are inevitable. As virtually all taxes distort economic activity, the aim of an efficient tax system is to raise taxes with the least misallocation of resources and hence with minimum DWL.

In perfectly competitive markets, factor payments and product price signals ensure production, consumption and product mix efficiency. In factor markets, wages equal the value of marginal product. In product markets, prices equal the marginal costs of production. Even when markets are competitive but not perfectly so, markets tend towards these equalities. On the other hand, most taxes create a wedge of some kind: between wages received and the market value of the marginal product or between product prices and marginal production costs or between the return on capital and the return on saving. These wedges change relative prices and alter the quantity and mix of goods produced and consumed.

Examples of distortions due to tax abound. Taxes on income reduce the return to work relative to the price of leisure. This is likely to reduce labour supply. Income taxes influence decisions on entry and exit from the labour force, the acceptance of welfare benefits instead of employment income and the kind of work undertaken. They also affect savings decisions because taxation of income from capital involves double taxation of income saved for future consumption relative to income consumed today. On the other hand, tax systems often discriminate in favour of some forms of savings such as owner-occupied housing and against others such as interest-bearing accounts, and thus affect investment decisions. Taxation of consumption often taxes some goods more than others and so affects the mix of goods consumed. Corporate tax structures may affect how businesses are organised (by incorporated or unincorporated forms) and how firms are financed (by debt or equity finance).

All such changes in behaviour represent changes from voluntary trading decisions. Aggregate welfare falls when individuals select less preferred production or consumption options. Given competitive markets, such distortions create a deadweight loss. A DWL is the excess of the total cost of a tax over the actual revenue paid.[8] Methods of quantifying DWL are described in Chapter 27.

Of course, not all markets are competitive and there are many market failures. Thus, taxation can be used to correct externalities (Chapter 13). Corrective taxes can raise revenue and improve the allocation of resources. For example, fuel taxes are corrective taxes that raise a large amount of revenue and reduce market-induced air pollution and global warming.

[8] In the US literature, a deadweight loss is often described as an "excess burden" where the tax is the basic burden.

Administration and compliance costs

Taxation involves the public costs of tax administration and the private costs of compliance with the tax system. The criterion of administrative simplicity requires that these costs be minimised.[9]

The costs of tax administration are the costs of designing, operating and changing the tax system. This is not simply a matter of collecting taxes. The costs of administering the tax system include the employment of many highly qualified people to deal with numerous disputed tax issues. Mikesell (1998) reports that expenditure by the US Internal Revenue Service equals 0.56 per cent of revenues collected. The administrative cost of the value-added tax in 12 OECD countries ranges from 0.32 to 1.09 per cent of collections. In Australia the estimated administration costs of the tax collecting authorities are about 1 per cent of revenue collected or 0.3 per cent of GDP (Freebairn, 2005).

Tax compliance costs are the costs of firms and individuals in dealing with the tax system. The costs include the costs of tax agents and other professional advisers as well as the time of internal staff and taxpayers taken in collecting and maintaining information for professional advisers and governments. These costs rise with the complexity of the tax rules, the number of tax rates, the variety of allowances or exemptions, the complexity of definitions, the supporting evidence required and so on. However, they are generally several times higher than public administrative costs. Sandford (1995) estimates that tax compliance costs equal about 2 per cent of GDP in New Zealand, the United Kingdom and the United States. In Australia, in 2008–09, 71 per cent of individual taxpayers employed a tax agent to assist with their tax return. Pope (1997) estimated that the cost of compliance with Commonwealth taxes was about 12 per cent of revenue collected. Including state taxes, Freebairn (2005) estimated that tax compliance costs are as high as 4 per cent of GDP in Australia. The introduction of the GST and the business activity statement in July 2000 increased compliance costs significantly for small and medium sized businesses.

Overall, expenditure on tax administration and compliance represents forgone consumption goods. For any given revenue target, these costs should be minimised.

Political responsibility

The criterion of political responsibility is normally interpreted as a requirement for transparent and certain taxation, rather than arbitrary and capricious taxes. This criterion is principally a statement of 'fair process'. Taxpayers should be able to plan their lives and businesses with certainty. However, the criterion also has efficiency implications. Countries with certain and transparent tax systems generally enjoy greater economic growth than countries with arbitrary and corrupt tax systems.

Conclusion: choosing a tax system

Choosing a tax system raises issues similar to other public policy choices. Equity and efficiency are critical issues, but administrative simplicity and political responsibility are also important.

The analysis is simplified if efficiency and equity can be considered separately. In some cases, one taxation instrument may achieve a given revenue target with less deadweight loss than another taxation instrument. In our analysis of tax efficiency in Chapter 27, we will look for Pareto improvements in the tax system. These are changes to the tax system that improve

[9] The criterion of administrative simplicity can be viewed as an efficiency criterion. Adam Smith regarded administration costs as part of the excess burden of taxation that should be minimised. However, in modern texts administrative simplicity is usually distinguished from the efficiency criterion.

the welfare of one taxpayer without reducing the welfare of another one. In so far as these improvements can be found, efficiency can be separated from equity.

More commonly a change in the tax structure benefits one individual at a cost to another. The equity effects must be traded against the efficiency effects. If we have a formal social welfare function, which incorporates interpersonal impacts, a socially preferred tax structure can be found. The application of such techniques to derive an optimal tax structure is discussed in Chapter 28. This is a useful exercise in analysing tax choices.

However, there are two main problems with these social welfare functions. First, they require interpersonal comparisons of welfare and thus ethical judgements. They cannot be derived from objective or technical analysis. Second, they bear little relationship to the day-to-day thinking of policy makers. As a practical matter, economists have a potentially important role in calculating the equity and efficiency consequences of tax systems, in describing efficiency–equity trade-offs and in explaining them to policy makers. These explanations often influence policy making. But tax is a highly sensitive issue and numerous interest groups compete for the attention of politicians.

Summary

- Government can obtain revenue by appropriating resources, taxation, social security contributions, printing money, borrowing, user charges or sale of assets. Most methods involve some coercion of individuals or firms. Taxation is by far the major source of revenue.
- There are three basic concepts in taxation. The tax base is the economic activity or item on which a tax is levied. The tax rate is the rate at which tax is levied on the tax base. The tax unit is the unit on which tax is levied.
- The two main tax bases are income and expenditure (or consumption). In OECD countries, income taxes account for just over 60 per cent of tax revenues. Consumption taxes account for about one-third of revenues. Property taxes (a tax on wealth) account for most of the balance.
- The tax rate structure is the complete set of tax rates in an economy. Tax rate structures are described as progressive, proportional or regressive. The average tax rate rises, is constant or falls with these three structures respectively. A progressive structure is the most common form.
- Many people try to escape tax. Tax avoidance is the legal rearrangement of economic activities to minimise tax liabilities. Tax evasion is failing to pay legally due taxes.
- A good taxation system should have four main characteristics. It should be equitable, efficient, administratively simple and politically responsible. In more formal analysis, the best tax system is the system that maximises social welfare based on a formal social welfare function.
- There are many concepts of equity, including ability to pay, horizontal and vertical equity, the benefit principle of fairness and the just deserts principle. Some concepts are consistent with others. But some imply different tax policies.
- Most taxes distort voluntary exchanges and have a deadweight loss (or excess burden). An efficient tax system is one that creates the minimum deadweight loss. However, some taxes can raise revenue and correct market inefficiencies.
- In developing tax policies economists have an important role in calculating the equity and efficiency consequences of tax systems, in describing efficiency–equity trade-offs and in explaining them to policy makers.

Questions

1. Given that taxation is painful and almost always imposes economic costs, why does not the government simply print money to finance its activities?
2. Under what circumstances will the choice of the tax unit affect the amount of tax paid? Given a progressive tax structure, what efficiency and equity considerations arise in assessing whether the tax unit ought to be the individual or the family?
3. Australian tax law says that business arrangements made with the dominant purpose of reducing tax are a form of tax avoidance that is illegal and will be disallowed. What kind of difficulties may arise with this kind of legislation?
4. Politicians and voters often view tax hypothecation as desirable. On the other hand economists tend to view tax hypothecation as either superfluous or as inefficient. Explain why economists think like this. Do you agree with the economists or the politicians?
5. Consider a tax system where individuals face a constant marginal tax rate of 25 per cent for every dollar they earn that exceeds $6000. Is this tax system an example of a proportional tax structure?
6. Consider an individual-based tax system which taxes income up to $30 000 at 15 per cent, and income that exceeds this amount at 30 per cent. Amy earns a salary of $45 000 and her husband Ben earns a salary of $60 000. Calculate their individual tax liability. If the tax unit were the family, calculate the joint tax liability of Amy and Ben. Which system would minimise their tax liability? Assuming that the tax system imposed was the family unit, what equity and efficiency implications arise?
7. What are attributes of a fair tax system? What are some alternative views of equity? Do the different notions of equity involve some contradiction? If so, how is this dealt with?
8. What does it mean to say that a tax system is 'efficient'? When might the tax system fail to achieve its goal of efficiency? Does taxation always drive resources away from their best use?
9. What determines the amount of tax evasion? Does tax evasion create inefficient outcomes?
10. How much resources should be devoted to tax administration?

Further Reading

Buchanan, J.M. (1987) 'The bridge between tax and expenditure in the fiscal decision process', paper reprinted in *Readings in Public Finance*, pp. 289–297, S. Baker and C. Elliot (eds), South-Western College Publishing, Cincinnati, Ohio.

Evans, C. and B. Tran-Nam (2010) 'Controlling tax complexity: rhetoric or reality?', pp. 439–463 in *Australia's Future tax System: The Prospects after Henry*, C. Evans, R. Krever and P. Mellor (eds), Thomson Reuters, Sydney.

Freebairn, J. (2005) *Equity in Sourcing Revenue: A Comparison of Alternative Tax Bases*, MIAESR Working Paper 11/205, Melbourne Institute of Applied Economics and Social Research, Melbourne.

Musgrave, R.A. (1985) 'A brief history of fiscal doctrine', pp. 1–60 in *Handbook of Public Economics*, Vol. 1, A.J. Auerbach and M. Feldstein (eds), Elsevier, North Holland, Amsterdam.

Gruber, J. (2014) *Public Finance and Public Policy*, 5th ed., Chapter 18, Worth Publishers, New York.

Noord, P. and Heady, C. (2001) *Surveillance of Tax Policies: A Synthesis of Findings in Economic Surveys*, Working Paper 29, Economics Department, OECD, Paris.

Organisation for Economic Cooperation and Development (2010) *Revenue Statistics 1965-2009*, OECD, Paris.

Rosen, H. and T. Gayer (2014) Public Finance, 10th ed., Chapter 14, McGraw-Hill, New York.

Slemrod, J. and Yitzhaki, S. (2002) 'Tax avoidance, evasion and administration,' pp. 1423–1470, in *Handbook of Public Economics*, Vol. 3, A.J. Auerbach and M. Feldstein (eds), Elsevier, North Holland, Amsterdam.

Stiglitz, J.E. (2000) *Economics of the Public Sector*, Chapter 17, W.W. Norton, New York.

Chapter

26

Incidence of Taxation

In general, the art of government consists in taking as much money as possible from one class of citizens to give to the other.

Voltaire

Incidence of Commodity Taxes ♦ Incidence of Taxes on Factors of Production ♦ General Equilibrium Analysis ♦ Further Issues in Tax Incidence ♦ Tax Incidence in Australia

Economic incidence of taxation
The change in real income as a result of taxation

Statutory incidence of taxation
The party responsible for paying a tax

A fundamental finding of the study of tax incidence is that the real burden of tax is often different from the statutory incidence. The real burden (or **economic incidence**) of a tax is the change in real income of any economic agent that results. The analysis of economic incidence shows how taxes change the prices of commodities and factors of production and how these changes in prices affect in turn the distribution of income. As we will see, the price changes depend on the conditions of demand and supply and the structure of markets.

When a tax is imposed the relevant legislation states the **statutory incidence**—who will be responsible for paying the tax. However, many taxes can be shifted. If government levies an excise tax on wine, retailers may pass on part, or all, of the tax in higher prices to wine drinkers. Shifting occurs especially with indirect taxes on commodities. When taxes are imposed on producers but borne in higher prices by consumers, taxes are shifted forwards. When taxes are imposed on consumers, but producers receive lower prices as a result, taxes are shifted backwards. Taxes on factors of production may also be shifted. Employers may pay a payroll tax on payments to labour, but then pay labour lower wages. A tax on the return to owners of capital may reduce the supply of capital and increase the cost of capital to users of capital.

In the first two sections below, we discuss the incidence of taxes on commodities and on factors of production. In both cases we examine the effects in the market where the tax is levied and in related markets. This analysis, known as partial equilibrium analysis, brings out the major effects of taxes. However, sometimes a tax affects prices in multiple markets. For example, a tax on a major commodity can affect incomes in factor markets and this in turn may affect the demands for other commodities. To analyse such general effects, we employ a general equilibrium (multi-market) model. This is the subject of the third section. The last part of the chapter discusses various other topics including equivalent taxes, the incidence of a general value-added tax and tax incidence in Australia.

Incidence of Commodity Taxes

We start by examining the incidence of a unit tax on a commodity (any good or service) produced in a competitive market. A **unit tax** is a given amount of tax for each unit purchased. We then examine the incidence of an *ad valorem* tax. An ***ad valorem*** tax is a tax on the value of commodities sold.

Unit tax
A given amount of tax per unit purchased

Ad valorem tax
A tax on the value of sales

Incidence of commodity taxes in a competitive market

Figure 26.1a shows the effect of a unit tax in a competitive market for wine. The market demand curve is shown as $D(P)$. The market supply curve is represented pre-tax by the $S(P)$ schedule. In a competitive market the supply curve is also a marginal cost curve. The pre-tax equilibrium is Q_1 litres of wine sold at price P_1. After a unit tax (T) is levied on wine producers, the supply curve shifts upwards to $S(P) + T$. For firms to produce the same quantity of wine with the tax as without it, the price of wine would have to rise by exactly the amount of the tax. At the new equilibrium, the gross price paid by consumers rises to $P_2 + T$, but the net price that producers receive after tax falls to P_2. The difference between the gross and net price is the tax. The quantity supplied (and consumed) falls from Q_1 to Q_2.

In Figure 26.1a, the direct burden of the tax (tax payments) equals area $(P_2 + T)ADP_2$. Although wine producers formally pay the tax, consumers bear the costs associated with the rectangle $(P_2 + T)ABP_1$ and producers pay the balance of P_1BDP_2. However, the tax revenues are not the full burden of the tax. With output falling from Q_1 to Q_2, there are additional losses of consumer and producer surplus given by areas ABC and BCD respectively. These losses are the deadweight loss (or excess burden) of the tax.

Now consider the outcome if the same unit tax were levied on consumers. As shown in Figure 26.1b, the market demand curve for wine would fall by exactly the amount of the tax to the new demand schedule $D(P) - T$. If consumers are willing to buy a million bottles of wine at \$12 a bottle before a \$1 tax is levied, they would be willing to pay \$11 a bottle for the same quantity of wine after the tax is introduced. Because the wedge between the gross price paid by the consumer and the after-tax price received by the producer is the same regardless of who pays the tax to the tax office, the post-tax quantity and price equilibrium is the same as in Figure 26.1a. The algebra is shown in Box 26.1 overleaf. This is an important result: the distribution of the tax burden is independent of the statutory incidence.

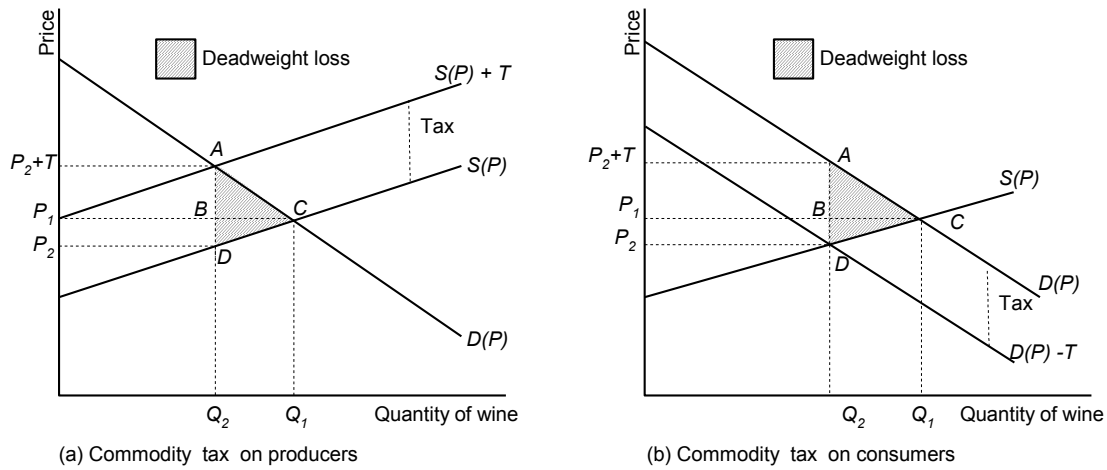


Figure 26.1 Commodity tax, prices and quantities in a competitive market

Box 26.1 Equivalence of a tax on producers and consumers

To show the equivalence of a tax on producers and consumers we first establish the pre-tax equilibrium price (P), based on linear demand and supply curves (Q_d and Q_s). We then examine how a unit tax (T) levied on consumers and producers separately affects the producer price (P^s) and the consumer price (P^c).

Suppose initially that

$$Q_d = a - bP^c \text{ and } Q_s = -c + dP^s \quad (26.1)$$

In equilibrium $$a - bP^c = -c + dP^s \quad (26.2)$$

and $$P^c = P^s = (a + c)/(b + d) \quad (26.3)$$

A tax on consumers reduces the price that they are willing to pay to producers.

So $$Q_d = a - b(P^s + T) \quad (26.4)$$

In equilibrium $$a - bP^s - bT = -c + dP^s \quad (26.5)$$

The price to producers is

$$P^s = (a + c - bT)/(b + d) \quad (26.6)$$

The price to consumers is

$$P^c = P^s + T \quad (26.7)$$

A tax on producers increases the price at which they are willing to supply consumers.

So $$Q_s = -c + d(P^c - T) \quad (26.8)$$

In equilibrium $$a - bP^c = -c = dP^c - dT \quad (26.9)$$

The price to consumers is

$$P^c = (a + c + dT)/(b + d) \quad (26.10)$$

The price to producers is

$$P^s = P^c - T \quad (26.11)$$

With quite simple manipulation it can be shown that the right-hand side (RHS) in Equation 26.6 = RHS in Equation 26.11 and that RHS in Equation 26.7 = RHS in Equation 26.10.

Thus, the gross price paid by consumers and the net price received by producers are independent of the party on which the tax is levied.

The key determinants of incidence of a commodity tax are the relative demand and supply elasticities for the commodity. The burden is borne by consumers or producers with the relatively more inelastic demand or supply respectively.

Two polar cases are shown in Figure 26.2. In panel (a), the tax is levied on producers, but demand is perfectly inelastic. In this case producers can pass all the tax on to consumers, with the consumer price rising from P^c_1 to P^c_2. In panel (b) the tax is levied on consumers, but the supply of the commodity is perfectly inelastic. In this scenario producers bear the whole cost, with the producer price falling from P^s_1 to P^s_2.

Now consider what would happen if demand or supply is perfectly elastic. If demand is perfectly elastic, firms cannot raise prices in response to a commodity tax. They have to reduce

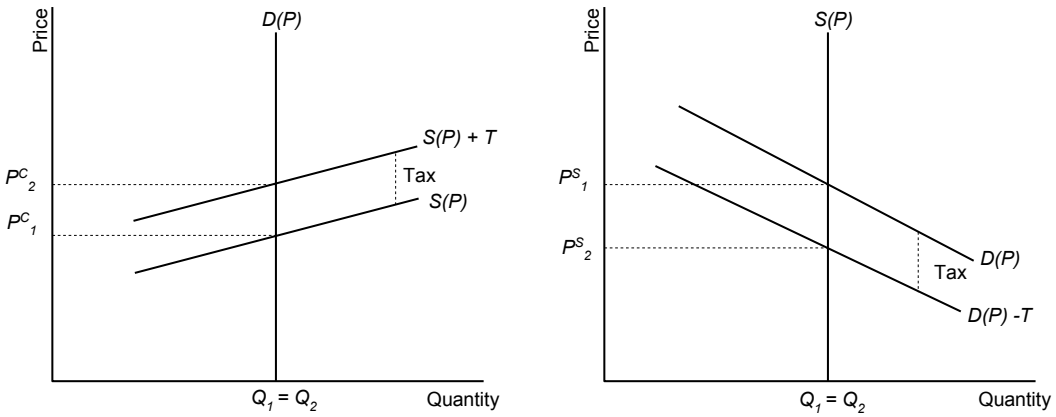


Figure 26.2 Incidence of commodity tax with inelastic demand and supply

Box 26.2 Incidence of a commodity tax

To estimate the incidence of a commodity tax we need to know how the price changes that result from the tax. If a commodity tax (T) is levied on producers, the change in price paid by consumers (ΔP^c) = $\Delta P^s + T$, where ΔP^s is the change in the price to producers, which is likely to be negative.

However, the percentage change in quantity must be the same for consumers and producers. Drawing on the standard elasticity formula, $\eta = \Delta P/P \times Q/\Delta Q$, the change in quantity formulae for consumers and producers are respectively:

$$\Delta Q/Q = \eta_d \times (\Delta P^s + T)/P \text{ and } \Delta Q/Q = \eta_s \times \Delta P^s/P$$

where d and s represent demand and supply respectively.

Note that η_d is negative and η_s is positive. Setting these two changes in quantity expressions equal to each other, the change in price to producers is given by:

$$\Delta P^s = [\eta_d/(\eta_s - \eta_d)] \times T$$

Equivalently in proportional terms:

$$\Delta P^s/T = [\eta_d/(\eta_s + \eta_d)] \times T$$

It may also be inferred from Equations 21.12 and 21.13 in the text that the relative impacts of the tax on the prices paid by consumers and producers are an inverse function of the respective elasticities:

$$\Delta P^c/\Delta P^s = \eta_s/\eta_d$$

output until their marginal cost plus the tax equals the consumer price. On the other hand, if supply is perfectly elastic, the consumer price rises by the full extent of the tax, and the quantity consumed falls, but the after-tax price received by producers is unchanged.

As shown in Box 26.2, the proportion of a commodity tax borne by producers is:

$$\Delta P^s/T = \eta_d/(\eta_s + \eta_d) \tag{26.12}$$

where ΔP^s is the change in the producer price, T is the unit tax, η is price elasticity and the subscripts d and s refer to demand and supply respectively. The burden borne by producers rises with the price elasticity of demand and falls with the price elasticity of supply. Similarly, the proportion of the tax borne by consumers is given by:

$$\Delta P^c/T = \eta_s/(\eta_s + \eta_d) \tag{26.13}$$

where ΔP^c is the change in the consumer price. The consumer burden falls with the price elasticity of demand and rises with the price elasticity of supply.

In practice, consumers usually bear more of the incidence of specific commodity taxes than do producers because the demand elasticity is less than the supply elasticity. Estimated demand price elasticities for commodities such as alcohol, tobacco and petrol, which often attract excise taxes and account for large amounts of public revenue, are usually low. For example, many studies have found that the price elasticity of demand for petrol is about –0.3 (Goodwin, 1992) and for tobacco about –0.4 (UK Department of Health, 1994). Factors of production are usually mobile at least between the supply of different goods and supply elasticities are high especially in the medium or long run. Accordingly, it is generally concluded that taxes on commodities are borne mainly by consumers.

Impacts on related commodities. So far, we have focused on the market on which the tax is imposed. When commodities are close substitutes the effects can flow over to a substitute market. Figure 26.3 overleaf shows how a tax on beer can affect suppliers and consumers of wine. In panel (a), a tax on beer shifts the supply curve to the left. The consumer price of beer rises from P^c_1 to P^c_2 and the quantity consumed falls from Q_1 to Q_2. As a result, in panel (b), the demand for wine, a substitute product, increases from D_1 to D_2 and both the price of wine and the quantity consumed increase. Suppliers of wine gain increased profits shown by area P_2ABP_1. Existing consumers of wine pay higher wine prices ($P_2 - P_1$) and indeed some may reduce their wine consumption. Consumers who switch from beer to wine also experience a loss of consumer surplus as beer was their preferred refreshment at pre-tax prices. Note that there may be a further market adjustment as the demand for beer may now rise following the

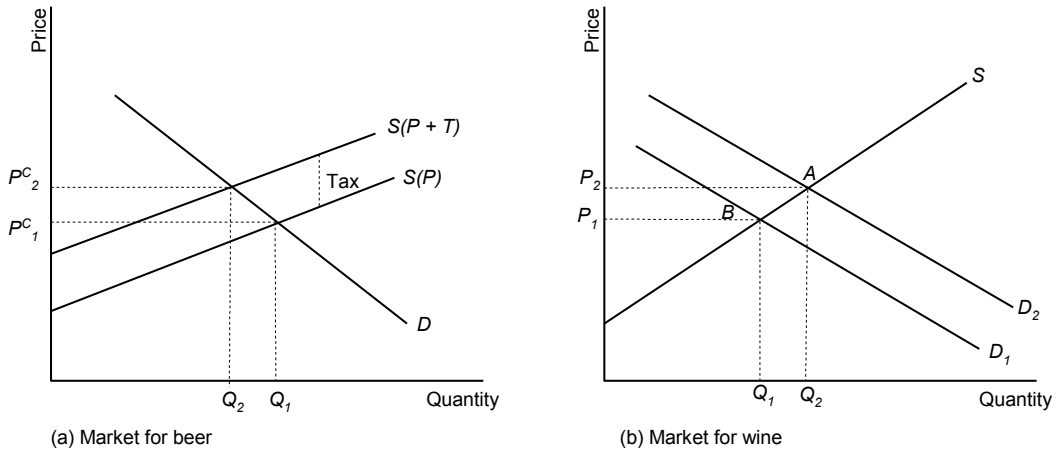


Figure 26.3 Effect of a tax on beer on markets for beer and wine

increase in the price of wine, and indeed further adjustments may continue until new market equilibria are achieved.

Incidence of *ad valorem* taxes. With an *ad valorem* tax, levied as a proportion of the price of the good being taxed, the tax rises as the price increases. This is illustrated in Figure 21.4, which shows three demand curves. Curve D_1D_1 is the initial demand with no tax. Curve D_2D_1 is demand with an *ad valorem* tax, which tilts the demand curve as price increases. Demand with a unit tax is shown by the D_3D_3 curve, which is parallel to the D_1D_1 curve. When the tax payment is the same with an *ad valorem* tax as with a unit tax, the effects of the two taxes are the same. As shown in Figure 26.4, the equilibrium price received by producers (P_2) and the quantity supplied (Q_2) are the same with both forms of tax.

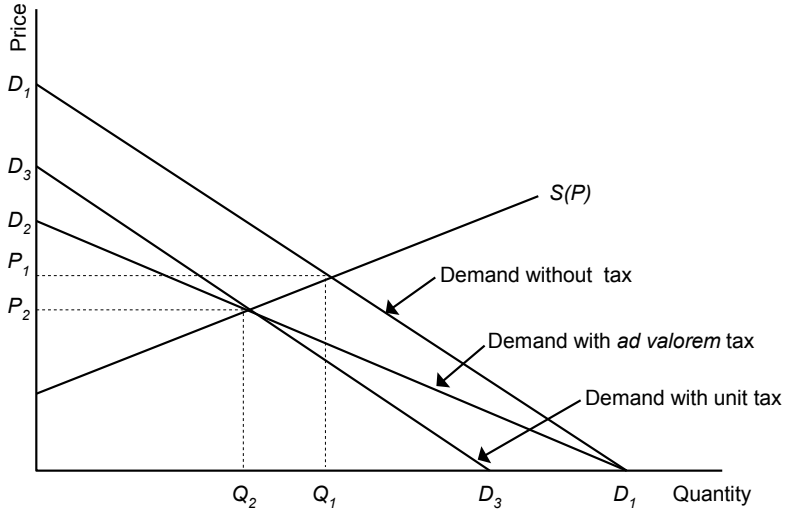


Figure 26.4 *Ad valorem* and unit commodity taxes

Of course, when the *ad valorem* tax is higher or lower in dollar terms than the unit tax, the outcome varies according to the amount of tax levied. Again, the distribution of the burden between consumers and producers depends on the relative price elasticities of demand and supply.

Traditionally governments levied unit taxes more often than *ad valorem* taxes. Unit taxes require data on quantities of items sold rather than on prices (or revenue) and quantity data are usually easier to obtain and monitor than prices. However, unit taxes are less equitable. It would scarcely be equitable to charge the same tax on a $10 bottle of wine as on a $50 bottle. Although unit taxes can vary with product quality, for example with the alcohol content of beer and spirits, the administration would generally be complex. As the variety of market goods rises, governments are increasingly using *ad valorem* taxes.

Commodity taxes with imperfect competition

The incidence of a commodity tax is different and more complex with imperfect competition. Typically, firms in imperfect competition face a downward-sloping demand curve. In the absence of price discrimination, this implies that marginal revenue is less than price and the marginal revenue curve sits below the demand curve. Figure 26.5a illustrates the case for a linear demand curve and a constant marginal cost schedule. With no commodity tax a profit-maximising firm produces up to the point where marginal revenue equals marginal cost (MC) and will produce output Q_1 at price P_1. Now suppose that a unit tax is imposed which increases the marginal cost of output to $MC + T$. The new equilibrium output is Q_2. The equilibrium price can be shown to rise by exactly half the tax from P_1 to P_2. Even though the firm has elastic supply at $MC + T$, the firm bears a substantial part of the tax.[1]

On the other hand, with a constant price elasticity demand curve as in panel (b), marginal revenue is a constant proportion of the price. In this case a unit tax increases the price by $[1/(1 - 1/\eta_d)] \times T$. Thus, if $\eta_d = 2$, the increase in price is twice the size of the tax. Indeed because, for a profit-maximising monopoly $\eta_d > 1$, the price always increases by more than the tax. In these cases consumers bear more than 100 per cent of the tax.

What happens if output is produced at increasing marginal cost? It can be shown that a commodity tax will lead to a smaller fall in output and a smaller increase in price. Indeed,

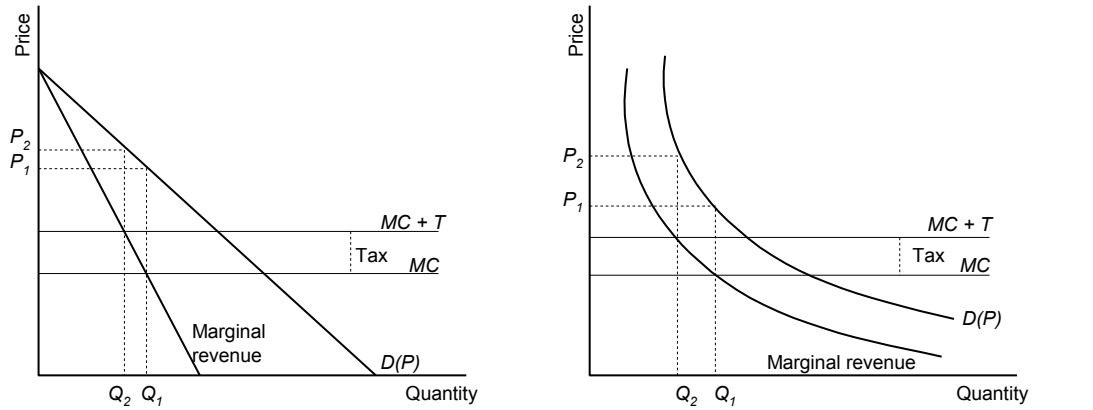


Figure 26.5 Incidence of a unit commodity tax with a monopoly

[1] For a formal derivation of the results in this subsection, see Stiglitz (2000), Chapter 18.

with perfectly inelastic supply there would be no change in output or price. As in a competitive market, the whole tax would be borne by the producer. Although this is not usual, it could occur when a plant, such as a steel mill, operates at full capacity.

Consider now an *ad valorem* tax. In a competitive market, for a given tax revenue, an *ad valorem* tax has the same price and output effects as a unit tax. However, in a monopoly market an *ad valorem* tax produces a smaller rise in price and smaller fall in output than does a unit tax. The reason is that an *ad valorem* tax is a tax on marginal revenue (MR) and, in a monopoly, MR < *P*. A unit tax reduces MR by the exact amount of the tax. An *ad valorem* tax reduces MR by less than the tax. Suppose that a product sells for \$100, produces MR of \$50 and is subject to a \$10 unit tax. The equivalent *ad valorem* tax on the product would be 10 per cent but this would reduce MR by \$5 (not \$10). Because a profit-maximising monopolist sets MR = MC, if MR falls by less than MC, there is a smaller impact on output and price and a smaller impact on consumers than with a unit tax.

In summary, the incidence of a commodity tax in monopolistic markets depends on several factors. The tax is borne mainly by consumers when the tax is a unit tax, the demand curve is non-linear, the monopolist can discriminate between customers and marginal cost is constant. The monopolist bears a greater proportion of the tax when the tax is an *ad valorem* tax, the demand curve is linear, the firm cannot discriminate between consumers and marginal cost is rising.

Incidence in oligopoly markets. Tax incidence in oligopoly markets is less well determined. Incidence depends again on how taxes change prices. However, there is no single explanation of price determination in an oligopoly market and no unique price equilibrium. Without an equilibrium price there is no basis for predicting how taxes will change price. Prices depend on how firms compete or collude with each other. In most models of firm behaviour in oligopoly markets, except the Bertrand model, firms expand output until MR= MC and P > MR as in the monopoly model. The incidence is here similar to that in the monopoly model described above. In the Bertrand model of oligopoly behaviour, P = MC and the incidence is likely to be as in in a competitive market.[2]

Incidence of Taxes on Factors of Production

Analysis of the incidence of taxes on the earnings of factors of production (labour, land and capital) follows similar lines to that for commodity taxes. The economic incidence depends on the relative elasticity of factor demand and supply, not on who nominally pays the tax. The relatively less elastic side of the market bears a higher share of the tax.

Three other general points may be noted. First, as with commodity taxes, incidence depends on the nature of the market. Most factor markets are competitive, with many buyers and sellers. However, some markets are not fully competitive, for example some labour markets are unionised or regulated, and this will affect the tax incidence. Second, most taxes on factors of production are *ad valorem* taxes. Labour is taxed on the value of its output or on wages paid, not on hours worked. Third, taxes on factors may be general or selective. For example, some payroll taxes and some land taxes apply only to certain payrolls and types of land respectively. However, as we will see, these taxes usually also affect the untaxed payroll labour or land.

Taxation of payments to labour

Labour may be taxed either by a personal income tax or by a group payroll tax. We consider first an **income tax**. This may be levied on a pay-as-you-earn (PAYE) basis with the tax

[2] See Varian (2006), Chapter 27.

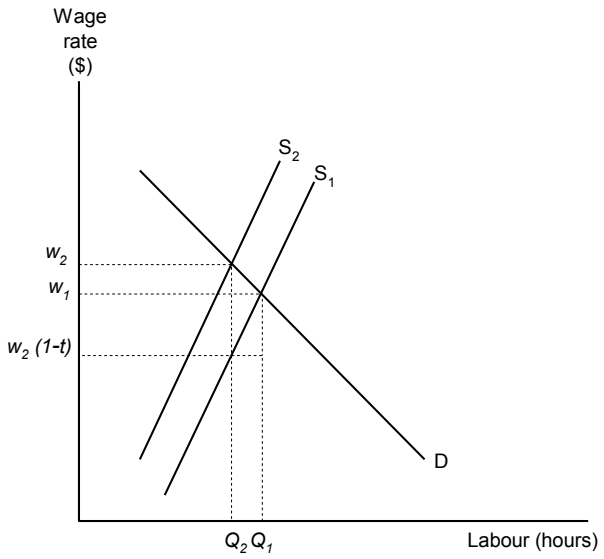


Figure 26.6 Incidence of income tax on labour

remitted by the employer or annually on the worker's declared income. The method of payment does not affect the analysis or the real outcome. Figure 26.6 depicts the demand and supply of labour as a function of the wage rate (w). The supply of labour (measured in labour hours) is shown as relatively inelastic. An income tax typically shifts the labour supply curve leftwards from S_1 to S_2 as workers offer less hours for lower take-home wages. In this case the gross wage rate shifts from w_1 to w_2, but the after-tax wage rate falls to $w_1(1 - t)$ where t is the tax rate. If labour supply is relatively inelastic, labour will bear most of the tax, but employers will also bear a small part in higher gross wages. Thus, although income tax is generally described as a direct tax on workers, some effects will be borne by employers. The quantity of labour hours will fall from Q_1 to Q_2.

We turn now to **payroll tax**. In many OECD countries, payroll tax is a general tax levied as a percentage of the total labour income of all employees. In this case the payroll tax works like personal income tax. For any given tax rate, the tax paid on the marginal worker is the same whether paid on the increment to total wage payments or as a tax on an individual income. The incidence effects are therefore as described for a personal income tax on labour.

However, in Australia the states levy a payroll tax as a selective tax on the total payroll of companies above a threshold amount. This raises fresh issues. Figure 26.7a overleaf shows the effects of the payroll tax on the taxed sector. The schedules D_1 and D_2 show the demand by large firms for labour without and with payroll tax respectively. Labour supply is more elastic than in Figure 26.6 because labour can work in the untaxed sector. Because labour supply is more elastic, work hours fall proportionately more and large firms bear more of the tax burden than with a general income tax. On the other hand, the burden on the employee is less with the take-home wage rate falling only from w_1 to w_2. But this is not the full story. Panel (b) shows the effects in the untaxed labour sector. Because employment and wages decline in the taxed sector, labour supply to the untaxed sector increases from S_1 to S_2. Employment in this sector increases to Q_2 and the wage rate falls to w_2. Indeed, if workers are indifferent between working in the two sectors, the wage falls by the same amount in the untaxed sector as in the taxed sector! On the other hand, firms in the untaxed sector gain from paying lower wages. However, if the labour supply curve is upward sloping, the fall in the wage rate from w_1 to w_2 will result in lower total employment.

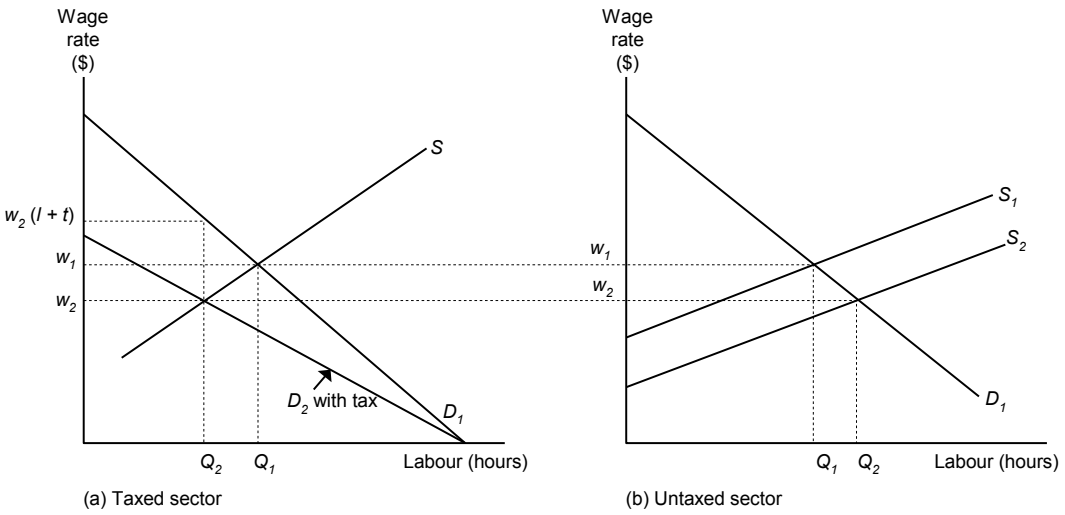


Figure 26.7 Incidence of selective payroll tax

Taxation of income from capital

Turning to taxation of income from capital, we again present two scenarios: with an inelastic and an elastic supply of capital. Figure 26.8a overleaf presents a fixed supply of capital that is independent of the rate of return on capital (interest rates). This could occur in the absence of foreign capital.[3] It also shows the demand for capital, with demand increasing as the price of capital falls. With no tax on income from capital, Q_1 capital would be supplied with a rate of return r_1. If a tax rate (t) is levied on the income, the quantity of capital supplied would be unchanged. However, suppliers of capital would receive a return of $r_1(1 - t)$. The whole burden of the tax would be borne by suppliers of capital.

Figure 26.8b presents a scenario with capital in perfectly elastic supply (S_1) due to the availability of foreign capital to a small open economy. In the absence of a tax on income from capital, Q_1 capital would be supplied with a return of r_1. In this case imposing a tax on income from capital would shift the supply curve upwards as shown to S_2. The amount of capital demanded would fall to Q_2 and the gross return on capital would rise to r_2. However, the net return to lenders would fall exactly to $r_1 = r_2(1 - t)$. Although a tax on income from capital may be intended partly as a tax on foreign capital, the tax does not change the after-tax return paid to foreign lenders, who simply lend less to the local economy. In this scenario, the whole tax is borne by local firms who pay r_2 for capital instead of r_1.

Taxation of returns to land

The total supply of land is perfectly inelastic. It follows that a general *ad valorem* tax on land has no impact on the quantity of land available. Such a tax reduces the value of land to the landowner and is borne entirely by the landowner. The tax cannot be shifted. The same applies to a tax on a natural resource such as an oil field or a diamond deposit. A general *ad*

[3] The domestic supply of capital is fixed if the substitution and income effects of interest rates on savings cancel out. As the interest rate rises, savers have more incentive to save because of the substitution effect but less need to save because of the income effect of higher returns.

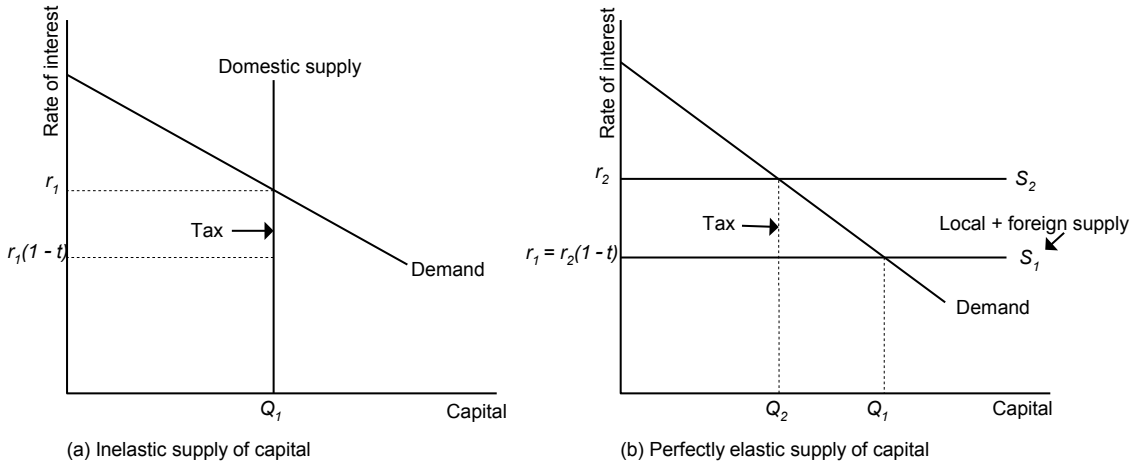


Figure 26.8 Incidence of taxation of capital

valorem tax on income directly attributable to a natural resource reduces its capital value by the discounted present value of the tax. This tax is borne wholly by the owner of the resource.

The actual incidence of taxation of land and natural resources is often more complex for two reasons. First, many taxes on land are selective; they vary with land use. When land has several uses, as it often does, the supply of land to any specific use is price elastic. Landowners switch to land uses with lower taxes. This reduces the supply of heavily taxed land. The users of this land will bear some of the tax by paying increased rents. On the other hand, the supply of land for untaxed uses will increase and so reduce the value of untaxed land. Thus, owners of untaxed land indirectly bear some of the tax on taxed land.

Second, nearly all land is improved and most exploitation of natural resources is made possible by the application of capital. Farmland is cleared, drained and made suitable for crops. Urban land is serviced with roads, water, sewerage services and so on. Thus, market prices of land and natural resources usually include the value of improvements and are payments for capital as well as for land. Nevertheless, in informed markets, in so far as the supply of capital is elastic and the supply of land is perfectly inelastic, a tax on the combined value will be borne entirely by the land component (i.e. it will reduce the value of the land). Capital must obtain the same return when combined with land as it would in other uses.

Tax capitalisation

So far, we have assumed that a tax is borne when it is levied. However, this is not necessarily true. Some taxes are borne before they are levied! This process is known as tax capitalisation. Tax capitalisation occurs when a stream of present and expected future taxes is incorporated into the present capital value of an asset. This commonly occurs with periodic land taxes. Suppose that land rent is $\$R$ per annum and that the rate of interest is r. The capital price (P) of the land equals:

$$P = R_0 + \frac{R_1}{(1+r)} + \frac{R_2}{(1+r)^2} + \ldots + \frac{R_n}{(1+r)^n} \qquad (26.14)$$

where land provides rent for n years. Now suppose that the rent is taxed at rate t each year, the price of the land becomes:

$$P^t = R_0(1-t) + \frac{R_1(1-t)}{(1+r)} + \frac{R_2(1-t)}{(1+r)^2} + \ldots + \frac{R_n(1-t)}{(1+r)^n} \qquad (26.15)$$

The difference between the asset price of land in Equations 26.15 and 26.14 is the discounted value of the future tax payments. The asset price falls by the present value of all future tax payments. In fact, it falls by $(1-t)$.

Tax capitalisation occurs whenever tax is levied on income from a durable asset. It applies to structures and land, financial instruments, licences to run hotels and taxis, and indeed to any asset whose income is taxed. Although future owners of the asset pay the periodic taxes to the tax authority, the owner of the asset when the tax is announced bears the burden of the expected changes in after-tax income. If there is doubt whether a tax will be legislated or retained, there will be less than a one-off full fall in the value of an asset when it is announced. On the other hand, if people fear that government may increase the tax later, the price of an asset may fall by more than 100 per cent of the present value of actual future tax payments!

General Equilibrium Analysis

Most of our discussion of tax incidence has focused on the impacts of taxes in single markets. We have also considered some related or multi-market effects. For example, we examined how a tax on beer can affect the price and output of wine and how a tax on payrolls of large firms can affect wages and employment in untaxed firms. These were first steps towards a general equilibrium (economy-wide) analysis but we were working primarily with single markets and using partial equilibrium (PE) models.

In many cases PE models can explain the full incidence of a tax. Moreover, the basic drivers of incidence (relative demand and supply elasticities) are the same in economy-wide models as in PE models. But a PE model cannot capture all the effects of large tax changes across many markets, for example the impacts of a tax on carbon. For such changes we are likely to need a general equilibrium (GE) model of the economy. These models can be very large and complex and run into thousands of equations in a computable GE model. However, we can gain important insights into the effects of a tax change on an economy by drawing on the simple but classic GE model proposed by Harberger (1974).

Suppose that an economy contains two factors of production, capital and labour, and produces two commodities, manufactures and services. In this model the total supply of capital and labour is fixed. However, capital and labour are mobile and can move between the manufacturing and service sectors. Also, in this economy all income is consumed; there are no savings. This economy is depicted in Figure 26.9.

In this economy there are nine possible *ad valorem* taxes.

- A general income tax. In the absence of savings, a general tax on income from labour and capital is equivalent to a uniform tax on expenditure on manufactures and services.
- Two selective factor taxes: a tax on labour or on capital.
- Two selective commodity taxes: a tax on manufactures or on services.
- There are also four partial factor taxes. A partial factor tax is a tax levied on a factor in a selective use. The four partial factor taxes are a tax on capital used to produce manufactures or services or a tax on labour in one of these activities.

To analyse the effects of these taxes, following Harberger we assume that factors of production are fully employed, the economy is competitive. and prices and wages are flexible. Suppliers of capital and labour seek to maximise their after-tax income. Therefore, the after-tax marginal return to capital and to labour is the same in manufacturing and services. Factors of production are paid the value of their marginal product. We also assume initially that consumers have similar preferences. This means that changes in the distribution of income between labour and capital do not affect the use of income.

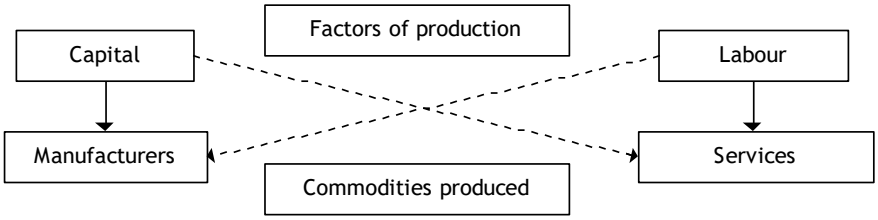


Figure 26.9 A simple general equilibrium model

We now examine the incidence of the four main types of *ad valorem* taxes starting with a **general income tax**. In this model, because factor supplies are fixed, this tax is borne wholly by income earners. The owners of the factors bear the cost; they cannot escape the taxes. Further, because capital and labour are the source of all income in this model, a general tax on income is equivalent to a uniform tax at the same rate on both factors of production. In addition, with no savings, income equals expenditure. With these assumptions, a general tax on income is equivalent to, and has the same incidence as, a uniform tax on consumption of manufactures and services.

Consider, secondly, **selective factor taxes**. Because labour supply is fixed labour bears the full cost of a selective tax on earnings of labour. With a tax on labour in manufacturing and services, labour has no incentive to switch from one sector to the other. Likewise, with a fixed supply of capital, a tax on capital is borne by owners of capital.

The incidence of **selective consumption taxes** is more complex. However, it is relevant to many economies, including Australia where the GST covers less than 60 per cent of all goods and services. Consider, for example, a tax on manufactures which raises their price. The consumption of manufactures falls and demand for services rises. Some capital and labour move from manufacturing into services. At this point, production technology is important. Suppose that manufacturing is capital intensive and services are labour intensive. Services will have to absorb a relatively large amount of capital. This causes the price of capital used in the services sector to fall. The size of the fall depends on the ease with which capital can be substituted for labour. The greater the substitutability of capital for labour, the lower the fall in price of capital. However, because capital used in services cannot earn less than capital used in manufacturing, the return to capital in both sectors must fall. In general, a tax on the output of one sector induces a decline in the relative price of the input that is used relatively intensively in that sector. Moreover, the greater the elasticity of demand for services, the greater the switch from manufactures to services and therefore the greater the decline in the relative price of capital.

Now allow consumer preferences to vary. Suppose that individuals with labour income prefer manufactures and individuals with capital income prefer services. If income from labour rises relative to income from capital, demand for manufactures will increase. A GE model would incorporate the effects of changes in the distribution of income on changes in demand for manufactures and services.

Fourth, we consider the incidence of a **partial factor tax**, for example a payroll tax on labour used in manufacturing. Such a tax has output and substitution effects. The output effect occurs because a tax on labour in manufactures raises the price of manufactures. This reduces consumption of manufactures and increases the demand for services. Labour and capital move from manufacturing to services. If manufacturing is capital intensive, a large amount of capital must be absorbed in the labour-intensive service sector and the price of capital falls relative to the price of labour. If manufacturing is labour intensive, labour must be absorbed

in the services sector and the relative price of labour falls. Thus, a tax on labour in manufactures has output effects that change the relative price of labour and capital, depending on the factor intensity of labour and capital in manufacture and services. The substitution effect arises because a tax on labour in manufacturing causes capital to be substituted for labour in manufacturing. This reduces the price paid to labour.

Combining output and substitution effects, if manufacturing is labour intensive the two effects work in the same direction. Both reduce the price of labour. If manufacturing is capital intensive, the switch in demand from manufactures to services increases the demand for labour and its price and reduces the price of capital. This output effect on the relative price of labour may more than offset the substitution effect. Thus, even though the tax is levied on labour, it can make owners of capital worse off. Conversely, a tax on capital in a labour-intensive sector can make labour worse off. The general point is that a tax on a factor in one sector ultimately affects the returns to factors in both sectors.

The Harberger model illustrates the key features of general equilibrium models, especially how a change in relative prices in one commodity or factor market can affect prices in other markets. More detailed models expand the range of consumers, factors and sectors that may be affected by taxes. They also relax the assumption that factor supplies are fixed and often add a time dimension to the behaviour of consumers and producers. Cordes and Watson (1998) provide a guide to extensions of the model.

Further Issues in Tax Incidence

We have seen that some taxes are equivalent to others. For example, a general tax on income is equivalent to a general tax at the same rate on all factors of production. We discuss below equivalent taxes and the implications for the overall incidence of a tax system. We see why a comprehensive income tax causes savings to be taxed twice. We also examine the incidence of a general value added tax (VAT or the equivalent goods and services tax, GST, in Australia), and see when the incidence is similar to a tax on income or one on consumption.

Equivalent taxes

When the incidence of two sets of taxes is the same, the taxes are called equivalent. Table 26.1 shows some tax equivalence relations in a single period with no savings. There are again two factors of production, capital (K) and labour (L), and two commodities, manufacturing (M) and services (S). There is an *ad valorem* tax rate (t) and Y is income. The table indicates six sets of equivalent taxes (one in each of the three rows plus one in each of the three columns). We have already met the equivalences in the last row and last column—the equivalence of income tax to a uniform and equal tax on either the factors of production or the consumption of manufactures and services (if there are no savings). Partial tax rates on capital (labour) used in manufacturing and services are equivalent to a general tax on capital (labour). Also, equal tax rates on capital and labour used in producing manufacturing (services) are equivalent to a general tax rate on manufactures (services).

Table 26.1 Tax equivalence relations with no savings

t_{KM}	and	t_{LM}	are equivalent to	t_M
and		And		and
t_{KS}	+	t_{LS}	are equivalent to	t_S
are equivalent to		are equivalent to		are equivalent to
t_K	and	t_L	are equivalent to	T_Y

Source: Rosen and Gayer (2014, p. 315) adapted from McClure (1971).

Introducing savings. So far savings have been excluded. Introducing savings complicates matters. A general tax on consumption is equivalent to a general tax on income that excludes income from capital. In effect this is equal to a uniform tax on income from labour (wages).[4] This can be shown using a two-period model. Suppose Emma earns a wage income (W_1, W_2) in two periods and saves nothing. The present value of her lifetime income, $PV(Y)$, is:

$$PV(Y) = W_1 + \frac{W_2}{(1+r)} \tag{26.16}$$

where r is her rate of time discount. Because Emma consumes exactly her wage income in each period, the present value of her consumption (C) is the same as the present value of her wage income.

$$PV(C) = C_1 + \frac{C_2}{(1+r)} = PV(Y) \tag{26.17}$$

Now suppose that Emma consumes less than her wage in period one. She would save ($W_1 - C_1$). Her consumption in period two would be the sum of her wage and her savings plus interest.

$$C_2 = W_2 + (W_1 - C_1)(1+r) \tag{26.18}$$

This implies that:

$$C_2 + C_1(1+r) = W_2 + W_1(1+r) \tag{26.19}$$

and

$$C_1 + \frac{C_2}{(1+r)} = W_1 + \frac{W_2}{(1+r)} \tag{26.20}$$

Given perfect bond markets and no tax, the present value of wage income equals the present value of consumption, whatever the pattern of consumption.

Now introduce taxation (t). With no savings the present value of the wage after tax is:

$$PV(Y) = W_1(1-t) + W_2\frac{(1-t)}{(1+r)} \tag{26.21}$$

To consider the effect of savings, we assume that Emma saves S_1 of her wage in period one and consumes these savings plus after-tax interest and her regular wage in period two. The present value of her consumption is then:

$$PV(C) = W_1(1-t) - S_1 + \frac{S_1}{(1+r)} + \frac{S_1 r(1-t)}{(1+r)} + \frac{W_2(1-t)}{(1+r)} \tag{26.22}$$

Emma gives up S_1 and gains $S_1[1 + r(1 - t)]/(1 + r)$. If the returns from savings are exempt from tax, Emma would gain as much as she had forgone. A tax on wages would be the same as a tax on consumption. Savings can be exempted by taxing an individual's labour income and exempting returns on capital or by taxing individuals on their total income less their savings, which is their consumption.

However, if the returns on savings are taxed the present value of consumption is less than the present value of wages. In effect, there is double taxation of savings. To illustrate this, consider the choice between consuming an after-tax income of $W_1(1 - t)$ in year one and investing this after-tax income at a rate of return r in perpetuity. Note that the return would be taxed in each year. The present value of the perpetual investment equals:

[4] In this analysis inheritances and bequests are ignored. It can be shown that a tax on labour income plus inheritances and gifts is equivalent to a tax on consumption plus bequests and gifts.

$$PV(C) = W_1\frac{(1-t)r(1-t)}{(1+r)} + W_1\frac{(1-t)r(1-t)}{(1+r)^2} + \ldots + W_1\frac{(1-t)r(1-t)}{(1+r)^{\infty}} \qquad (26.23)$$
$$= W_1\frac{(1-t)r(1-t)}{r} = W_1(1-t)(1-t)$$

Equation 26.23 shows that savings are taxed twice.[5] Savers consume less of their lifetime income in terms of present values and pay more in tax than do non-savers. We now consider the incidence of a general value-added tax (VAT). The incidence depends on how investment is treated.

Incidence of a general value-added tax

In the most common form of VAT, used in Australia and the European Union, capital goods are effectively exempt from VAT. Accordingly, VAT works as a consumption tax. However, it can be designed as an income tax equivalent inclusive of savings.

An example may illustrate these statements. Suppose that a shirtmaker purchases wool and machinery, manufactures shirts and sells the shirts to retailers. Table 26.2 shows the value-added process and the implications for the value-added tax. The shirtmaker purchases wool and machinery valued at $3000 and sells the shirts to a retailer for $6000. The retailer sells the shirts to customers for $8000. Column 3 shows the value added at each stage in this process. Value added is defined here as the difference between the firm's sales and purchases. In effect the shirtmaker writes off the full cost of the machinery purchased in this accounting period. The total value added in this process is $8000. Column 4 shows the VAT paid with a 10 per cent tax rate applied at each stage of production.

In the Australian (GST) system the shirt manufacturer would invoice the retailer $600 for GST and claim a rebate of $300 for GST paid on inputs. The total GST paid is $800, which equals 10 per cent of the value of the retail sales. With this design, the GST (or the VAT) is a consumption tax. It is equivalent to a general sales tax on consumption goods.

This definition of value added underestimates the real value added of the shirtmaker because it overestimates the real cost of the machinery. Suppose that the shirtmaker can deduct only the amount by which the machinery depreciates, say 20 per cent in this period, or $400. The real value added, and income, of the shirt manufacturer is $4600. This is $6000 (sales) less $1000 for materials and $400 for use of machinery. Total value added in the whole process would be $9600, which is the real value of income earned. Also, if the shirtmaker paid $460 VAT on the real value added of $4600, total VAT would be $960. Expressed another way, the real income of all factors in this period is $9600. This output

Table 26.2 Example of a value-added tax for a shirt manufacturer

	Purchases	*Sales*	*Value added*	*VAT @10%*
Inputs				
Material (wool)	0	1000	1000	100
Machinery	0	2000	2000	200
Total inputs	3000	3000		
Manufacturer	3000	6000	3000	300
Retailer	6000	8000	2000	200
Total			8000	800

[5] The present value of any perpetual income stream (y) equals y/r, where r is the discount rate.

comprises $8000 of consumption goods and $1600 of investment goods. If firms can deduct only the amount by which investment goods depreciate, the tax base becomes total income inclusive of investment output. With this tax design, the VAT is equivalent to a uniform income tax. It is also equivalent to a general sales tax on consumption and investment goods.

Balanced budget incidence

So far, we have considered only the incidence of taxes. To obtain a complete picture of tax incidence we should also account for the way in which taxes are spent. Tax incidence analysis that takes account of both the incidence of taxes and the benefits received from revenue raised is known as balanced budget incidence.

Where the benefits accrue mainly to low income households, taxes may be proportional or even regressive and the balanced budget incidence can be progressive. Suppose that all taxpayers are taxed a given percentage of their income to provide health care services only for households on below average incomes. The tax would be proportional, but the balanced budget incidence would be progressive. In the section below, we look at the distribution of cash and various other benefits as well as the distribution of taxes.

Tax Incidence in Australia

To show the estimated incidence of most taxes as well of benefits we draw on the ABS (2007) analysis of household incomes in 2009–10 (which appear to be the latest such estimates). The results are shown by equivalised private income quintile in Table 26.3. Equivalised income allows for the different sizes and composition of households. Using the modified OECD scale, in these calculations the first adult in a household is given a weight of one, each extra person who is 15 years or older is allocated 0.5 points and each child under the age of 15 is allocated 0.3 points. Equivalised household income is total household income divided by the sum of the household's equivalence points.

Private income includes income from labour and capital. Social assistance benefits in cash are direct Australian government benefits. Social transfers in kind are the value of government transfers in kind for education, health, housing and social security. This totalled just over 50 per cent of all Commonwealth, state and local government expenditure.

Table 26.3 Distribution of household income, benefits and taxes by equivalised private income quintile, 2009-10

	Lowest quintile %	*Second quintile* %	*Third quintile* %	*Fourth quintile* %	*Highest quintile* %
Private income	3	9	16	24	48
Benefits					
Social assistance benefits in cash	58	26	10	4	2
Social transfer in kind	30	22	18	15	14
Total	39	24	16	12	10
Taxes					
Taxes on income	0	4	12	23	61
Taxes on production	14	15	18	22	31
Total	6	9	14	23	49
Final income	13	14	16	21	36
Equivalised private income	2	10	17	25	46
Equivalised final income	13	14	17	21	35

Source: ABS, *Government Benefits, Taxes and Household Income*, Cat. No. 6537.0, 2009-10.

The distribution of direct taxes was derived from income tax statistics. The distribution of indirect taxes, including the goods and services tax, was estimated based on expenditure data obtained from the 2009-10 *Household Expenditure Survey*. Overall 60 per cent of taxes on production were allocated to households. The study assumed that households bear all the costs of both income tax and consumption taxes and that there is no shifting of the taxes to employers or to suppliers of goods and services. Thus, it may not reflect the complete or real economic incidence of taxation that we have discussed in much of this chapter.

It may be noted that in Table 26.3 the data in most of the table refer to the households in each quintile. The last two rows are slightly reweighted based on the 20 per cent of all individuals in each quintile. This is described as a person weighting and gives equal weighting to people in large households to those in smaller ones although the incomes are still based on the concept of an equivalised household.

Taking the weighting in the final two rows, in 2009-10 individuals in households in the highest quintile received 48 per cent of all private income and those in the lowest quintile received only 3 per cent. After direct and indirect taxation and the distribution of benefits, the percentages change to 35 per cent and 13 per cent respectively.

Overall taxes were broadly proportional to private income. The income tax was highly progressive but taxes on production were regressive. However, the provision of social assistance in cash and in kind was highly progressive, especially benefits in cash.

Summary

- Economic tax incidence is often different from statutory incidence. Statutory incidence indicates who is legally responsible for paying a tax. Economic incidence shows which party actually bears the tax.
- In a competitive market, the burden of a commodity tax depends on the relative price elasticities of demand and supply. Consumers bear most of the tax when demand is inelastic and supply elastic. Producers bear most of the tax when supply is inelastic and demand elastic. Taxes are borne to the extent that they cannot be escaped.
- A monopolist, or any firm with some market power, can pass on more of a commodity tax to consumers when it can discriminate between customers, the marginal revenue curve is non-linear, the supply curve is elastic, and the tax is *ad valorem* rather than a unit tax.
- The analysis of the incidence of taxes on factors of production is similar to the analysis of commodity tax incidence. The effective incidence of a tax on a factor depends on the relative elasticities of demand and supply, not on who nominally pays the tax.
- The owners of land, capital and labour bear most of the tax when supply is inelastic and demand elastic. Users of factors bear most of the tax when their demand for factors is inelastic and the supply is elastic.
- Tax capitalisation occurs when a stream of present and future taxes is incorporated into the present capital value of an asset.
- When the incidence of two sets of taxes is the same the taxes are called equivalent taxes. A general tax on income is equivalent to a general tax on all factors of production. It is also equivalent to a comprehensive and uniform sales tax on all output, including investment goods. A general tax on consumption is equivalent to a general tax on labour income (which excludes income from savings).
- Partial equilibrium analysis examines the incidence of price and wage changes within markets. General equilibrium analysis examines the incidence of price and wage changes across the economy. It shows how a change in relative prices in one commodity or factor market can affect prices in other commodity or factor markets.
- In Australia, the overall tax system is broadly proportional. The income tax is progressive, but consumption taxes are regressive. However, the distribution of benefits, especially cash benefits, is highly progressive.

Questions

1. Why is the statutory incidence of taxation often irrelevant when determining the actual effects of taxation? What determines the economic incidence of a commodity tax?
2. Under perfect competition to what extent does the incidence of a commodity tax depend on whether the tax is a unit tax or an *ad valorem* tax? What difference will an *ad valorem* tax have under monopoly production?
3. Will an income tax be borne entirely by workers or will firms bear any of the tax?
4. Who bears the incidence of a partial payroll tax levied on firms with a labour payroll in excess of one million dollars?
5. What is tax capitalisation and when is it likely to occur?
6. Assume that a good is fixed in supply at 30 units. Demand for the good can be represented by the demand equation $Q^d = 50 - 4P$ where P is the price per unit. If government imposes a tax on the producer equal to \$4 per unit determine:
   i. The price paid by consumers before and after the tax is imposed.
   ii. The price producers receive before and after the tax is imposed.
   iii. The amount of revenue raised as a result of the tax.
7. Suppose that demand for shirts is represented by the equation $Q^d = 80 - 3P$ and supply of shirts is given by $Q^s = 40 + P$. If a commodity tax of \$2 per shirt is imposed on production of shirts, determine:
   i. The pre-tax equilibrium price and quantity combination.
   ii. The post-tax equilibrium price and quantity combination.
   iii. The burden of tax borne by consumers and producers.
8. Suppose that Bruce earns \$40 000 each year, and saves a quarter of that income in year 1 for consumption in year 2.

   If the rate of interest is 5 per cent and the tax on wages or consumption is 20 per cent determine the following:
   i. The present value of Bruce's income and consumption without taxation.
   ii. The present value of Bruce's income and consumption with tax.
   iii. What inference can be drawn about taxation equivalence?
9. The auction house Sotheby's traditionally charged a percentage of the sale price as a commission on both vendor and purchaser and was heavily criticised for double charging. Show that the final price to the consumer is the same regardless of whether Sotheby's charges the whole commission to the vendor or to the purchaser or splits the commission equally or in any other proportion between them. What is the implication for commodity taxation?
10. An economy has two factors of production, labour and capital. Labour is in inelastic supply (there is no immigration). Capital is in perfectly elastic supply (as in a small open economy). Drawing on the Harberger general equilibrium model, show how all taxes, whether on labour or capital or output, will fall on labour. How does this outcome change with migration?
11. Suppose that there exists a general consumption tax such as a VAT or GST that covers two-thirds of all consumption goods. Are firms likely to pass all the tax on to consumers? When would they not do so?
12. What are the main differences between an income tax, a social security levy on employers and a payroll tax? Do they have any different real economic incidence?
13. What methods would you employ to determine the incidence of a carbon tax? What results would you expect?

Further Reading

Australian Bureau of Statistics (2007) *Government Benefits, Taxes and Household Income,* Cat. No, 6537.0.

Cordes, J.J. and Watson, H.S. (1998) 'The theory and practice of tax incidence', pp. 93–134 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Creedy, J. (2001) 'Tax modelling', *Economic Record*, 77, 189–202.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapter 19, Worth Publishers, New York.

Rosen, H. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapter 13, McGraw-Hill, New York.

Seligman, E.R.A. (1923) 'The effects of taxation', *Political Science Quarterly*, 38, 47–61.

Stiglitz, J. (2000) *Economics of the Public Sector*, Chapter 18, W.W. Norton, New York.

Warren, N., Harding, A. and Lloyd, R. (2005) 'GST and the changing incidence of Australian taxes: 1994–95 to 2001–02', *eJournal of Tax Research*, 3, 117–149.

Chapter

27

Taxation and Efficiency

There can be no economy where there is no efficiency.

Disraeli, Letter to Constituents

The Deadweight Loss of Taxation ♦ Taxation of Labour Income ♦ Taxation of Income from Capital ♦ Selective Commodity Taxes ♦ General Taxation of Consumption ♦ Actual Deadweight Losses of Taxation

Most taxes change behaviour, distort economic activity and create a deadweight loss. This chapter describes how the main forms of taxation cause a deadweight loss and how this loss can be quantified. Following an introduction to the deadweight loss of taxation, we examine the deadweight losses due to income taxes on labour and capital and the losses associated with consumption taxes. The final section discusses the overall deadweight loss of a taxation system.

The Deadweight Loss of Taxation

A **deadweight loss** (DWL) is an economic loss for which there is no offsetting benefit. When associated with taxation, DWL is sometimes described as excess burden. In this chapter, we use the terms DWL and excess burden interchangeably but most often use the term DWL.

Deadweight loss
An economic loss with no offsetting benefit—sometimes called excess burden

An efficient tax imposes burdens on individuals by reducing their income, but it does not cause an excess burden associated with behaviour change. In other words, an efficient tax has income effects but no substitution effects. A head tax is an example of an efficient tax because it reduces an individual's income by a given amount independently of any action that he or she might take. Such a tax is a transfer payment from taxpayer to government. One person's tax is offset by government revenue gain. There is no DWL to the community.

To illustrate the concept of DWL, consider a tax on income. Suppose that Emma is willing to work 20 hours a week for a minimum annual payment of $20 000 and that, in line with the value of her marginal product, she receives $24 000. Emma has a surplus from work of $4000 a year. If Emma's income were taxed at a rate of 25 per cent, she would receive only $18 000. Because this after-tax income would not compensate for the loss of her household production and leisure, Emma changes her behaviour and does not take on paid employment. Accordingly, she would pay no tax. Nevertheless, the tax imposes a DWL on society. Emma has a net welfare loss of $4000. But there is no offsetting benefit to anyone else.

To quantify DWL, we draw on our discussion of welfare measurement in Chapter 6. Welfare measures may be based on the principle of compensating or equivalent variation (CV or EV). Adopting the CV principle, measures of welfare changes are based on the expenditure

that would maintain the existing level of welfare (utility) in society *before* any change is made. The cost of a tax increase to someone is the minimum dollar amount that she would be willing to accept as compensation for the increase in tax and be no worse off than before. The benefit of a tax reduction is the maximum amount that she would be willing to pay for the reduction and be at least as well off as before.

With the EV principle, measures of welfare changes are based on the expenditure that would maintain the level of real income *after* taxes change. The cost of a tax increase to someone is the maximum amount that she would be willing to pay to prevent it, given their new after-tax income. The benefit of a tax reduction is the compensation that she requires to forgo the reduction in tax. The benchmark here is the individual's level of real income (or utility) after the proposed change.

These measures of welfare change reflect an individual's loss or gain, inclusive of income effects. When a tax is imposed, the loss to an individual is the sum of the loss of income (the burden) and the loss over and above the tax paid (the DWL or excess burden). The DWL is due to the substitution effect of a tax. To illustrate this with another example, consider a tax on beer which causes Charles to reduce his beer consumption from 12 to 10 bottles a week. The weekly burden to Charles is the extra cost of the 10 bottles of beer (the income effect) plus the loss of surplus from forgoing the extra two bottles (the substitution effect). However, the tax that Charles pays is income to government and will provide benefits or lower tax to Emma. There is no net cost arising with this transfer payment. On the other hand, the DWL (or excess burden) of the tax associated with the substitution effect has no offsetting benefit.

Substitution effects occur when quantities change, be they quantities of labour or consumption, in response to changes in relative prices (holding real income constant). However, changes in prices may not change quantities when substitution and income effects work in opposite directions. For example, a tax on labour encourages individuals to *substitute* from market goods purchased from labour income which is taxed to leisure and home production which are not taxed. The *income* effect arises because individual utility is a function of market goods and leisure and, as consumption of market goods falls, the demand for leisure falls and individuals work more. DWL is caused by the substitution effect and to measure this loss substitution and income effects must be separated.

Given that nearly all taxes create some DWL, an efficient tax system is the one that creates least total DWL. This tax system would equate the marginal DWL per dollar of tax raised across all revenue sources. Of course, such an efficient tax system would not be socially optimal. To determine social optimality, which depends on equity as well as efficiency, it would be necessary to compare and weight gains and losses to all individuals in society. This complex exercise is discussed in the next chapter.

In this chapter we discuss efficiency issues and examine the DWLs of various taxes without making interpersonal comparisons of utility. Tax A is more efficient than tax B if, for a given level of taxpayer welfare, it raises more tax revenue than tax B. Tax A is also more efficient than tax B if, for a given level of tax revenue, it leaves at least one taxpayer better off. In both cases, tax A would make either the taxpayer or the tax recipient better off without making the other worse off. Formally, it is Pareto efficient.

Taxation of Labour Income

A tax on labour income may affect many aspects of labour supply. These include participation in the workforce (entry or exit choices), choice of occupation, hours of work and work effort. Most literature (for example, the major Australian review by Dandie and Mercante, 2007) focuses on the effects of income tax on hours of work and we follow this practice here, but we also note some of the other effects of income tax on labour.

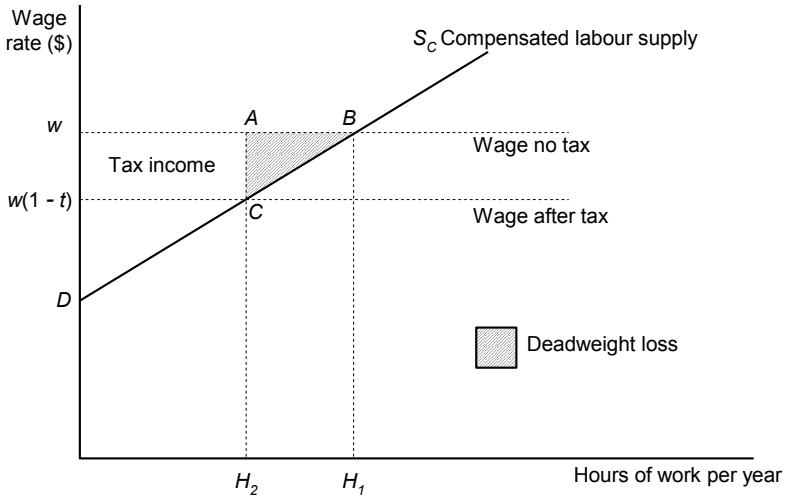


Figure 27.1 Deadweight loss of a tax on labour

The DWL of a tax on labour is shown in Figure 27.1. In a competitive market, the wage rate (w) for Emma is represented by a horizontal schedule; it does not vary with Emma's hours of work. Emma's supply curve is here a **compensated labour supply curve**. This shows the lowest wage for which Emma is willing to work for each extra hour, holding her utility from consumption of market goods and leisure constant. Thus, it shows the substitution effect on her labour supply of changes in the relative price of market goods and leisure, holding utility constant As the wage rate falls, Emma reduces her work. Without any tax, Emma enjoys total surplus equal to triangle wBD, which is the difference between her wage and the minimum amount that she requires as compensation for loss of leisure. Now if a tax is imposed on the wage at a rate (t), Emma receives an after-tax wage of $w(1- t)$ and reduces work hours from H_1 to H_2. She pays a tax equal to area $wACw(1- t)$. But she also loses surplus equal to triangle ABC. This is the DWL due to the tax-induced fall in the labour supply. It is the amount by which Emma's loss of welfare exceeds the transfer to government.

Compensated labour supply
A supply curve that shows how quantity of labour supplied varies with the wage rate, holding utility constant

Income and substitution effects

Unlike the compensated labour supply curve, the ordinary (uncompensated) labour supply curve includes income effects. A person's ordinary labour supply curve may be perfectly inelastic even though his or her compensated labour supply curve is upward sloping.[1] As Emma's net wage falls, the relative price of leisure falls and Emma substitutes leisure for market goods. But the income effect works the other way because the demand for leisure falls as Emma's income falls. Overall, a tax on labour income may have little or no effect on Emma's hours of work. However, there may still be a DWL due to the substitution effect.

Figure 27.2 overleaf illustrates the substitution and income effects. The line GL shows Emma's initial budget constraint with hours of leisure on the horizontal axis and consumption of market goods (or income) on the vertical axis. A tax on labour income rotates her budget line to G_2L. The tax reduces her potential consumption of goods, but it does not affect the total amount

[1] An individual's labour supply curve may even be backward bending. As the wage falls an individual may work more hours if the income effect of the fall is greater than the substitution effect.

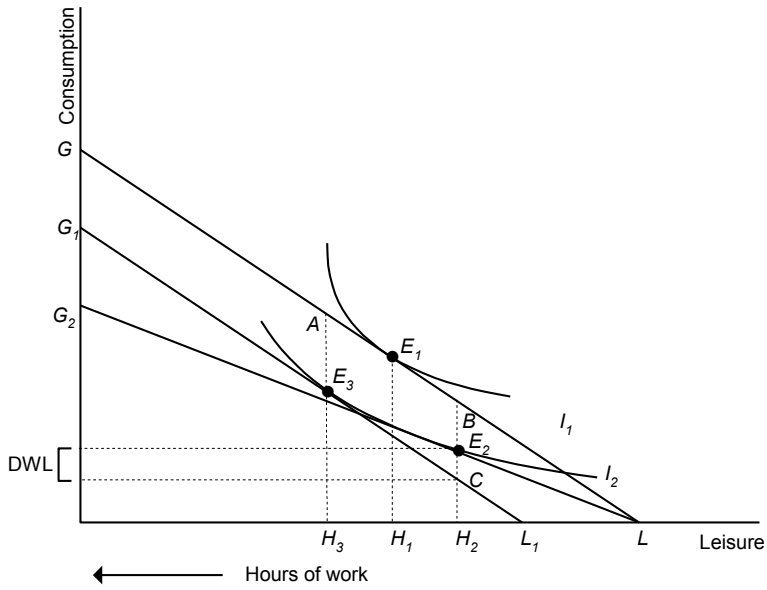


Figure 27.2 Tax on labour: income and substitution effects

leisure that she may enjoy. Initially Emma chooses position E_1 and works H_1 hours (note that hours of work are measured from point L). After a tax on labour income is introduced, Emma chooses E_2 and works H_2 hours, which is a slight fall in hours of work. Now the move from E_1 to E_2 can be decomposed into income and substitution effects

The income effect is measured by holding the relative price of leisure and other goods constant (thus excluding any substitution effects) and drawing a line parallel to the initial budget line (GL) that is tangent to the indifference curve I_2 at E_3.

The parallel line G_1L_1 represents the lump sum tax that would leave Emma on the same indifference curve as at E_2. Given a lump sum tax, Emma would choose the combination of goods and leisure shown by point E_3. The move from E_1 to E_3 represents the income effect. At this point, Emma works more hours (H_3) than at E_1. However, the tax on labour causes Emma to substitute leisure for work. The substitution effect is reflected in the move from E_3 to E_2. What then is the DWL of a tax on labour income compared with a lump sum tax? Emma enjoys the same level of utility at E_3 and E_2 even though she pays more tax at E_3. At E_3, she pays AE_3 tax. At E_2, she pays BE_2 tax. The revenue loss (E_2C) is the equivalent variation measure of the DWL of the tax on labour.

Individuals may also adjust to a tax on labour income by changing occupations. These changes may also generate a DWL. Suppose that Emma works in a stressful marketing job for a gross income of $60 000 a year but could work in publishing for $50 000 a year which she would enjoy more. Emma requires $6000 to compensate her for working in marketing compared with publishing. Before tax, she has a surplus of $4000 from working in marketing. However, suppose that, after tax, Emma would earn $40 000 from marketing and $35 000 from publishing (this implies a marginal tax rate of 50 per cent on income over $50 000). Emma now chooses to work in publishing. As a result, she pays $15 000 tax and bears an excess burden of $4000 (the loss of the surplus she obtained from her marketing job). Put another way, a lump sum tax of $15 000 would raise the same amount of revenue as the tax on labour, but Emma would stay in marketing and continue to enjoy a net benefit of $4000.

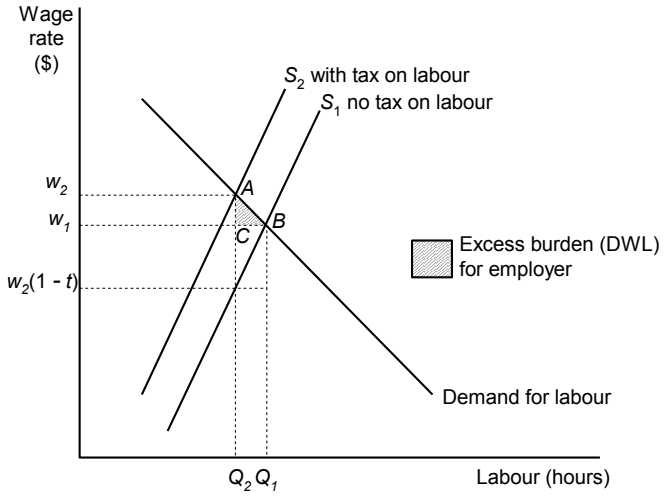


Figure 27.3 Tax on labour: excess burden on employers

So far, we have examined the excess burden on workers of a tax on labour. But as we saw in Chapter 26, a tax on labour may increase gross wages. Consequently, firms may also bear an excess burden. Figure 27.3 shows a market demand curve for labour along with an ordinary (uncompensated) supply curve for labour. A tax on labour shifts the supply curve upwards from S_1 to S_2. Total hours worked fall from Q_1 to Q_2. The gross wage rises from w_1 to w_2. The labour tax paid by firms is given by area w_2ACw_1. The excess burden borne by firms equals area ABC.

Quantifying deadweight loss

Assuming an elastic demand for labour and a linear compensated labour supply curve, as in Figure 27.1, the DWL of a tax on income from labour is given by:

$$DWL = 0.5\ (Q_1 - Q_2)\ (w_1 - w_2) \tag{27.1}$$

where Q is the (compensated) quantity of labour supplied, w is the net wage received and subscripts 1 and 2 refer to the quantities and wages before and after-tax respectively. Thus to estimate DWL, we need to estimate the change in the net wage received and the compensated changes in quantity of labour supplied.

To understand the factors that determine the deadweight loss it is helpful to express Equation 27.1 in terms of the compensated elasticity of supply (η_{cs}) and the tax rate (t), which is the ratio of the tax (T) to the initial wage. Now the elasticity of compensated supply (η_{cs}) is:

$$\eta_{cs} = (\Delta Q/Q_1)/(\Delta w/w_1) \tag{27.2}$$

So
$$\Delta Q = (\Delta w/w_1)\ \eta_s Q_1 \tag{27.3}$$

Also
$$\Delta w = (w_1 - w_2) = T \tag{27.4}$$

so
$$DWL = 0.5\ (\Delta w/w_1)\ Q_1\eta_{cs}T = 0.5\ Q_1 w_1 \eta_{cs} t^2 \tag{27.5}$$

The DWL is the product of half the initial wage payments (Q_1w_1), the compensated elasticity of labour supply and the square of the tax rate.[2] If the payments base is initial wage

[2] If the demand curve for labour is not perfectly elastic, DWL includes a loss of employer surplus as in Figure 27.3. However, there is a smaller fall in labour employed. The total DWL is then equal to $0.5wQt^2/((1/\eta_d)+(1/\eta_{cs}))$ where η_d is the elasticity of demand for labour (Bishop, 1968). This expression equals Equation 27.5 when $\eta_d = \infty$.

payments as in Equation 27.5, the estimate is a compensating variation measure of welfare loss. If the base is post-tax wage payments, the estimate is an equivalent variation measure. In either case, the critical determinants of DWL are the labour supply elasticity and the tax rate. DWL rises proportionately with the elasticity of supply and with the square of the tax rate. The latter is an important result because it implies that broad-based low tax rates on a comprehensive tax base are more efficient than narrowly based high tax rates which collect a similar amount of tax revenue.

Average deadweight loss
Total deadweight loss divided by total tax revenue

To illustrate the DWL of a tax on labour income, suppose that Charles is willing to work 100 hours a month at $30 per hour, the tax rate is 40 per cent and Charles' ordinary labour supply elasticity is 0.25. Given these parameters, Charles will reduce his work hours by 10 per cent (work 90 hours) and pay $1080 in tax per month (90 hours × $30 × 0.4). Assume also that Charles' compensated supply elasticity is 0.5. Applying Equation 27.5, total DWL is $120. **Average DWL** is 11.1 cents per dollar of tax revenue. If the tax rate were 50 per cent, DWL would rise to $187. Although the tax rate rises by 25 per cent, the DWL rises by just over 50 per cent. Thus, the **marginal DWL** is much higher than average DWL.

Marginal deadweight loss
The change in deadweight loss with a marginal increase in a tax

Elasticity of labour supply. Apart from tax rates, the compensated labour supply elasticity is a major determinant of DWL of tax on labour income. The higher this elasticity, the higher the DWL. However, estimates of labour supply elasticities vary widely with the type of labour—with age, sex, household composition, part- or full-time work and occupational context. They may also vary because of differences in data sets and estimation methods.

Labour supply elasticities are estimated most often from cross-sectional individual workforce data. Hours of work of individuals are regressed against after-tax wage rates, non-labour income and a set of control variables such as education, marital status, number of children and so on. These other factors affecting labour supply must be accounted for in order to identify accurately the effect of income on labour supply.

Gruber (2016, Chapter 21, pp.663-4) describes two other methods of estimating labour supply elasticities—experimental evidence and quasi-experimental evidence. As an example of experimental evidence, he cites a randomised experiment in the United States in which government randomly assigned low income individuals to alternative benefit-tax schemes between 1968 and 1976. The schemes varied in the size of the grant provided to individuals and the rate at which the grant was withdrawn as income rose, which is an implicit tax rate. Analysts were able to estimate how labour hours responded to the implicit tax rates. The estimated overall elasticity of labour was quite low at only about 0.1.

As an example of quasi-experimental evidence, Gruber cites analysis of labour supply before and after the 1986 *Tax Reform Act* in the United States. In general, labour supply before and after a major tax change provides useful data for analysis of labour supply. However, it is important to identify and take account of concurrent changes and other characteristics of individuals that may affect the labour supply.

Another complication is the distinction between uncompensated and compensated labour supply. The latter includes only the substitution effect. In the cross-section regression approach described above, the change in labour supply with respect to non-labour income (the coefficient on non-labour income) reflects only the income effect. On the other hand, the coefficient on after-tax labour income includes income and substitution effects. Subtracting the former coefficient from the latter gives an estimated substitution effect. If the income effect is negative, which it often is, the ordinary labour supply elasticity is lower than the compensated supply elasticity. For example, if the estimated ordinary labour supply elasticity is 0.2 and this includes an income elasticity of labour supply of say –0.2, the compensated supply elasticity would be 0.4.

Table 27.1 Estimates of uncompensated labour supply elasticities with respect to wages

Population group	*Australian studies*		*Other countries*[a]		
	studies (no)	*Range*	*Median*	*Studies (no)*	*Range*
Married men	5	-0.19 to 0.26	0.00	3	-0.29 to 0.24
Married women	11	-0.19 to 1.3	0.18	12	-0.37 to 0.71
Single men	1	0.28	0.28	1	0.63
Single women	1	0.34	0.34	1	0.82
Lone parents	3	-0.15 to 1.48	0.20	5	0.11 to 1.44

(a) Canada, New Zealand and the United Kingdom.
Source: Dandie and Mercante (2007).

Traditionally empirical studies (e.g. Heckman, 1993; Blundell, 1996) found that the ordinary (uncompensated) labour supply elasticity with respect to the after-tax wage is only marginally positive at about 0.1 for the primary income earner (most often males) and varies from 0.5 to 1.0 for secondary income earners. After taking out income effects, compensated labour supply elasticities may be 0.2 to 0.3 points higher than uncompensated supply elasticities. However, the results vary with individual and labour supply circumstances.

Dandie and Mercante (2007) reviewed estimated uncompensated elasticities in several Australian and other studies from Canada, New Zealand and the United Kingdom. The results, summarised in Table 27.1, are consistent with traditional findings though it should be noted that the groupings (married men, married women etc.) are not necessarily synonymous with primary and secondary earners. Note also that uncompensated labour supply elasticities can be negative where the income effect is greater than the substitution effect, whereas compensated labour supply elasticities are highly unlikely to be negative.

In a more recent study, drawing on data from Sweden in the 1980s, Blomquist and Selin (2010) estimated that the compensated labour supply elasticities with respect to wage rates were about 0.15 for males and 0.50 for females, with uncompensated elasticities only some 5 percentage points higher. However, Blau and Kahn (2007) found that married women's labour supply elasticities fell by just over 50 per cent between 1980 and 2000, which reflected significant changes in opportunities and in preferences towards work, children and leisure.

Finally, recall that taxation can affect labour supply in many ways, for example workers may move to less taxed, lower paid, jobs or put in less work effort. A significant consequence is that the elasticity of tax revenue to tax changes is almost always higher than the elasticity of labour supply hours. Blomquist and Selin (2010) estimate that the elasticity of tax revenue to tax changes for men is about 0.20 and for women rises to about 1.0.

Differential taxation of labour income

We have assumed so far that all labour income is taxed in a similar way, although it may be taxed at different marginal rates. We now consider the effect of taxing different types of labour income at different rates.

The most extreme differential occurs with home labour which avoids any taxation. In Figure 27.4 overleaf, panel (a) shows work in the market and panel (b) work at home. The value of the marginal product of labour (VMP) falls as more hours are worked. In competitive markets with no tax, if individuals are indifferent to where they work, they allocate time to equate VMP_M and VMP_H, where subscripts M and H indicate market and home respectively. The real wage in both sectors would be w_1. If the VMPs are not equal, an individual can increase his or her real income by doing more work in the sector with the higher VMP.

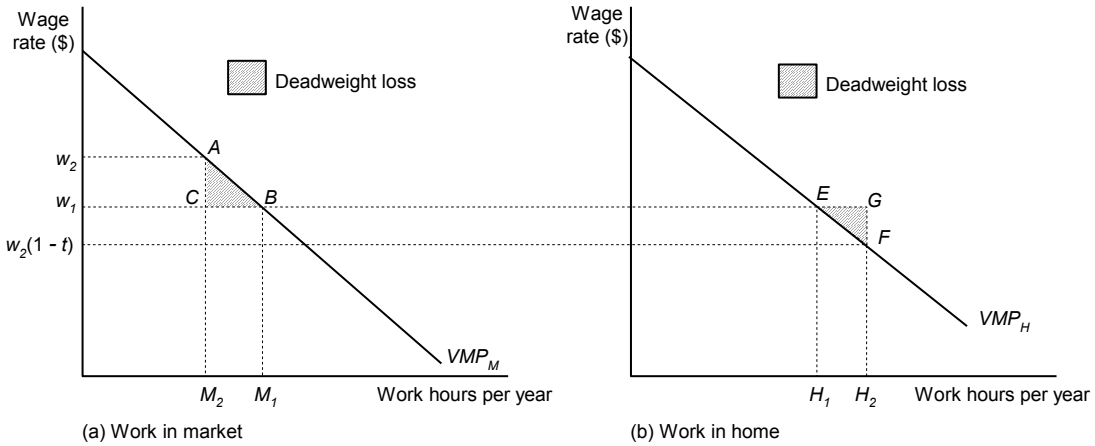


Figure 27.4 Deadweight loss of a differential tax on labour

If a tax is imposed on market income, the initial equilibrium no longer maximises take-home income. To maximise this income, individuals will reallocate time to the home sector. This shift causes a downward move along the VMP_H curve. A new equilibrium occurs when the after-tax return in the two sectors is equal, that is, when $VMP_H = VMP_M\ (1 - t)$ where t is the tax rate on market income.

Hours worked in the home increase from H_1 to H_2. If total working hours are fixed, hours worked in the market sector fall by an equal (offsetting) amount from M_1 to M_2. This is inefficient because the value of output lost in the market (area ABM_1M_2) is greater than the gain in output at home (area EFH_2H_1). The net loss in value of output is the DWL of the differential tax.[3] Equivalently, the DWL equals areas $ABC + DEF$. Boskin (1975) estimated that this loss equalled between 6 and 13 per cent of tax revenues in the United States. A similar method of analysis would apply to any differential taxation of labour income, for example a payroll tax on large firms and exemptions for small ones.

Taxation of Income from Capital

A comprehensive income tax taxes labour income that is saved and the returns from the savings (the capital). Like other taxes, a tax on income from capital has a substitution and an income effect. Such a tax increases the relative price of future consumption and so causes present consumption to be substituted for future consumption. On the other hand, because the tax reduces future consumption, individuals may save more to offset this loss. Thus savings may not fall because the income effect of a tax on savings may offset the substitution effect.[4] However, there would still be a DWL loss associated with the substitution effect.

To analyse the effect of a tax on income from capital, we examine how such a tax may affect consumption over time. Suppose Emma earns $50 000 in period 1 and expects to receive a pension of $20000 in period 2. And suppose that the real rate of interest is 5 per cent per period. If Emma saves, say $20000, she could consume $30 000 in the first period and $41 000 ($20000 + ($20000 × 1.05)) in the second. Assuming Emma cannot borrow against her pension, her consumption choices before tax are illustrated by the FF line in Figure 22.5, where

[3] This assumes that individuals are indifferent between working outside and inside the home.

[4] Suppose that $U = f(C_n, C_f)$ subject to $C_n + C_f/(1+r) = W_n + W_f/(1+r)$ where C is consumption, W is wages, n and f denote now and future, r is the discount rate. A tax on savings reduces total income and both C_n and C_f.

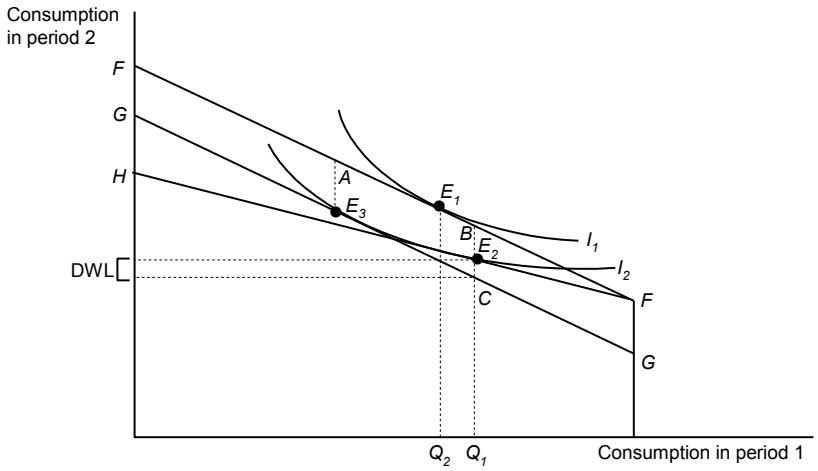


Figure 27.5 Excess burden of a tax on income from capital (savings)

consumption in periods 1 and 2 is shown on the horizontal and vertical axes respectively. By deferring part of her consumption, Emma can choose any point along this line. Emma must give up $(1 + r)$ units of marginal consumption in period two for an extra unit of consumption in period one.[5]

To maximise her utility, Emma trades present against future consumption. If a marginal unit of present consumption is valued at more than $(1 + r)$ future units, she would consume more in period 1. Emma maximises her utility when the marginal rate at which she is willing to substitute future for present consumption equals the rate at which she can exchange future consumption for present consumption, which is at point E_1. The convex indifference curves imply that, as Emma's consumption in either period falls, the marginal rate at which she is willing to substitute consumption in the other period rises. Now suppose that interest income is taxed. The budget constraint rotates to curve FH and Emma chooses position E_2 which sits on indifference curve I_2. In this example, Emma increases her consumption in period 1 but she reduces it in period 2.

To identify the welfare effects, income and substitution effects must be distinguished. As we have observed, a tax on future income reduces Emma's lifetime income and, with no change in relative prices, would generally reduce consumption in both periods. On the other hand, the tax makes future consumption more expensive relative to present consumption. Thus, Emma is likely to substitute present for future consumption.

To show the DWL, we draw a budget line GG parallel to the initial budget constraint and tangent to the new indifference curve I_2. With this budget constraint, Emma would choose point E_3. This represents the income effect with no substitution effect because the price of future consumption is unchanged. Now with a lump sum tax, Emma would pay AE_3 tax. On the other hand, with the tax on interest income, Emma pays BE_2 tax. Thus, for the *same* level of taxpayer welfare after the tax is imposed, a lump sum tax would raise additional tax revenue E_2C. This is an equivalent variation measure of the DWL of the tax.

Some complications may be observed. First, most taxes on interest income are levied on nominal rather than real income. This results in high real marginal tax on income.[6] Second, income from different forms of capital is often taxed at different rates. A high tax rate on one

[5] The method of analysis for taxation of income from capital is similar to the analysis of taxation of income from labour. The difference is that in this case consumption in period 2 is taxed instead of labour.

[6] Suppose that the nominal rate of interest is 10 per cent, the tax rate is 40 per cent and the inflation rate is 5 per cent. The real return on $\$100 = \$(100 + 10 \times 0.6) / 1.05 = \$100.95 - \$100 = \$0.95$. The real return after tax is less than 20 per cent of the real return without tax.

form of capital may divert savings into another form of capital with a lower gross return. Because this is a less productive investment, there is a DWL associated with this switch in investment.

Finally, not all the burden of the tax on income from capital is borne by savers. If this tax reduces savings and causes interest rates to rise, firms would pay more for capital. There is a DWL for firms who now employ less capital. The analysis of this loss is similar to that for firms that employ less labour because a tax on labour has increased wages (see Figure 27.3). The size of the DWL depends on tax rates and the elasticities of demand for and supply of capital for business.

The impact on savings of a tax on income from capital can be estimated by time series regression analysis in which quantity of aggregate savings is explained by the expected real after-tax rate of return to savings, disposable income and other relevant factors. A positive relationship between aggregate savings and the after-tax rate of return variable implies that a tax on income from capital would reduce savings. However, there are various problems with this kind of analysis. One is that estimates of expected real after-tax returns on savings require estimates of expected rates of inflation. A second is the variety of after-tax rates of return available in different investment vehicles. In addition, the regression analysis may not distinguish between income and substitution effects.

Such difficulties may explain the variety of estimates of the elasticity of savings with respect to the after-tax interest rate. In an early study of the responsiveness of savings to interest rates, Boskin (1978) estimated that the interest elasticity of savings was in the range of 0.2–0.4. Attanasio *et al*. (1995) reported an estimated elasticity of 0.67. Auerbach and Slemrod (1997) surveyed the literature and concluded that the overall elasticity of savings to interest rates is very low or zero because the income and substitution effects cancel out. This would imply that a tax on savings would have negligible impact on borrowing costs and investment. Hence there would be no deadweight loss from reduced investment by firms.

Nevertheless, from a saver's perspective changes in after-tax interest income have contrary income and substitution effects. A low total elasticity of savings with respect to the after-tax interest rate could be consistent with a pure substitution effect in the order of 0.3–0.4. This could result in a significant deadweight loss.

Selective Commodity Taxes

We now consider the DWL associated with selective commodity taxes. Figure 22.6a shows an ordinary (uncompensated) market demand curve for a commodity and two supply curves (S_1 and S_2). A unit tax (AC) shifts the supply curve from S_1 to S_2. The price to consumers rises from P_1 to P_C, while producers now receive P_S (the after-tax price). Consumption falls from Q_1 to Q_2. The tax paid (rectangle P_CACP_S) is a transfer from taxpayers to tax recipients. However, the losses of consumer surplus (AEB) and producer surplus (BEC) are a deadweight loss because there is no offsetting benefit.

To be exact about the DWL, we should distinguish between the income and substitution components of the demand curve for a commodity. Panel (b) of Figure 27.6 has the same elements as panel (a) but includes the compensated demand curve. For a price increase (fall) the compensated demand lies above (below) the ordinary demand curve. If a commodity tax increases the consumer price from P_1 to P_C, the effect on consumption can now be broken into substitution and income effects. Holding real income constant, the substitution (relative price) effect causes consumption to fall from Q_1 to Q_S. The income effect causes consumption

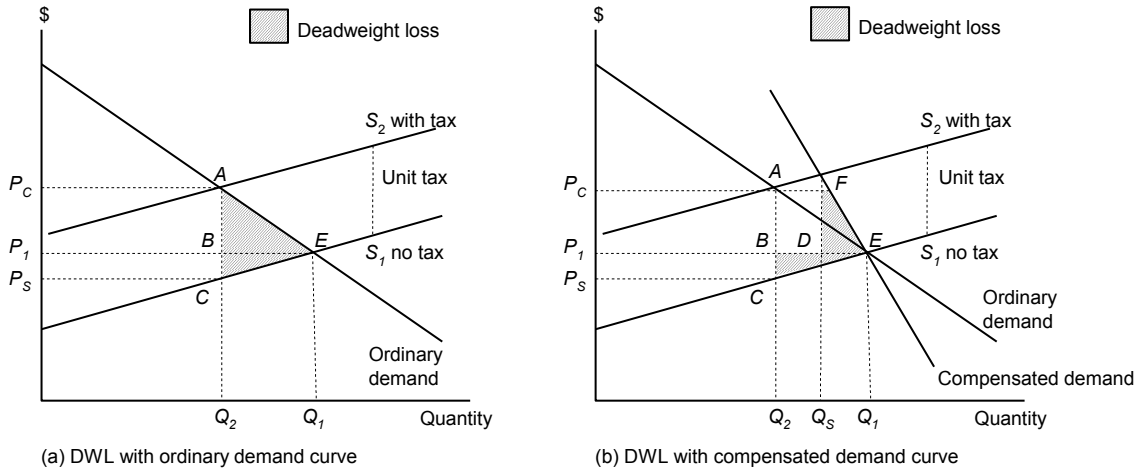


Figure 27.6 Deadweight loss of a selective commodity tax

to fall from Q_S to Q_2. The compensating variation measure of the DWL for consumers equals triangle *FED* instead of *AEB*. The loss of producer surplus remains area *BEC*.

However, as we saw in Chapter 6, for small price changes there is little income effect and differences between the uncompensated consumer surplus measure of DWL and the compensating (or equivalent) variation measure of loss are small.

Estimating deadweight loss using indifference curves

The concept of DWL can be illustrated using indifference curves and budget constraints. The DWL is the difference between the consumer's burden with a lump sum tax (the income effect) and the burden with a commodity tax (which includes the substitution effect as well as the income effect). The equivalent variation can be estimated in two ways. Holding after-tax utility constant, we can estimate the extra revenue that a lump sum tax would raise compared with a commodity tax. Alternatively, with tax revenue constant, we can estimate the extra expenditure required to make an individual as well off with a commodity tax as with a lump sum tax.

We consider the former approach first. Figure 27.7a overleaf shows Emma's initial budget constraint as line *GF*. Emma can spend her income on food or on other goods. The indifference curve I_1 shows the maximum utility that Emma can obtain given her budget constraint and she chooses position E_1. Now suppose that government levies a lump sum tax on Emma and her budget line shifts to G_1F_1. This is parallel to *GF* as the relative price of food and other goods has not changed. Emma now selects the E_2 combination of goods, which maximises her utility given her new budget constraint. The vertical distance between the budget constraints (AE_2) measures the tax revenue collected.

Now consider what happens if government taxes food. Emma's budget line rotates from *GF* to GF_2. Because the price of food has increased relative to other goods, the new budget curve is steeper than *GF*. Holding Emma's utility constant (along the I_2 indifference curve), Emma chooses the E_3 combination of goods. The substitution effect induces Emma to reduce her food consumption from Q_2 with the lump sum tax to Q_3. Now tax revenue equals BE_3. Holding Emma's welfare constant along I_2, there is a loss of tax revenue equal to $AE_2 - BE_3$, which is equal to *NM* on the vertical axis. *NM* is a measure of the DWL of the tax on food.

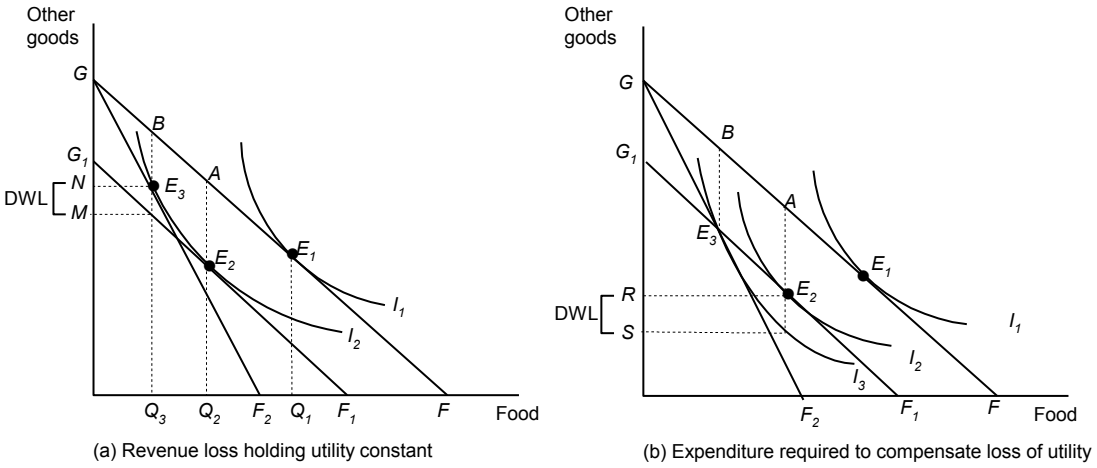


Figure 27.7 Estimating the deadweight loss of a commodity tax for an individual

Given that AE_2 is a lump sum tax, the fall in tax revenue occurs because Emma substitutes other goods for food.

Panel (b) of Figure 27.7 shows what happens when government has a revenue target AE_2. To reach the revenue target with a commodity tax, government would have to increase the tax on food. The GF_2 curve would be steeper than in panel (a). Holding tax revenue constant, Emma would be on indifference curve I_3 instead of I_2. At her new equilibrium point E_3, Emma pays the same amount of tax with a commodity tax as with a lump sum tax, but she has less utility. Holding revenue constant, the vertical difference between the utility curves I_2 and I_3 is a measure of the deadweight loss. An expenditure of RS on other goods would be required to compensate Emma for the loss of welfare.

Quantifying deadweight loss

Following our previous analyses, DWL rises with the elasticities of the compensated demand and supply curves. This is illustrated with two polar examples in Figure 27.8. Panel (a) shows a perfectly elastic supply curve and two compensated demand curves, representing elastic demand (D_E) and inelastic demand (D_I). Assume that a commodity tax shifts the supply curve from S_1 to S_2 and that the consumer price rises from P_1 to P_2. The DWL with D_I equals area BCD. With more elastic demand D_E, DWL increases to area ACE.

Panel (b) shows a perfectly elastic demand curve and two supply curves, an elastic curve (S_E) and an inelastic one (S_I). Assume that a commodity tax shifts the demand curve down from D_1 to D_2 and that the price to suppliers falls from P_1 to P_2. The DWL with S_I is given by area BCD. With more elastic supply S_E, DWL increases to area ACE. The loss rises with the elasticities of demand and supply because these elasticities drive the changes in quantity demanded or supplied in response to changes in price.

Evidently, if the compensated demand schedule for a commodity is linear and supply is perfectly elastic, the deadweight loss is given by:

$$DWL = 0.5\,(Q_1 - Q_2)\,(P_2 - P_1) = 0.5\,\Delta Q\,\Delta P \qquad (27.6)$$

where Q and P are the quantity and consumer price respectively, subscripts 1 and 2 refer to quantities and prices before and after tax respectively and Δ represents the change in quantity or price. To quantify the loss, estimates of the changes in price and quantity as a result of the tax rate are required.

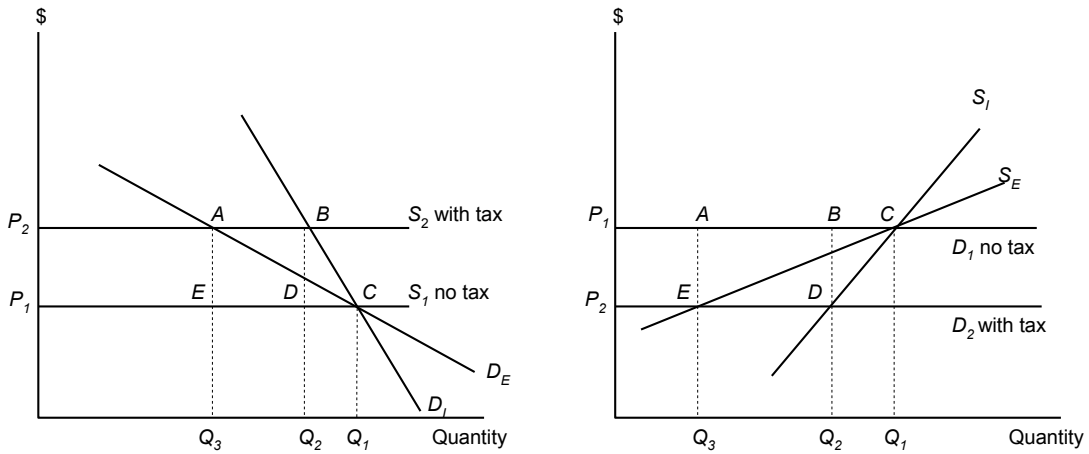


Figure 27.8 Deadweight loss and demand and supply elasticities

To understand the factors that drive DWL it is helpful to express Equation 27.6 in terms of the compensated elasticity of demand (η_{cd}) and the tax rate ($t = T/P_1$) where T is a unit tax. Following the procedure that allowed us to derive Equation 27.5 from Equation 27.1, the loss in this case can be estimated as:

$$DWL = 0.5\ (\Delta P/P_1)\ Q_1\eta_{cd}T = 0.5\ Q_1P_1\eta_{cd}t^2 \qquad (27.7)$$

where η_{cd} is the compensated elasticity of demand. Equation 22.7 indicates that DWL rises directly with the elasticity of demand and the square of the tax rate.[7]

Holding other factors constant, doubling a commodity tax results in a quadrupling of DWL. Suppose that a tax of \$1 per unit is imposed on a market in which 100 units are sold for \$10 per unit and that the price elasticity of demand is –1.0. The DWL would be approximately \$5 (equal to $\frac{1}{2}\Delta Q\Delta P = \frac{1}{2} \times 10 \times 1$). If a tax of \$2 per unit is imposed, DWL would be \$20 (equal to $\frac{1}{2} \times 20 \times 2$).[8]

This finding has significant implications. First, if compensated demand elasticities across commodities are similar, DWL is minimised by setting low tax rates across a broad range of goods rather than high tax rates across a narrow range of goods. If two commodities have similar compensated demand elasticities, to raise a given amount of revenue the DWL from taxation is minimised by an equal rate of tax on both commodities. The DWL with a tax of 10 per cent on both commodities will be one-half the loss with a 20 per cent tax on one commodity and no tax on the other. Second, the marginal DWL from raising an extra dollar of revenue exceeds the average DWL of taxation. The marginal loss is the policy relevant cost of raising taxes for public projects.

---

[7] When the supply curve is not perfectly elastic the excess burden also contains a loss of producer surplus. The deadweight loss is then equal to $0.5PQt^2/((1/\eta_d) + (1/\eta_s))$ where η_s is the elasticity of supply (Bishop, 1968). This expression equals Equation 22.7 when $\eta_s = \infty$.

[8] This is an approximation. If a demand curve has a constant elasticity of –1 all along the curve, the demand curve would be a (non-linear) rectangular hyperbola.

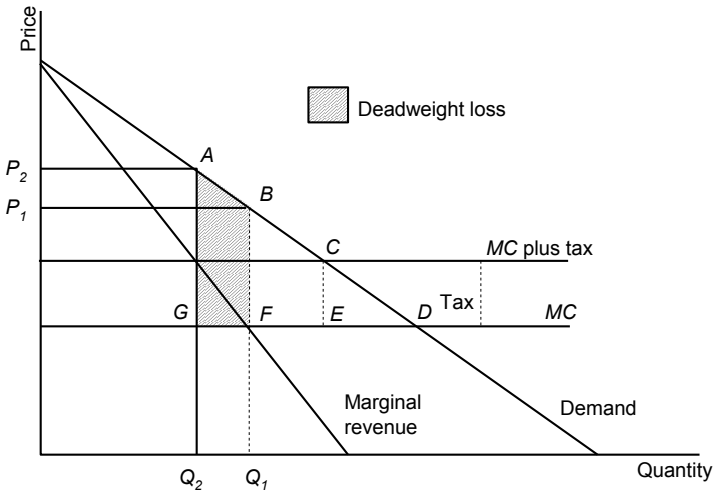


Figure 27.9 Deadweight loss of a tax on a monopoly

Other efficiency effects of selective commodity taxes

So far, we have assumed implicitly that government levies a partial consumption tax in a competitive market with no externalities and with no impacts on other markets. These assumptions are now relaxed.

A commodity tax with a monopoly. Figure 27.9 shows linear demand and marginal revenue curves in a monopoly market along with a constant marginal cost curve. For ease of illustration the ordinary demand curve is assumed to be similar to the compensated demand curve. A commodity tax *CE* is now imposed. If the market were competitive the DWL would equal area *CDE*. However, because a profit-maximising monopolist would equate marginal revenue to marginal cost, the tax would raise the price from P_1 to P_2 and consumption would fall from Q_1 to Q_2. The loss would be the much larger area *ABFG*, which is the difference between the price that consumers would be willing to pay for the commodity and the marginal cost of supply. Because price exceeds marginal cost in a monopoly market, a commodity tax has a greater DWL in this market than in a competitive market.

A tax on a commodity that causes a negative externality. As we saw in Chapter 13, a tax that reflects the marginal damage costs of a negative externality is called a corrective tax because it can correct the externality and produce a net benefit. Consider fuel consumption, which causes air pollution. Figure 27.10 shows the marginal private and social cost schedules (*MPC* and *MSC* respectively). The latter includes the cost of air pollution. With no tax on petrol, fuel consumption is Q_1, compared with the efficient quantity Q_2. The net social cost equals triangle *ABC*. If a corrective tax equal to *AC* is levied, the price rises to P_2 and consumption falls to Q_2. The corrective tax eliminates the *ABC* deadweight loss.

A tax on an untaxed substitute for a taxed commodity. This may also reduce the effect of an existing distortion. Suppose that initially beer is taxed but wine is not. A tax on wine may reduce the distortionary impact of the tax on beer. Figure 27.11 shows the initial compensated demand and supply for beer (D_1 and S_1). A tax on beer shifts supply to S_2, reduces output from Q_1 to Q_2, and causes a DWL of area *ACE*. After a tax is imposed on wine, the demand for beer rises from D_1 to D_2 and the quantity consumed increases from Q_2 to Q_3. The deadweight loss has fallen to

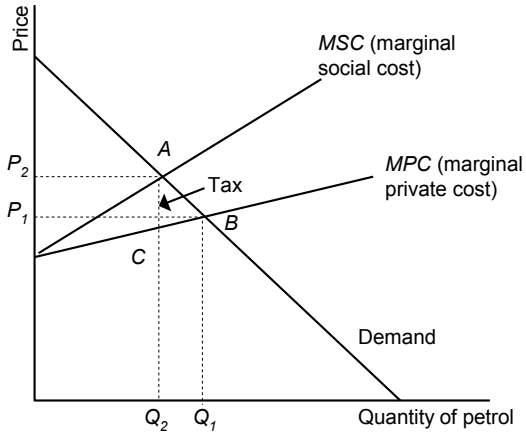


Figure 27.10 Benefit of a corrective tax

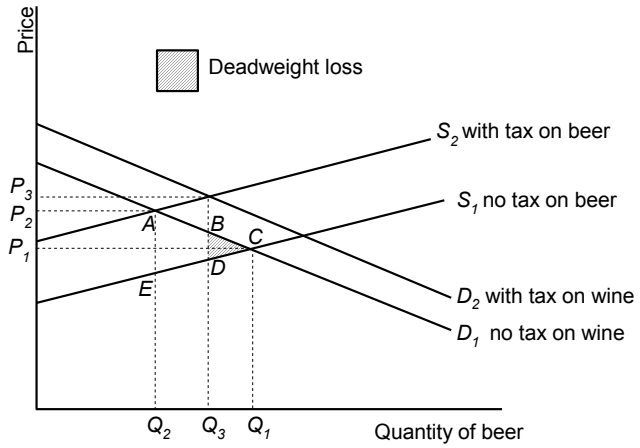


Figure 27.11 Deadweight loss in a second-best world

area *BCD*. As it happens, in this example consumption is still below the efficient level of Q_1. But the tax on wine has reduced the DWL of the initial tax on beer. This reduction in DWL from taxation of beer may partly or even fully offset the DWL associated with the introduction of a tax on wine.

General Taxation of Consumption

A general tax on consumption can take several forms. It can be a general expenditure tax (i.e. a tax on income less savings), a general tax on retail sales or a general value-added tax that excludes capital expenditures. Being equivalent taxes on a similar tax base, the tax rate would be the same in each case. Likewise, the DWL of each kind of general consumption tax would be similar unless there were some fiscal illusions. We also saw in the last chapter that a general tax on consumption is equivalent to a tax on labour. We would expect it therefore to have similar impacts—it would affect the supply of labour but not the supply of savings.

Consider first the impact of a general consumption tax on labour supply. Suppose that Emma earns $20 an hour, or $40 000 a year if she works full time, and that the average income tax rate is 25 per cent. The tax would raise $10 000 revenue. Replacing the income tax with a general consumption tax, to raise equivalent tax revenue, government could levy a 33 per cent tax on consumption expenditure ($30 000 × 0.33 = $10 000). In both cases Emma would receive $15 of goods in return for an hour's work. For a given tax revenue the after-tax rate of substitution between work and leisure is the same with a consumption (expenditure) tax as with an income tax. In both cases, Emma has to decide whether to forgo an hour of leisure for $15 worth of goods.

It follows that, for any given tax revenue, a general expenditure tax has the same effect on labour supply as a tax on labour income. This assumes no fiscal illusion—that is, Emma understands the real (after-tax) purchasing power of an hour's work. If people perceive a tax on income to be a greater burden, it could be a greater deterrent to work than an equivalent tax on expenditure.

Now consider the effect of a general consumption tax on savings. Ignoring inheritances and bequests, suppose that Emma has only her own income and spends all of it in her lifetime. And suppose that Emma earns W_1 in year 1 and W_2 in year 2. In scenario A, she consumes all income when it is earned. In scenario B, she saves S_1 in year 1 and consumes these savings plus interest in year 2. With a tax on consumption the present value (PV) of Emma's consumption is the same in the two scenarios (C_A and C_B) despite the different consumption patterns.

Allow that consumption is taxed when it occurs at rate (t). Then,

$$PV(C_A) = W_1(1-t) + \frac{W_2(1-t)}{(1+r)} \tag{27.8}$$

$$PV(C_B) = (W_1 - S_1)(1-t) + \frac{W_2(1-t)}{(1+r)} + \frac{S_1(1+r)(1-t)}{(1+r)} \tag{27.9}$$

where r is the rate of interest. The present value of consumption is the same in Equations 27.8 and 27.9. Likewise, the present value of the tax paid is independent of the expenditure pattern.

With a consumption tax, savings are taxed only once—when they are spent. Thus, a tax on consumption, like a tax on wages, does not distort an individual's pattern of consumption (and savings) over time.

Incomplete general consumption taxes. Few general consumption taxes are comprehensive. In the European Union many goods are exempt from VAT. In Australia, the GST covers 57 per cent of expenditures. This differential tax treatment distorts the pattern of consumption.

Suppose that an economy consists of two goods, manufactures and services, and that manufactures are taxed whereas services are not. Panel (a) in Figure 27.12 shows the market for manufactures. The tax on manufactures shifts the supply curve for manufactures from S_1 to S_2. There is a DWL of triangle ABC, similar to that for a selective tax on one commodity. Panel (b) shows that the tax on manufactures raises the demand for services from D_1 to D_2. The quantity and price of services rises. However, there is no additional DWL in the market for services. The increase in the price of services is a transfer payment from consumers to producers. The full cost imposed on consumers who shift from manufactures to services is reflected in triangle ABC in panel (a).

Estimates of the DWL of consumption taxes are more complex when many different tax rates on commodities exist. It might be expected that a tax system with several commodity tax rates would distort consumption patterns more and cause a higher aggregate DWL than would a uniform tax on consumption. However, as we show in Chapter 28, an efficient commodity tax system may include differential tax rates that allow for differences in demand elasticities.

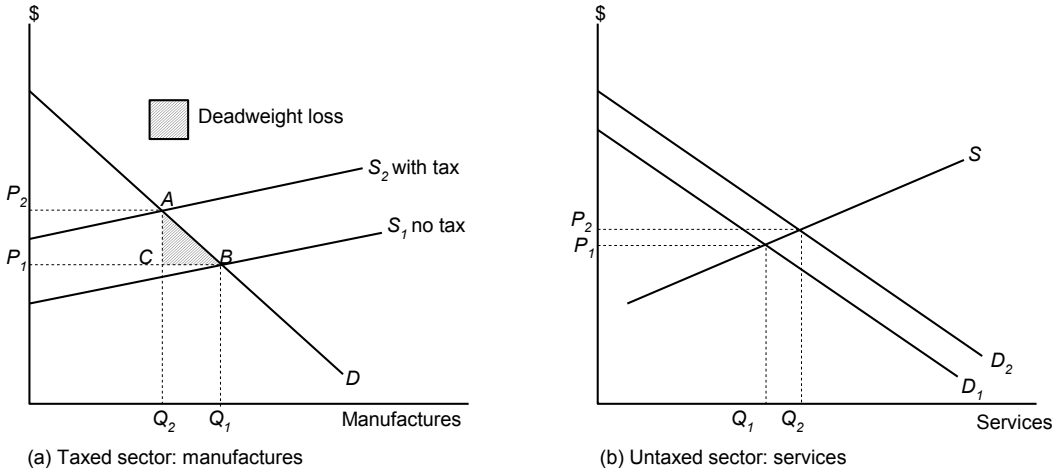


Figure 27.12 Deadweight loss with an incomplete general consumption tax

Moreover, the DWL of a commodity tax on one or a few goods cannot be considered in isolation from other commodity taxes. The overall deadweight loss due to the tax system is not necessarily the sum of the individual distortions.

It may be reasonable to assess the impact of a particular tax holding constant impacts in other markets. However, to estimate the overall distortion of a consumption tax system the net impacts in all markets need to be considered simultaneously.

Actual Deadweight Losses of Taxation

Several studies have attempted to estimate total, average and marginal DWLs of major tax forms and indeed a total DWL for all taxes across an economy. For example, for Australia, Campbell and Bond (1997) estimated that the marginal DWL of an extra dollar of income tax revenue on labour incomes was between 19 and 24 cents in the dollar. Freebairn (1998) concluded that the DWL of the income tax on labour supply was about 2–3 per cent of GDP and that the total DWL of taxes on savings and investment could be 1–2 per cent of GDP. In the United States, Aaron and Gale (1996) estimated that the efficiency gains of replacing the hybrid tax system with a comprehensive income or expenditure system would be equivalent to between 2 and 5 per cent of GDP.

Table 27.2 overleaf presents detailed estimates of average and marginal DWLs for Australia taxes, prepared by KPMG Econtech (2010) using a computable general equilibrium (CGE) model for the Henry Tax Review (2010). The main findings are:

- Several taxes have a marginal DWL of less than 10 per cent, including the petroleum resource rent tax, taxes on land and the GST.
- The estimated marginal DWL of personal income tax is 23 per cent, of corporate income tax is 40 per cent and of payroll tax is 41 per cent.
- The marginal DWL of royalty taxes and taxes on insurance and gambling is over 60 per cent.

Key determinants of the estimated DWLs are the mobility and narrowness of the tax base. When a tax base is mobile, the taxed item can move. When a tax base is narrow, users of the taxed item can substitute untaxed substitutes. Thus DWL rises with mobility and narrowness.

Table 27.2 Estimated deadweight losses of Australian taxes (cents per $ of revenue)

Tax	*Marginal DWL*[a]	*Average DWL*
Tobacco excise (allowing for externalities)	-8	-23
Petroleum resource rent tax	0	0
Municipal rates	2	1
GST	8	6
Land taxes	8	6
Alcohol excise and WET	9	7
Fuel taxes	15	10
Stamp duties exc. real property	18	18
Luxury car tax	20	9
Labour income tax	24	16
Conveyancing stamp duties	34	31
Motor vehicle registration	37	32
Motor vehicle stamp duties	38	38
Corporate income tax	40	23
Payroll tax	41	22
Insurance taxes	67	47
Royalties and crude oil excise	70	50
Gambling taxes	92	54

(a) Actually for 5 per cent increases in each tax, so the estimates are not strictly marginal DWLs.

Source: KPMG Econtech (2010).

In interpreting these results, some caveats are necessary. First, as KPMG Econtech details, numerous assumptions underlie these estimates, including estimates of critical demand and supply elasticities. The zero estimate for the petroleum resource rent tax assumes only excess profits are taxed. The DWL estimates for land tax are underestimated because the modelling did not account for all the variations in these taxes. Other simplified assumptions about actual tax structures were unavoidably adopted. The results of such economic modelling depend on these input assumptions.

Second, the aggregate DWL of the tax system cannot be calculated simply by adding up the separate DWLs of all taxes. One tax or set of taxes may offset the effects of another tax. A CGE model of the economy is required that can compare the economy-wide output and DWLs under alternative tax systems. Such a model should include prices and demand and supply functions for each major sector as well as measures of income and output. KPMG (2010) does not provide an estimate of the DWL of all taxes. Using a CGE model, Diewart and Lawrence (1994) estimated that the DWL of taxation in New Zealand amounted to 5.5 per cent of GDP. Freebairn (1998) concluded, without using a CGE model, that the total DWL of taxation in Australia could be as high as 6 per cent of GDP and that compliance costs could be an additional 4 per cent of GDP.

Finally, estimates of the DWL of a taxation system relative to a costless lump sum system of taxation are of limited value. As we have observed, an individualised lump sum system of taxation based on individual capabilities is not feasible. A general (non-individualised) system of lump sum taxes does not satisfy the fundamental equity principle that tax should be paid in accordance with ability to pay. In practice, all tax systems are based on, and distort, the activity of firms or individuals. Thus, all tax systems have some DWL. To be policy relevant, the estimated cost of a tax system should be compared with the lowest DWL of a *feasible* tax system alternative. The evidence presented above suggests that the gains from an improved tax system in Australia could be substantial, but they are lower than the total DWL of the tax system.

Summary

- Most taxes change behaviour and cause resources to be reallocated. Unless a tax corrects for market failure, it results in an inefficient use of resources. This inefficiency is described as a deadweight loss or excess burden.
- Taxation of labour income influences workforce participation, supply of labour hours, choice of occupation and work effort. All such changes have deadweight losses.
- Most attention has been paid to the impact of taxation on labour hours. The deadweight loss rises with the compensated supply elasticity. This is typically quite low for primary earners but can be around 1.0 for secondary earners.
- Taxation of capital income encourages individuals to substitute present for future consumption, which has a deadweight loss. However, savings may not fall if the income effect of lower after-tax returns on capital encourages people to compensate by increasing savings.
- A tax on a commodity causes a deadweight loss when a consumer substitutes other goods for the taxed good. The loss rises with the elasticities of demand and supply and with the square of the tax rate.
- A general tax on consumption effectively taxes labour income and discourages market output. However, a general tax on consumption is rare.
- Most so-called general consumption taxes are selective taxes and have a deadweight loss as consumers substitute non-taxed or lightly taxed goods for more heavily taxed goods.
- The deadweight losses of Australian taxes vary greatly from around zero to over 50 per cent. Overall, the taxation of income and consumption in Australia may have an aggregate deadweight loss in the order of 5–6 per cent of GDP.
- The potential gains from an improved tax system are therefore substantial. However, it will never be possible to achieve a tax system free of any deadweight loss.

Questions

1. Explain the difference between the income and substitution effects of commodity taxation. Why is the deadweight loss of taxation associated only with the substitution effect?
2. What factors determine the size of the deadweight loss of taxation? Why does the marginal deadweight loss from raising one more dollar of tax revenue exceed the average deadweight loss?
3. What are the practical implications of your answers to (2) for determining income and consumption taxes?
4. (From Rosen and Gayer.) In the formula for deadweight loss given in Equation 22.5, the tax rate is less than one. Thus squaring the tax rate reduces the deadweight loss. Using t^2 in the formula rather than t makes the tax rate less important. Discuss.
5. What efficiency implications arise from the non-taxation of goods produced outside of the market?
6. Suppose that Craig is willing to work 40 hours a week at \$40 per hour. Craig faces a tax rate of 40 per cent. His ordinary supply elasticity for marginal work hours is 0.5 and his compensated supply elasticity for marginal work hours is 0.7. Determine:
   i. The hours Craig will work given the tax rate and the amount of tax Craig will pay.
   ii. The deadweight loss of each dollar of marginal tax collected.
7. Suppose that a good is fixed in supply at $Q^S = 12$, and that demand for the good is given by $Q^D = 50 - 6P$. The government imposes a tax of \$2 per unit on the consumer. Determine:
   i. The pre-tax price paid by consumers and the amount received by producers.
   ii. The post-tax price paid by consumers and the amount received by producers, and the share of the tax burden borne by producers and consumers.
   iii. The amount of government revenue raised by the tax.
   iv. The deadweight loss of the tax.
8. The government imposes a tax of \$2 per unit on a commodity Z. The supply of Z is perfectly elastic, and the compensated price elasticity of demand is –0.2. Currently 100 units of Z are sold at a price of \$4 per unit. Determine the deadweight loss that would arise as a result of the tax on Z.
9. Suppose that the community demand for beer is given by $Q^D = 50 - 2P$, where Q is thousands of litres of beer consumed each week. The supply for

beer is given by $Q^S = -50 + 10P$. The government imposes an excise tax of $1.0 per litre of beer consumed payable by beer retailers (approximately the current Australian rate on small barrels of beer). Determine:

i. The pre-tax and post-tax equilibrium price and quantity of beer.

ii. The tax burden borne by the sellers and consumers.

iii. The excess burden of the tax on the consumption of beer.

10. Suppose that, as a population policy, government taxes families for each child in excess of two in the family. Consequently, no family has more than two children and government collects no revenue from the policy. Does this mean that the policy has no deadweight cost?

11. Does a tax on income from capital create a deadweight loss when there is no change in the amount of saving?

12. It is often commented that the 'whole is not always equal to the sum of its parts'. What implication does this statement have for analysing the total deadweight loss of (i) commodity taxation and (ii) all taxes?

13. Can subsidies, such a subsidy on private health care insurance, be regarded as negative taxes and therefore be expected to have similar deadweight losses to positive taxes?

14. Government is considering raising taxes by 10 per cent on alcohol consumption, all food, mobile phones, laptop devices, land values and high income earners. Which of these taxes is likely to have the highest or least deadweight loss and why?

Further Reading

Blomquist, S. and Selin, H. (2010) 'Hourly wage rates and taxable labor income responsiveness to changes in marginal tax rates', *Journal of Public Economics*, 94, 878–889.

Creedy, J. (2004) 'The excess burden of taxation', *Australian Economic Review*, 37, 454–464.

Dandie, S. and Mercante, J. (2007) *Australian Labour Supply Elasticities: Comparison and Critical Review*, Working Paper 2007–04, Treasury, Canberra.

Diewert, W.E., Lawrence, D.A. and Thompson, F. (1998) 'The marginal costs of taxation and regulation', pp. 135–171 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Feldstein, M.S., (2008) *Effects of Taxes on Economic Behaviour*, Working Paper 13745, National Bureau of Economic Research, New York.

Gruber, J. (2016*) Public Finance and Public Policy*, 5th ed., Chapter 21, Worth Publishers, New York.

Hausman, J.A. (1985) 'Taxes and the labour supply', pp. 213–264 in *Handbook of Public Economics*, Vol. 1, A.J. Auerbach and M. Feldstein (eds), North Holland, Amsterdam.

Johansson, H., Brys, A. and Vartia (2008) *Tax and Economic Output*, Economics Department Working Paper, No. 620, OECD, Paris.

Keene, M.P. (2011) 'Labour supply and taxes', *Journal of Economic Literature*, 49, 961–1075.

KPMG Econtech (2010) *CGE Analysis of the Current Australian Tax System*, Final Report to the Henry Tax Review.

Rosen, H. and Gayer, T., 2014, *Public Finance*, 10th ed., Chapter 15, McGraw-Hill, New York.

Chapter

28

Optimal Taxation and Tax Reform

There is nothing more difficult to carry out, nor more doubtful of success, nor more dangerous to handle, than to initiate a new order of things. For the reformer has enemies in all who profit by the old order, and only lukewarm defenders in all those who would profit from the new order.

Machiavelli, *The Prince*

Optimal Taxation: Core Concepts ♦ Optimal Taxation of Income ♦ Optimal Commodity Taxation ♦ Taxes on Producer Goods and International Trade ♦ An Optimal Tax System ♦ Tax Reform

In this chapter we turn to policy questions. What is the best system of tax? And, how can an existing tax system be improved? Answers to these questions require not only an understanding of the effects of alternative taxes but also normative judgements. Traditionally economists evaluated tax systems using various equity and efficiency criteria. More recently economists have sought to determine more formally the tax system that maximises social welfare using a social welfare function. The former approach is more pragmatic; the latter more rigorous. In our discussion below we draw on both approaches.

The first section discusses optimal income taxation concepts. We then discuss optimal taxation of income. Importantly, the term 'optimal income taxation' is used in the economics literature to refer to a set of tax rates **and** income transfers (or negative taxes) across individuals that maximises social welfare. Thus, it is concerned not just with taxation but with optimal income redistribution. The third section discusses an optimal commodity tax system which is a preferred set of tax rates across commodities. We then briefly examine the role of taxes on producer goods (also known as intermediate goods) and international trade. In the fifth section we discuss the optimal overall system of taxes. The last section discusses tax reform—how we can move from a flawed system towards a better one.

Optimal Taxation: Core Concepts

Traditional public finance asked: what is the best way to raise a revenue target? Three strategies were proposed. One was the principle of **equal absolute sacrifice—**each person should make the same sacrifice in utility (not money) terms. Suppose that government needed to raise \$100 from two individuals where A is rich and B is poor and that the loss of three dollars to A is equivalent in utility to the loss of one dollar to B. The principle of equal absolute sacrifice implies that A should pay \$75 in tax and that B should pay \$25.

Although this may appear to be fair, it implies an increase in relative inequality. Suppose that A has 10 units of utility and B has 5 units and that each loses one unit of utility in paying

the tax. After tax, A has 9 units of utility and B has only 4 units. The less well-off person (B) has become relatively worse off. This change in relative wellbeing can be avoided by the principle of **equal proportional sacrifice**. In this case, each would be taxed such that they sacrificed an equal proportion of their total utility.

However, so long as a marginal dollar has more value to one individual than to another, neither absolute or equal proportional sacrifice is welfare maximising. Social welfare is maximised by minimising the total loss of welfare (or utility). This is achieved by the principle of **equal marginal sacrifice**. So long as a marginal dollar has less utility to A than to B, the marginal tax should be levied on A. This continues until the marginal tax dollar has the same utility to A and B.

As we saw in Chapter 7, under certain assumptions equal marginal sacrifice implies equality of after-tax incomes. The assumptions are that social welfare is the sum of individual welfares, individuals have similar utility functions with respect to income, these utility functions exhibit diminishing marginal utility of income and total income is constant. For any revenue target, welfare would then be maximised by levying a 100 per cent marginal tax on the highest income individuals until the target is obtained.

A critical assumption underlying this conclusion, and the case for equal absolute sacrifice, is that the tax rate does not change labour supply. If taxation reduces hours worked there is less total income to allocate. If this assumption does not hold, optimal tax system design should allow for the effect of taxation on labour supply and tax revenues.

Optimal tax design becomes more complex if the objective is broadened from achieving a revenue target at least cost to finding the optimal combination of income taxes and grants to individuals or households. The former objective is essentially concerned with how to fund public goods. The latter objective includes redistribution aims. The optimal tax problem is now how to choose both the tax rate and the level of grants, subject to the labour supply constraint, that maximise social welfare. A social welfare function is needed to determine the welfare implications of dollar transfers between individuals.

Figure 28.1 illustrates the problem. This shows how the grant (G) per person can vary with tax rates with a fixed (perfectly inelastic) labour supply and with a more typical labour supply that falls as tax increases. A minimum tax rate (t_o) is required to finance provision of general government goods. At t_o tax rate, G is zero. As the tax rate increases tax revenue rises and G rises. If the labour supply were perfectly inelastic the grant would rise linearly with the tax

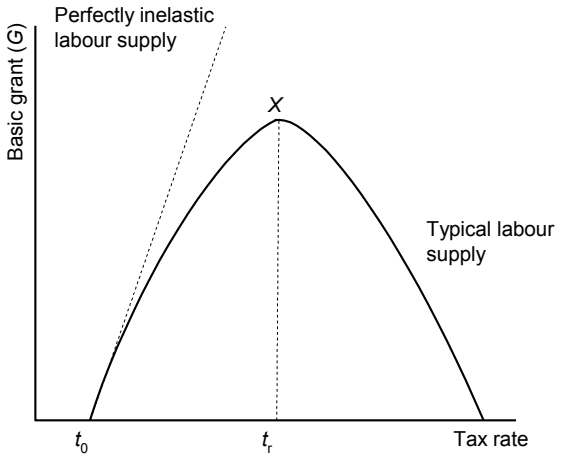


Figure 28.1 Tax rates and redistribution

rate. However, if labour supply falls as the tax rate increases, revenue increases less than linearly with the tax rate.

When tax rates are greater than t_r, revenue falls. Point X is a Rawlsian solution, where the income of the least well-off person (the basic grant) is maximised. However, at points just to the left of X the tax rate is increasing with little increase in G and the marginal burden to taxpayers may be viewed as exceeding the marginal benefit to grant recipients. The aim of optimal tax studies is to find the optimal tax rate between t_o and t_r (including possibly the end points).

To understand the issues, we need first to understand the concepts of linear and non-linear tax structures. These structures are illustrated in Figures 28.2 and 28.3. In both figures individuals receive a grant until their income reaches $\bar{Y}$. But they also pay effective tax with incomes below $\bar{Y}$ as the grant is withdrawn.

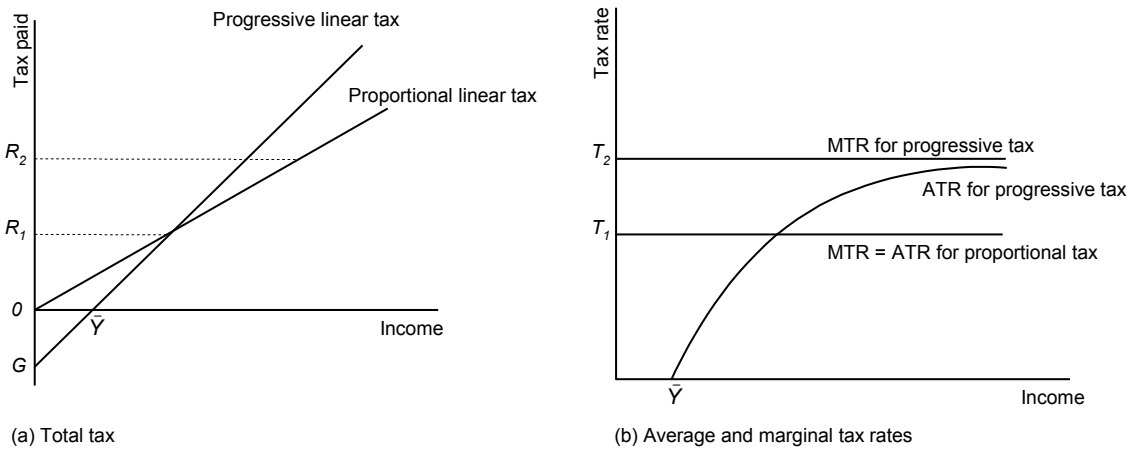


Figure 28.2 Proportional and progressive linear taxes

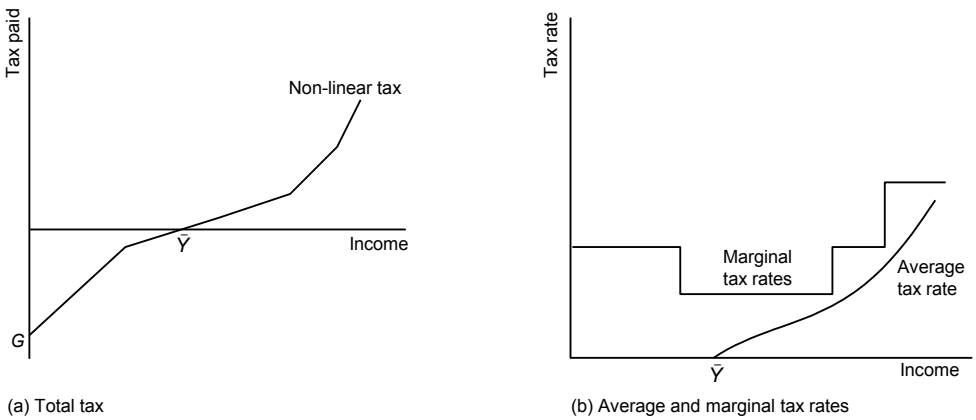


Figure 28.3 Typical progressive non-linear tax structure

Linear and non-linear tax structures

A tax structure is **linear** if it has a constant marginal tax rate (MTR) at all levels of income. As previously notes, this tax structure is also called a **flat tax** system.

As shown in Figure 23.2a, a linear tax structure can be proportional or progressive. In a **proportional linear tax system** tax paid is proportional to income. This means that the MTR equals the average tax rate (ATR) at all levels of income.

Linear tax
Has a constant marginal tax rate at all levels of income. It may be proportional or progressive

In a **progressive linear tax system** MTR is again constant but ATR increases with income. This can be achieved by combining a uniform lump sum grant (G) to all individuals along with a proportional income tax. The grant that individuals receive in this system is sometimes described as a **negative income tax**.

In Figure 23.2a, the MTR is the slope of the tax schedules. The steeper the slope, the higher is the MTR. Panel (b) shows the MTRs and ATRs for the progressive and proportional tax.

On the other hand, a **non-linear tax structure** is defined by varying MTRs. Figure 23.3a shows a non-linear tax schedule. Figure 23.3b shows the equivalent ATRs and MTRs. The effective MTR for low–medium earners is often high as they lose benefits as well as paying taxes as their income increases.

Negative income tax system
A tax system that includes income grants as well as taxes

All tax schedules have two basic features: efficiency and equity. Efficiency is a function of marginal tax rates. Individual decisions, and hence resource allocation, depend on the relevant MTR. On the other hand, equity depends on average tax rates. The tax that a person pays is a function of her income and ATR. Therefore, an optimal income tax system must aim to combine low MTRs that are efficient along with increasing ATRs that are equitable.

Non-linear tax structure
Contains varying marginal tax rates

Now consider the implications of proportional and progressive linear tax schedules. As shown in Figure 28.2, for all levels of income MTR is higher with a progressive tax than with a proportional tax. It follows that, for any amount of tax revenue and any linear tax structure, the deadweight loss (DWL) will generally be greater with a progressive tax than with a proportional tax. On the other hand, ATR rises with income with a progressive tax but not with a proportional tax.

In practice, most tax schedules are non-linear with varying MTRs. This is likely to be equitable but may result in significant efficiency costs. In an optimal tax structure, the ATR for high-income earners should be maximised to achieve equity, but the MTR minimised to achieve efficiency. The task of optimal tax studies is to find the best compromise. As we will see this has led some analysts to conclude that a progressive linear system could be preferred.

Optimal Taxation of Income

In this section we discuss separately the optimal taxation of income from labour and from capital as they raise different issues, albeit that equity may be concerned with total income.

Optimal taxation of labour income

Most formal economic analysis of optimal income tax focuses on labour income. We follow this approach and discuss optimal linear and non-linear tax systems respectively. With a **linear tax system,** the tax revenue available to pay for public goods (R) after paying the basic grant (G) to all adults (N) equals:

$$R = tY - NG \tag{28.1}$$

where Y is aggregate taxable income. If G is zero, this is a proportional tax system. If G is positive it is a progressive linear tax system. If grants are paid, this is known as a negative income tax system. The optimal tax problem is to determine the values of t and G that maximise social welfare, subject to various constraints.

Formally the aim is to maximise social welfare subject to four constraints shown in Equations 28.3 to 28.6.

$W = W(u_1 \ldots u_n)$	the social welfare function	(28.2)
$R + NG = \Sigma_i t w_i h_i$	government revenue requirement	(28.3)
$u_i = u(y_i, l_i)$	individual utility functions	(28.4)
$y_i = w_i h_i (1 - t) + G$	individual income ($y_i = c_i$)	(28.5)
$l_i = T - h_i$	time constraint	(28.6)

where y is individual income, c is consumption, l is leisure, h is hours of work, T is time, there are $i = 1 \ldots n$ members of the community and the other variables are as before. The solution depends upon the specifications of the social welfare function (SWF) and individual utility functions. The other three equations are identities. Critically, labour supply is determined by maximising individual utility (Equation 28.4) subject to Equations 28.5 and 28.6. In this model, labour is the only source of income and there are no savings or income from savings.

In a classic study of an optimal linear income tax, Stern (1976) adopted the SWF:

$$W = \frac{1}{(1-\epsilon)} \int_0^x u(c, l)^{1-\epsilon} f(w) \mathrm{d}w \tag{28.7}$$

where ϵ is an equity parameter and the other variables as above. This SWF is similar to the general SWF in Chapter 7 (Equation 7.6). In Stern's model, individuals have a common constant elasticity of substitution utility function:[1]

$$u(c, l) = \left(\alpha (T - h)^{-\mu} + (1 - \alpha) c^{-\mu} \right)^{-\frac{1}{\mu}} \tag{28.8}$$

where $\mu = (s\text{-}1)/s$ and s is the elasticity of substitution, α is the share parameter, and their budget constraint is $c = wh(1–t) + G$, which is similar to Equation 28.5. Stern calculated the results for various assumptions about government revenue requirements, labour supply elasticities and equity parameters. For example, given a general revenue requirement of 20 per cent of GDP, a compensated labour supply elasticity of 0.4 and $\epsilon = 2.0$, Stern estimated an optimal constant MTR of 54 per cent. But even with this high MTR, the estimated minimum grant income is only one-third of average income. The results are very sensitive to the equity parameter. With the same set of assumptions but with no specific equity allowance in addition to allowances for the falling marginal utility of income (i.e. $\epsilon = 0.0$), the optimal MTR falls to 25 per cent.

Atkinson (1995) also estimated an optimal linear tax. He assumed a SWF where the social value of an additional unit of income to an individual with a wage rate w is proportional to $w^{-\epsilon}$, where ϵ is a measure of equity.[2] If ϵ is 1.0 the social value of marginal income is inversely proportional to a person's income. The higher the value of ϵ the greater the weight attached to a dollar going to a low-income person. Atkinson showed that, with this formulation of the redistribution objective, the optimum tax rate (t^*) is given by:

$$\frac{t^*}{1 - t^*} = \underbrace{\frac{1}{\eta_{cs}}}_{\text{efficiency}} \underbrace{\left[1 - \left(1 + v^2 \right)^{-\varepsilon(1+\eta_{cs})} \right]}_{\text{equity}} \tag{28.9}$$

[1] A utility function with a constant elasticity of substitution has the property that the ratio between the proportional changes in relative quantities (of consumption and leisure) and the proportional changes in their relative prices is always the same.

[2] Atkinson used the expression w^{-y}, but for consistency of expression we use ϵ for the equity parameter here.

where η_{cs} is again the compensated labour supply elasticity, v is the coefficient of variation in wage rates and ϵ is the equity parameter.[3] This formula indicates that efficiency is related solely to the elasticity of the compensated labour supply. The larger this elasticity the greater is the DWL due to taxation and the lower the optimum tax rate.

In this formulation, the equity component depends on the values of v, ϵ and η_{cs}. The greater the inequality of incomes and the higher the value of ϵ, the higher is the optimum tax rate. Assuming a conservative equity value of 0.5 for ϵ, a coefficient of variation of wages of about 0.4 and a labour supply elasticity of 0.3, the optimal MTR would be 21 per cent. However, this makes no allowance for the government's general revenue-raising requirements. This requirement, say 20 per cent of GDP, cannot simply be added to the optimum tax rate that results from Equation 28.9.

All such studies seek to maximise social welfare subject to a government budget constraint. Most studies test the sensitivity of the optimal constant MTR to variations in government's general revenue requirements, the labour supply function and the equity parameter. The range of results is wide. The studies find that the optimal tax rate varies from as low as 17–18 per cent (with no general revenue requirement) to 70–80 per cent when the labour supply elasticity is very low and the equity parameter is high. However, common findings are that the optimal MTR is higher for:

- a high government general revenue requirement;
- greater inequality in pre-tax wages;
- high values of ϵ, the degree of concern for inequality; and
- low values of the compensated elasticity of labour supply.

Optimal income tax with a non-linear taxation system. Many studies have examined optimal tax rates with a non-linear tax structure following a similar approach, that is, by maximising social welfare as embodied in a SWF subject to various constraints. However, the studies have produced conflicting results. Some find that the optimal MTR should be an inverted U-shape with low MTRs on both low and high incomes. The logic here is that, for equity reasons, individuals on very low incomes should lose very little of the marginal dollar received and for efficiency reasons individuals with high marginal productivity should not be discouraged from working, while taxes on medium levels of income have less impact on marginal labour supply. Others have found that a U-shaped schedule, with high effective MTRs (allowing for loss of grants) on low and high incomes, is optimal. The logic here is that grants should be highly targeted and withdrawn rapidly as income increases and that equity requires high MTRs on high-income individuals.

In an early study of a non-linear tax structure, Mirrlees (1971) found that the MTR for the person with the highest income should be zero. At the margin, she will be indifferent between work and leisure, so no extra revenue is gained by raising her MTR above zero and there is no reason to distort her labour supply decision. This conclusion that the MTR at the top income should be zero has little practical value because it applies only to the highest income. More usefully, Mirrlees calculated complete optimal income tax schedules for various specific cases. These results indicated a slight decline in the MTR over most of the income distribution. However, the decline is so slight that the optimal income schedule approximated a linear tax schedule (i.e. a constant MTR).

Some other studies support Mirrlees' general findings. For example, Tuomala (1990) found that rapidly declining MTRs are desirable. Slemrod *et al*. (1994) examined a tax system with

[3] For this case study Atkinson assumed a constant elasticity of labour supply and that the variation of wage rates was lognormal. Wages were the only source of income.

two MTRs and concluded that individuals in the higher income bracket should face a lower MTR than those in the lower income bracket. The finding is based on the estimated labour supply elasticity of high-income individuals such that a lower MTR increases both labour supply and tax revenue (and so increases benefits to low-income households). Although MTRs fall in their model, ATRs rise with income so that their tax structure is progressive. In general, if an equal amount of tax can be raised with a lower MTR, the lower MTR is preferred because it minimises the DWL. High-income individuals achieve a higher level of utility with no loss of revenue to the low-income group.

On the other hand, studies by Diamond (1998) and Saez (2001) found that a U-shaped tax schedule could be optimal and preferred to a linear tax schedule because of its greater ability to redistribute income. This ensures a targeted distribution of grants. The results are sensitive to assumptions about the market distribution of income, labour supply elasticities (assumed to be low in the benefit phase out income region) and the government's redistributive tastes.

Conclusion. With perfect information and no tax administration costs or avoidance activity, a non-linear (labour) income tax would dominate a linear income tax as a variable MTR would be more efficient and more equitable than a single MTR. However, the benefits of a non-linear system depend on the distribution of labour income, labour supply elasticity and the importance attached to equity relative to deadweight losses.

In a review article, Heady (1996) concluded that a linear tax schedule is generally a reasonable approximation of an optimal tax schedule, with non-linear tax schedules providing only small welfare gains. A linear income tax has a high administrative convenience, especially the ability to withhold taxes from multiple sources. It can also achieve significant redistribution. If combined with personal tax allowances it can also be progressive over most of the income distribution.

Finally, we should note that the analyses above deal only with labour income. Personal income tax is more complicated when it includes taxation of income from capital.

Optimal taxation of income from capital

Taxation of income from capital (such as rents, term deposits and shares) is controversial. Some economists (see Mankiw *et al*. 2009) contend that income from capital should not be taxed, especially income from new capital. There are three main reasons for this. First, a tax on capital is a tax on capital equipment which is an input to the production of future output. This raises interest costs, reduces investment and distorts production away from capital goods. This is contrary to the Diamond-Mirrlees result discussed later in this chapter that intermediate goods should not be taxed. Second, as we saw in Chapter 27, a tax on income from capital discriminates against deferred consumption and distorts inter-temporal consumption choices. Third, as we will see in Chapter 31, in a small open economy country some of the tax on foreign capital is borne mainly by workers in the local economy.

However, there are strong counter arguments, many of which are made by Auerbach (2006). First, taxation of income from existing (old) capital and taxation of pure economic profits has no deadweight loss. Second, taxation of income from capital, including corporation tax, provides for up to a quarter of all tax revenue in many countries. If this tax revenue were forgone, most of the revenue would have to be obtained by increasing some other distortionary tax, such as taxation of labour income. Third, taxation of income from capital is usually equitable because the income accrues to higher income persons, though not always as it also accrues to elderly non-income earning individuals.

On the whole, the arguments for taxation of income from capital seem to be stronger than the arguments against it. However, there may be a case for exempting from taxation normal returns to new capital and for reducing taxation on international capital. As we will see, recent major tax reviews in the UK and Australia, have recommended reforms along these lines.

Optimal Commodity Taxation

In this review of commodity taxation, we start by describing an efficient commodity tax system. This is the commodity tax system that would minimise the DWL of commodity taxes for any revenue target. We will then review equity issues.

In Chapter 27 we saw that the DWL of commodity taxation rises with the elasticity of demand or supply. It follows that, to minimise DWL, higher taxes should be imposed on goods in inelastic demand or supply. Further, to minimise the total DWL of commodity taxes the taxes should be set so that the marginal DWL from the last dollar of revenue raised from each commodity is equal. If this is not so, total DWL can be reduced by raising the tax rate on the commodity with the smaller marginal loss and by reducing it on the commodity with the higher marginal loss. To find the tax rates that equate the marginal DWL for each commodity, we need to determine (1) the quantity changes that result in the lowest total DWL and (2) the tax rates that produce these quantity changes.

Consider a simple example of DWL shown in Figure 28.4. The compensated demand curve is linear and there is a constant marginal cost curve. The DWL from a unit tax (T) on commodity X (area ABC) equals:

$$DWL = 0.5T\ (X_1 - X_2) = 0.5T\ \Delta X \tag{28.10}$$

where ΔX is the change in X consumed. Given that tax revenue is TX_2, the ratio of the DWL to revenue raised is:

$$\frac{Deadweight\ loss}{Revenue\ raised} = \frac{0.5T\Delta X}{TX_2} = \frac{0.5\Delta X}{X_2} \tag{28.11}$$

If there is no tax initially and a small ($1) unit tax is imposed, Equation 23.11 can be interpreted as the ratio of marginal DWL to marginal revenue.

A similar equation would apply to other commodities Y and Z. Therefore, to equate the marginal DWLs of the last dollar of revenue raised for each commodity, the following equality must hold.

$$\frac{0.5\Delta X}{X_2} = \frac{0.5\Delta Y}{Y_2} = \frac{0.5\Delta Z}{Z_2} \tag{28.12}$$

where Y_2 and Z_2 are the amounts of Y and Z consumed after the marginal tax is imposed.

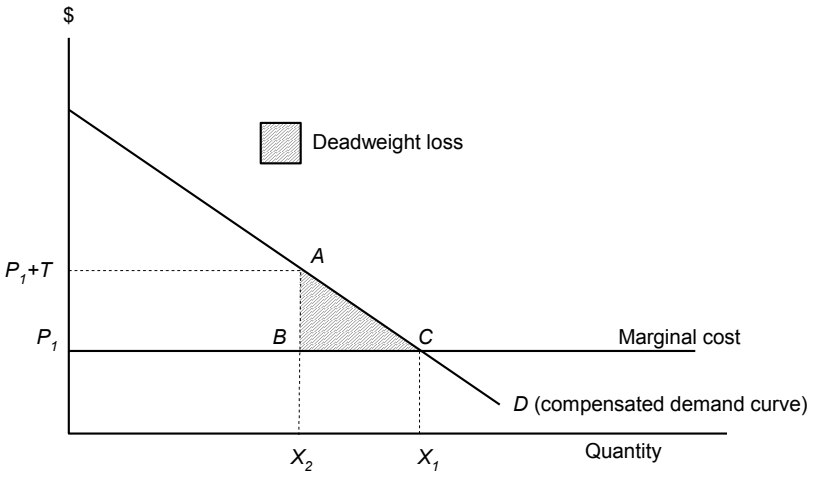


Figure 28.4 Deadweight loss revisited

Multiplying Equation 28.12 by two produces:

$$\frac{\Delta X}{X_2} = \frac{\Delta Y}{Y_2} = \frac{\Delta Z}{Z_2} \tag{28.13}$$

The equi-proportional principle. Equation 28.13 implies that, to minimise the DWL of a set of commodity taxes, the percentage reduction in the quantity of each commodity consumed must be the same. Strictly speaking, this is true for small changes where X_2 is close to X_1 etc.

Ramsey (1927) showed that, if there are no income effects and if demand curves are linear, a given amount of revenue is raised from selective commodity taxes with least DWL when taxes produce equal proportional reductions in the consumption of each good. Although these are strong conditions, Atkinson and Stiglitz (1980) show that the principle of equal proportional reductions has quite general application.

The inverse elasticity rule. To determine the tax rates that are consistent with this equi-proportional principle, recall the standard demand equation:

$$\frac{\Delta Q}{Q} = \eta_{cd} \frac{\Delta P}{P} \tag{28.14}$$

where η_d is here the compensated demand elasticity. Now if $\Delta Q/Q$ is held equal to a constant k for all commodities as required by Equation 23.13, it follows that

$$\frac{\Delta P}{P} = k \frac{1}{\eta_{cd}} \tag{28.15}$$

where k reflects the revenue requirement. Note that $\Delta P/P$ is the tax rate (t). Equation 28.15 implies that, to achieve equal proportional changes in consumption, prices must be increased in inverse proportion to the compensated elasticity of demand. This is known as the inverse elasticity rule.[4] It follows that, to minimise DWL, tax rates should also be inversely proportional to the compensated demand elasticities. Tax rates should not be uniform. They should be higher for commodities with a low elasticity of demand.[5,6]

We have assumed so far that the supply of commodities is perfectly elastic. If supply is included in the analysis, the efficient tax rate (t^*) is:

$$t^* = T/P = k\ (1/\eta_{cd} + 1/\eta_s) \tag{28.16}$$

where T is the per unit tax, P is the producer price after the tax has been levied, k is the proportionality factor that depends on the total revenue that government seeks to raise and η_s

---

[4] These results are similar to the optimal price mark-up rule to achieve a financial result in Chapter 17.

[5] The exposition in the text broadly follows Rosen and Gayer (2014, pp. 349-51, see footnote). Rosen and Gayer show that the result can be obtained more rigorously as follows. Suppose the welfare objective is to minimise total excess burden from two commodities subject to a revenue constraint. The total excess burden for two commodities X and Y is $0.5\,\eta_x P_x X t_x^2 + 0.5\,\eta_y P_y Y t_y^2$, where t is the tax rate. The revenue constraint is $P_x X t_x + P_y Y t_y = R$, where R is the required revenue. The problem is to choose t_x and t_y to minimise the total excess burden subject to $R - P_2 X t_x P_y Y t_y = 0$. Set up the Lagrangian expression:

$$L = \frac{1}{2}\eta_x P_x X t_x^2 + \frac{1}{2}\eta_y P_y X t_y^2 + \lambda[R - P_x X t_x - P_y Y t_y]$$

where λ is the Lagrange multiplier. Setting $\delta L/\delta t_x = 0$ yields $\eta_x t_x = \lambda$ and $\partial L/\partial t_y$ yields $\eta_y t_y = \lambda$. Hence $\eta_x t_x = \eta_y t_y$. This implies that tax rates should be inversely related to elasticities.

[6] Strictly this rule applies only if goods are unrelated in consumption, i.e. the compensated demand for each commodity is independent (the cross-price elasticities are zero). The implications of non-zero cross-price elasticities are usually considered to be minor.

is the supply elasticity. When supply is perfectly elastic (∞), the efficient tax rate is inversely proportional to the elasticity of compensated demand.

Other efficiency considerations

In practice Ramsey's inverse elasticity of demand theorem has limited application to taxation as distinct from optimal price mark-ups for public utility pricing. The theorem was designed to show how to raise a given revenue target from a limited set of commodity taxes. However, for any given revenue target a limited set of taxed commodities must always have higher tax rates (and therefore higher DWL) than a broader set of taxed commodities with lower tax rates. Given that most developed economies other than the United States have a general consumption tax (although with quite large exemptions), the technical and policy question is whether to adopt differential commodity taxation with a broad-based consumption tax.

There are two main arguments for adopting a uniform commodity tax. Technically, under certain conditions a uniform commodity tax is efficient. The literature here is quite complex (see for example Deaton, 1979) but the underlying idea is that if a uniform commodity tax does not change either individuals' proportionate allocation of income to goods or the amount of leisure they take, then a uniform commodity tax is non-distorting. Secondly, administrative simplicity points against detailed differential commodity taxation. Comprehensive specific commodity taxes would require detailed knowledge of compensated demand elasticities, which are generally not well known. The administration and compliance costs of a differential commodity tax system would be much higher than for a uniform commodity tax. Differential commodity taxes are also open to business and political manipulation.

A uniform, or indeed, any tax on consumption may discourage work relative to leisure. However, it is hard to tax leisure. To deal with the work–leisure distortion, Corlett and Hague (1953) proposed that goods that are consumed jointly with leisure (e.g. sporting goods) should be taxed more highly than other goods such as food and clothing, which could indirectly reduce the demand for leisure. While this would be efficient, it is unlikely that any tax authorities have tried consciously to implement such a strategy. However, in so far as they tax luxury items more heavily than necessities, they may have stumbled towards it.

A different issue arises with externalities and demerit (or merit) goods. As we have seen, there are strong arguments for taxing negative externalities like carbon emissions and traffic congestion and for taxing demerit goods like tobacco and possibly drugs of various kinds. There may also be arguments for subsiding or lower taxes on merit goods like gymnasiums or fitness providers.

These kinds of considerations have led many economists, notably including the influential Mirrlees Review (MR, 2011) of taxation, to advocate basically a uniform commodity system with exceptions only for major externalities and demerit goods.

Equity issues

The inverse price elasticity rule for efficient commodity taxation states that goods in inelastic demand should be more highly taxed. However, this may not be equitable. Price elasticities may be low for necessities like basic foods and transport and high for luxuries such as fillet beef and sailing boats. This observation is often used as an argument against application of the inverse price elasticity rule.

This argument is sometimes extended. Arguably, for equity reasons, goods that are consumed heavily by the poor should bear lower tax rates. If low-income households spend a greater proportion of their income on food, for example, than do high-income households then a tax on food is regressive. This view has led to adoption of zero taxation of many goods, notably food and health services, in the VAT or GST commodity tax systems in many countries.

However, there are major counter arguments. The most important one is that commodity taxation is only part of an overall tax system. As Kaplow (2008), MR (2011) and many other economists have argued, income taxation can meet any redistributive goals including any inequities that may arise from commodity taxation. Moreover, it is better at meeting these goals. Income taxation can be targeted on income levels adjusted for household differences. Commodities are often purchased by a wide variety of households. In addition, any loss of tax revenue from these exemptions has to be compensated from other tax sources, often raising the MTR and creating efficiency costs. Differential commodity taxation also complicates administration and compliance with all the associated administrative and legal expenses.

Taxes on Producer Goods and International Trade

Many governments tax producer goods (intermediate goods traded between producers) and internationally traded goods. Taxes on companies are administratively convenient and can raise a large amount of tax revenue from relatively few tax sources. Taxes on imports protect labour and capital in competitive domestic sectors. On the other hand, taxes on exports increase the local supply of goods and thus help to control local prices.

Despite these arguments, taxes on producer and traded goods are poor taxes. Consider first **producer goods**. As Diamond and Mirrlees (1971) demonstrated in an often-cited article, people's level of welfare depends on the income they receive from selling labour and capital and the prices they pay for the goods they consume. Government can control the prices of final goods independently of the prices of intermediate goods and, subject to behavioural responses, can achieve any desired distribution of real income. Government cannot improve welfare further by taxing intermediate goods.

On the contrary, taxing producer goods is likely to distort relative prices of inputs to the production process and reduce productive efficiency and hence welfare. As seen in Chapter 3, productive efficiency requires that the marginal rate of technical substitution between any two inputs is the same for all firms that use those inputs and that the marginal rate of transformation between any two outputs is the same for all firms producing those outputs. To achieve this, the prices of producer goods should reflect their real opportunity cost. Any tax on producer goods will distort relative prices and break the requirements for productive efficiency. It also distorts corporate structure because inter-company transactions are taxed whereas intra-company transactions are not. Turnover taxes increase the tax on small companies that buy and sell each other's goods. Large firms have a government-created competitive advantage because they pay less for inputs than do small firms.

It follows that government should not tax trade between firms or alter the prices that producers face in inter-firm trades. Also, government should not tax company turnover. Taxes on income (including return on capital) or consumption can achieve desired welfare outcomes more efficiently than can taxes on production. Taxes on intermediate goods are inefficient and distort production. There are also better instruments to obtain distributional objectives.

International trade. The production efficiency theorem also implies that a small country that has no influence on world trade prices should not tax imports or exports. International trade enables a country to convert exports into imports. This allows a country to consume more goods than it could do otherwise. Import taxes should be replaced by domestic taxes on the same goods. For any given amount of revenue raised, there will be less DWL.

Figure 28.5 overleaf demonstrates the DWL. The S_L and S_I schedules represent local and international supply of goods respectively. With no tariff the price would be P_1 and Q_1 amount of goods would be consumed, with Q_4 goods supplied locally. With the tariff shown, the price would rise to P_2 and consumption would fall to Q_2, with Q_3 goods produced locally. The government would collect $ABED$ in tariff revenue, but this would be a transfer from

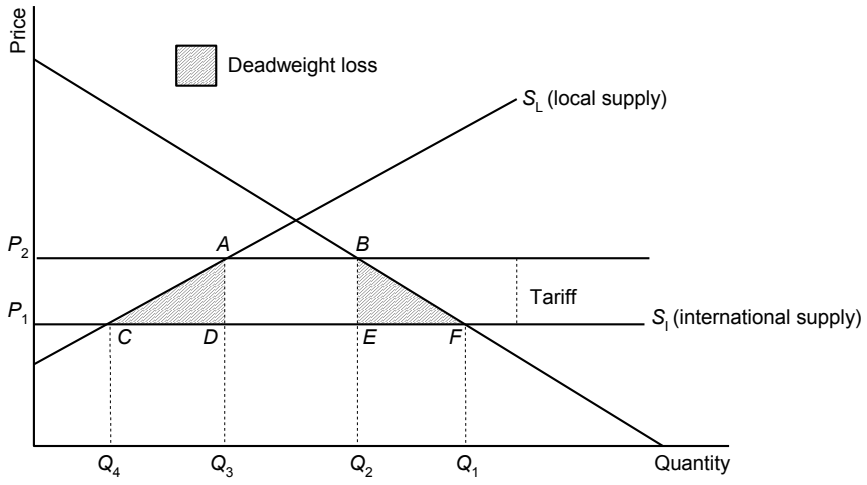


Figure 28.5 Deadweight loss of tax on international trade

consumers to government. The DWL would be the sum of triangles *ADC* (inefficient production) and *BFE* (loss of consumer surplus).

These conclusions about non-taxation of intermediate goods and international trade depend on certain assumptions. First, wages are flexible, and removal of trade distortions does not create unemployment. If wages are inflexible, they may not reflect true opportunity costs. Second, firms are competitive and do not make economic profits (these are profits over and above a normal return to capital). Alternatively, government can tax economic profits (as distinct from accounting profits). Third, the country is a price taker. Taxes on trade can improve a country's real income when a country is a price maker rather than price taker. Fourth, government can tax incomes on final goods in such a way as to produce the desired social welfare outcome, including compensation in some cases for workers who would lose from policy changes such as a reduction in import duties.

Arguably, if one or more of these assumptions do not hold, there may be a case for taxation of producer goods or international trade, or both, especially when the administrative costs of such taxes are low. However, given that taxes on producer goods and international trade almost always reduce productive efficiency, proponents of such taxes need to show that the welfare gains more than offset the economic losses.

An Optimal Tax System

The analysis of optimal taxation above is limited in several ways. First, income and consumption taxes were analysed as separate systems. Fundamentally, what matters is the efficiency and equity of the tax system overall. Individual taxes do not need to be equitable if the system overall is equitable. Second, our analysis of income taxation focused on taxation of labour income. This did not fully resolve the appropriate taxation of income from capital. An optimal tax system needs to consider how to treat savings and, more formally, how to maximise social welfare over time.[7] Third, only light consideration was given to administration and compliance costs. Fourth, we ignored political factors. For example,

[7] Atkinson and Stiglitz (1980, Chapter 14) discuss a formal approach to determining optimal taxation in an inter-temporal model.

differential commodity taxation creates opportunities for rent seeking (by obtaining favourable tax rates) by special interest groups. Some public choice economists argue that, given the tendency of government to act as Leviathan, welfare is maximised by limiting the power of government, including limiting available tax instruments.

In this section we review the major features of the main tax instruments (albeit as separate instruments) and then discuss how these instruments might be combined to create an overall optimal tax system.

Main tax options

Table 28.1 overleaf shows the main tax bases and options along with some comments on advantages and disadvantages. The comments are only indicative of possible effects. The merits or otherwise of a tax often depend on the circumstances and on how it fits in a package of taxes.

An individualised **head tax** based on ability would be equitable but is not feasible. A uniform head tax would have no DWL but have low revenue-raising capacity because it is constrained by the income of the least well-off taxpayer and inequitable. Thus, the main tax bases are income, consumption and wealth.

A general **income tax** would tax all sources of income at the same rate. The tax structure could be linear or non-linear. In practice, most income tax systems are hybrids. They tax income from different sources at different rates and have non-linear tax structures.

Four forms of **consumption tax** are identified. There are three forms of general consumption tax. These are an expenditure tax (a tax on all income less saving), a general consumption tax (GST or VAT) with investment goods excluded and a general retail sales tax. As we saw in Chapter 26, a general consumption tax is equivalent to a tax on labour. An expenditure tax could have a linear or non-linear tax structure. Despite its potential efficiency and equity advantages, such as tax has never been implemented. A VAT system usually has high compliance costs. On the other hand, with a retail tax it can be difficult to distinguish consumer goods from producer goods. In practice, commodity taxation is usually a hybrid differential tax system, with some commodities not taxed and others taxed at various rates.

Wealth taxes are taxes on the capital value of assets. Taxation of land and natural resources is common. It is sometimes easier to tax wealth (such as property values) than income or consumption. However, taxes on income or consumption can achieve most outcomes that wealth taxes can achieve and taxes on wealth may represent double or even triple taxation.

Features of an optimal tax system

In principle, a complete optimal tax system can be modelled as an exercise in social welfare optimisation, designed to balance efficiency and equity, subject to work–leisure preferences and present–future consumption preferences and so on (see for example Atkinson and Stiglitz, 1980). However, the results are sensitive to assumptions about (1) income differences between households, (2) the structure of household preferences over commodities and leisure, (3) the tax instruments available, including negative tax policies and (4) the nature of the social welfare function, including equity preference.

Nevertheless, some key principles emerge from the many reviews of optimal taxation.

1. All taxes should be viewed as part of an overall taxation system. Moreover, because direct personal income taxes and benefits can be designed to achieve any desired distributional outcomes, other tax instruments should be designed to achieve efficiency.
2. Normal returns to new capital investment should be exempt from income taxation.
3. Taxation should be comprehensive and neutral. A broad tax base reduces tax rates. To avoid distortions, similar activities and goods should be taxed in a similar way.

Table 28.1 Main tax instruments with some advantages and disadvantages

Tax instrument		*Possible advantages*	*Possible disadvantages*
1	**Head tax** Individualised on ability or uniform	If tax individualised on ability, equitable and efficient	Individualised head tax not feasible Uniform head tax not equitable and has low revenue raising capacity
2	**Income tax on labour and capital**	Can be based on personal income and other individual circumstances	Taxes labour supply and other factors of production Distorts inter-temporal consumption decisions Encourages non-market activity
2a	Linear general income tax Taxes equally all sources of income Can include negative income tax and be progressive	May distort less than a partial income tax system Low administrative cost Low compliance costs	May be less equitable than more progressive non-linear tax system
2b	Non-linear general income tax Taxes equally all sources of income Marginal tax rates usually rise with income	May be more equitable than (2a)	High marginal rates may distort use of resources May have high administration and compliance costs
2c	Non-linear differential income tax Taxes different sources of income, e.g. labour and different types of capital, at different rates Marginal tax rates usually rise with income	By treating savings differentially may be more efficient and equitable May be more equitable generally than (2a) and (2b)	Can significantly distort income-gaining activities and investments High marginal rates inefficient Higher administration and compliance costs than (2a) or (2b) May encourage rent seeking
3	**Consumption taxes** General consumption tax is equivalent to taxation of labour income	Do not distort inter-temporal consumption	Tax labour supply indirectly Can be based only indirectly on individual circumstances Encourage non-market production
3a	General expenditure tax Tax on all forms of income less saving Can be linear or non-linear	Treats savings efficiently and equitably Treats all income in same way Can be implemented equitably	More difficult to administer than general income taxes
3b	General value-added tax Taxes value added at all stages of production for all consumption goods at same rate but usually exempts investment expenditures	May capture income that would otherwise evade tax	Discourages market production (encourages home production) High compliance cost
3c	General retail tax Taxes all goods at the same rate but only at retail stage	May capture income that would otherwise evade tax	Discourages market production Difficult to distinguish between retail and non-retail purchases
3d	Differential commodity taxes Taxes commodities (by VAT or at retail stage) at different rates	Corrective taxes are efficient May be equitable	May be neither efficient nor equitable High compliance costs Encourages rent seeking
4	**Wealth taxes** Often imposed at differential rates on different assets	Efficient if taxes on fixed assets May be more equitable than other forms of taxes	May discourage investments High taxation of savings

4. Tax revenue is raised most efficiently by taxing factors of production or goods with inelastic supply or demand respectively. This implies that fixed resources may be taxed efficiently. Labour with a high elasticity of supply should be taxed lightly at the margin.
5. Producer (intermediate) goods should not be taxed.
6. Taxes should be targeted at problems. Thus, major negative externalities or demerit goods may be taxed.
7. To avoid distortions and minimise administration, tax should be as simple as possible. Any advantages of differential taxes should be weighed against complications such as defining and policing boundaries. Thus, there should be few departures from general uniform taxes.

The highly credentialed Mirrlees Review (MR, 2011) of taxation in the UK drew essentially on these principles. Their conclusions form the basis of the following description of an optimal tax system.

MR recommended that the tax system should be progressive and neutral (non-distorting). The idea that the tax system overall should be progressive is of course a normative one but an idea that doubtless has widespread public support (although apparently not from the Tea Party in the United States).

For the income tax, MR recommended that there should be a simple and clear progressive tax rate schedule. This would include a single benefit to support those with low income and/or high needs. There would then be two or at most three rising tax rates (as distinct from one MTR). However, the schedule could reflect evidence of responsiveness to tax rates especially where the rates could not affect behaviour. Thus, there could be lower MTRs for mothers of school age children and for people around retirement age.

Income from all sources of labour and all sources of capital should be taxed according to the same rate schedule. Taxation of capital and labour at a similar rate maintains their relative prices, reduces distortionary substitution of one factor for another and encourages an efficient allocation of both labour and capital. It also discourages high-income individuals from adopting artificial schemes to convert income into capital or vice versa to reduce tax.

However, all costs of generating income, including the normal cost of generating savings, would be deductible. MR would allow savers to deduct the opportunity cost of the capital, defined as the return on a government bond rate, from their return on capital. Thus, there would be no tax on the normal return on capital. This system would allow capital to be taxed at the same rate as labour but avoid the double taxation of savings without all the *ad hoc* tax reductions given to savings in many tax systems, including the Australian tax system.

Turning to commodity taxes, it may be asked why they are needed if income tax can provide sufficient revenue. One reason is that they treat income from savings equitably with income from labour. However, we have just seen that there are other ways of treating savings fairly. Second, commodity taxation captures some income that would otherwise evade tax. In practice, if not in the ideal world of optimal tax, many people arrange their affairs, legally or otherwise, to escape tax. Thus, commodity taxation can contribute to a more equitable tax outcome. Third, consumption taxes affect labour supply less directly than income taxes. Assuming some fiscal illusion, for given revenue a combination of consumption and income tax may distort the labour supply by less than income tax does.

With regards to the structure of commodity taxation, MR takes as a starting point that a uniform commodity tax (via VAT or GST) should be charged on all final consumption by households. A broad taxation base lowers tax rates. There should also be equivalent taxes on hard-to-tax housing and financial services. Zero and reduced rates of tax on sales of commodities would generally be avoided. There may be lower taxes (or some equivalent benefits) for child care and educational investments and higher taxes on tobacco or alcohol. But the list of exceptions should be small. General differential taxation of commodities is too

difficult. Taxes to correct market failures or taxes on damaging goods are also justified. MR picks out carbon and congestion taxes. On the other hand, there should be no taxes on business inputs, business turnover or transaction taxes.

Finally, turning to taxes on wealth, MR argues that there is a strong case for taxing pure economic rents. These taxes are not distorting. Thus, a tax on land is justified if this can be separated from buildings and from improvements to land. Carefully designed taxes on natural resources may also tax pure economic rents. MR also supported on equity grounds taxing transfers of wealth, gifts and bequests, particularly between generations. However, the review noted that there are significant practical difficulties.

As shown in Box 28.1, the Henry Tax Review (2010) in Australia produced similar recommendations to the MR recommendations.

Tax Reform

Optimal tax principles and recommendations provide critical guides to tax reforms. However, there are some special problems in implementing tax reforms. We discuss these problems immediately below before discussing some strategic responses.

First, we note a technical problem. A key feature of tax reform is that it involves a choice between two or more highly imperfect systems. From the second-best theorem we know that when an economy has two or more distortions, removal of one distortion may not reduce the total DWL. Also, a change to a tax may affect have positive efficiency effects but negative equity effects or the reverse. Thus, it may not always be clear whether a tax change improves overall social welfare. In any case, reform packages may be necessary. Rather than assessing each tax reform separately, changes to the tax system should be assessed as a whole.

Box 28.1 Key recommendations of the Henry Tax Review

The Australian tax system has too many taxes and too many complicated ways to deliver policy objectives. The personal tax structure combined with the personal grants system should be the sole way to deliver progressivity in the tax system.

Revenue raising should be concentrated on four efficient and robust tax bases: personal income tax, business income, private consumption and economic rents from land and natural resources. Other taxes should be maintained only if they efficiently address social or economic costs. The Henry Review highlighted taxes on tobacco, alcohol and gambling, and road user charges. Over time all other taxes should be eliminated. This includes insurance taxes, payroll taxes, property transfer taxes, resource royalties, taxes on superannuation contributions in the super funds, income taxes on all government benefits, and fuel and vehicle registration taxes.

The personal income tax system would include all forms of income including employer superannuation contributions, but not government benefits. There would be a higher tax-free threshold of about $25 000 and a simple, transparent two-step tax scale beyond that. To allow for inflation there would be an arbitrary 40 per cent discount for interest income, net residential rents and capital gains.

There would be three levels of primary support payments. These would be pensions for the aged, disabled and carers, lower-rate participation allowances for those of working age, and assistance payments for young persons and students. Means-tested withdrawal rates would reflect different work expectations. Family assistance would be paid through a simplified single program.

The company income tax rate should be maintained to the lower end of small to medium OECD economies with a reduction to 25 per cent in the medium term.

The Henry Tax Review argued strongly for a broad-based land tax but rather confusingly said that there could be a threshold and different rates based on the value per square metre of land. It also recommended resource rent taxes to replace mineral royalties and road user charges to replace other vehicle and fuel taxes.

The Review also recommended that one-off consumption taxes, such as taxes on insurance, should be eliminated and that a more efficient tax on financial services be developed. Restricted by its terms of reference, the Review did not recommend any changes to the GST that currently covers less than 60 per cent of all final goods and services.

Second, related to the first point but perhaps more important in practice, tax reforms will nearly always produce winners and losers. Many of the losers may oppose the tax change. If a tax reform generates an aggregate net welfare gain, it should be possible to compensate losers where this is deemed appropriate. When the fringe benefits tax was introduced, income tax rates were reduced. When the GST was introduced, pensions were increased and various state taxes were eliminated. Thus, both efficiency and equity can be achieved.

However, compensation can be difficult in practice. For example, a carbon tax (even with some exceptions) will affect millions of individuals in some way. It would be hard to achieve full compensation. For a start, the winners and losers may not always be clear. This occurs especially when gains or losses are capitalised. Capitalisation occurs when changes in asset values reflect the impacts of any actual or expected taxes on the after-tax income stream. Asset prices adjust so that investors obtain a similar after-tax return on assets of similar expected risk. Note that asset prices reflect expected tax changes. Introducing an unexpected tax on an asset may be viewed as unfair to investors who have paid a high price for an asset. It is also important that compensation does not set up distorting incentives. Finally, introducing taxes without compensation is especially difficult with short three-year political cycles as applies to the Australian government.

Third, multi-government systems or fiscal federalism are a major obstacle to tax reform. The Australian government collects 80 per cent of all taxes and consequently dominates policy making at sub-national as well as at national level. The recommendations of the Henry Tax Review could increase this already excessive vertical fiscal imbalance. The Review recommended abolishing a raft of state and territory taxes including payroll, royalty, insurance, property transfers and motor vehicle taxes and it is not clear how these will be replaced by an extended land tax (assuming this remains a state tax) and road user charges. The Review was remarkably light on how vertical fiscal imbalance would be addressed.

Strategies for tax reform. What strategies for tax reform emerge? First, tax reform needs to be based on clear principles and a good road map. The Mirrlees and Henry tax reviews provided both principles and a road map. Related to this, tax reform should be viewed and assessed as part of a tax system.

Moreover, both the Mirrlees and Henry reviews recommended that distributional objectives be achieved wholly through the income tax and transfer system and that all other taxes should be assessed against efficiency criteria. Acceptance of these principles is a major constructive first step towards tax reform.

Second, once these principles are agreed, the next step is to design a workable and simple income tax/transfer system that meets at least most of the distributional objectives. This is likely to be a progressive system but one with relatively few tax rates. To be efficient, income would be defined to include most, or all, components of income. Some deduction for the costs of earning income, including child care and costs of savings, as per MR, may be allowed.

Third, the tax reform program would identify and progressively eliminate the taxes with the highest deadweight losses. The consumption tax base would also be broadened to minimise tax rates, since deadweight losses rises by the square of the tax rate. For administrative and compliance simplicity, the tax rate structure would be fairly uniform other than for some important exceptions.

Fourth, the transition to the new tax system may have to include income compensation for clear cases of hardship or inequity. One way in which this can be done is by slow transition. This reduces inequity associated with unexpected tax changes. But it also risks policy stagnation or in some cases policy reversals.

Summary

- The optimal tax system is sometimes viewed narrowly as the system that raises the required revenue at least deadweight loss. More broadly, it is the system that maximises social welfare with the optimal combination of taxes and income transfers.
- Tax systems are efficient and deadweight loss is minimised when marginal tax rates are low.
- On the other hand, equity rises with average tax rates. Thus, redistribution objectives usually involve rising marginal tax rates.
- Optimal tax rates rise with government's general revenue requirement, wage inequality and concern for equity. They fall with higher (compensated) labour supply elasticities.
- Income taxes and transfers can achieve all distributional objectives. Thus, other taxes should be designed to minimise deadweight losses.
- Some studies find that linear tax systems (those with a constant marginal tax rate) are close to the optimal form of income tax. But others find that a variety of non-linear tax schedules are optimal.
- For a selective set of commodity taxes, the deadweight loss is minimised when tax rates are inversely proportional to demand elasticities.
- However, deadweight loss is also minimised by a broader set of taxes with lower tax rates. Also, differential taxes increase compliance and rent-seeking costs. Therefore, a broadly uniform commodity tax rate is generally preferred along with corrective taxes and taxes on demerit goods.
- Taxes on producer goods and international trade are generally inefficient. Government should not alter the prices that producers face in trades between themselves or that local citizens face in trades with foreigners.
- Major tax reviews (chaired by Mirrlees in UK and Henry in Australia) found that tax should be based on four major tax bases: income (personal and business), consumption and wealth (or land and natural resources) along with some special taxes on major demerit goods and externalities. Other taxes should be eliminated.
- Most proposals for tax reform are based on base broadening, more uniform tax rates, reducing the discrimination against savings, tax simplification and rewriting tax statutes to reduce tax evasion.
- Tax reform is complicated because there are many winners and losers. Losers may receive income compensation so long as this does not produce adverse incentives. Also, the transition process may be slow, providing that this does not result in a stalled or reversed reform process.

Questions

1. 'Distributional objectives can be achieved simply by raising marginal income tax rates.' Discuss.
2. Marginal tax rates are constant with proportional and progressive linear income tax schedules. What then is the main difference between these two schedules?
3. Income from capital should not be taxed. Discuss.
4. Why should only final goods be taxed and not intermediate goods?
5. Consider a small country which levies a tax on imports equal to \$2 per unit of import. Suppose that the supply of the domestic producer is given by $Q^s = 50 + 2P$ while domestic demand is given by $Q^d = 90 - 3P$.

   If the world price for the imported good is \$3, determine the following:

   i. The quantities supplied by domestic producers and by imports if there is no import tax.

   ii. The revenue that government raises from the tax on imports.

   iii. The deadweight loss due to the import tax.
6. The compensated demand curve for good A is given by $Q^d_A = 50 - 2P$. Supply is perfectly elastic and the marginal cost of producing a unit of A is equal to \$10. A unit tax of \$2 per unit is imposed on good A. Suppose the observed demand for a second good B is $Q^d_B = 100 - 7P$, supply is perfectly elastic and marginal cost is equal to \$10. Determine the following:

   i. The change in quantity of good A consumed as a result of the tax.

   ii. The ratio of the deadweight loss to revenue raised for good A.

   iii. The quantity of good B consumed.

   iv. The tax rate to be imposed on good B that will minimise the deadweight loss.

7. Some optimal tax studies find that, given a choice between income and commodity taxation, the optimal solution is complete reliance on income taxation. Should consumption be taxed as well as income in an optimal tax system? Give reasons.
8. A uniform commodity tax is simply an increase in linear income taxation. True or false? Why?
9. Draw on public choice theory to explain why most governments fail to index income taxes. Does this have any implications for optimal tax design?
10. Individuals with low incentives to work should be taxed at lower marginal tax rates. Discuss
11. What key principles do the Mirrlees and Henry tax reviews share?
12. What are some differences of substance or emphasis between the Mirrless and Henry tax reviews? What reasons might there be for these differences?
13. What are the key issues arising in tax reforms?
14. What impacts, if any, do international factors have on optimal taxation?

Further Reading

Auerbach, A.J. (2006) *The Choice between Income and Consumption Taxes: A Primer*, Working Paper 12307, National Bureau of Economic Research, Cambridge, MA.

Auerbach, A.J. (2006) 'The future of capital income taxation', *Fiscal Studies*, 27, 399–420.

Benham, F. (1942) 'What is the best tax system?', *Economica*, IX, 115–126, reprinted in *The Foundations of Public Finance*, Vol. 1, 1996, 124–135, Edward Elgar, Cheltenham.

Blacklow, P. and Ray, R. (2002) 'Optimal commodity taxes in Australia', *Australian Economic Review*, 35, 45–54.

Feldstein. M., (1976) 'On the theory of tax reform', *Journal of Public Economics*, 6, 77–104.

Creedy, J. (2009) 'Personal income taxation: from theory to practice', *Australian Economic Review*, 42, 496–506.

Creedy, J. (2010) 'Tax and transfer tensions: designing direct tax structures', *Australian Economic Review*, 43, 103–113.

Gruber, J. (2014) *Public Finance*, 5th ed., Chapter 25, Worth Publishers, New York.

Henry, K. (Chair) (2010) *Australia's Future Tax System: Report to the Treasurer*, Commonwealth of Australia, Canberra, www.taxreview.treasury.gov.au

Kaplow, L. (2008) *The Theory of Taxation and Public Economics*, Chapters 4 to 6, Princeton University Press, Princeton, NJ.

Mirrlees, J. (Chair) (2011), 'Conclusions and Recommendations for Reform', Chapter 20 in *Tax by Design: the Mirrlees Review*, Institute for Fiscal Studies, London.

Mankiw, N.G., Weinzierl, M. and Yagan, D. (2009) 'Optimal taxation in theory and practice', *Journal of Economic Perspectives*, 23, 147–174.

Rosen, H. and Gayer, T. (2014), *Public Finance*, 10th ed. Chapter 16, McGraw-Hill, New York.

Saez, E., Slemrod, J.B, and Giertz, S.H. (2009) *The Elasticity of Taxable Income with respect to Marginal Tax Rates: A Critical Review*, Working Paper 15012, National Bureau of Economic Research, Cambridge, MA.

Stiglitz, J. (2000) *Economics of the Public Sector*, Chapters 20 and 25, W.W. Norton, New York.

Chapter

29

Government Borrowing and Debt

Be not made a beggar by banqueting upon borrowing.

Ecclesiasticus 18:33

Budget Deficits and Public Debt: Meaning and Measures ♦ Funding a Budget Deficit ♦ Economic Effects of Budget Deficits ♦ The Burden of Deficits and the Public Debt ♦ Borrowing and Debt Policies

The use of debt to finance capital investment is a long-standing practice in the public sector as in the private. Clearly it is fair that taxpayers who receive the benefits of public investment should pay for much, if not all, of the investment over the life of the asset. Also, since the 1930s, following Keynes, there has been widespread support for running deficits on the government's current account in times of economic recession as a means of increasing aggregate demand, economic activity and employment without creating inflation.

On the other hand, there is widespread belief that budget deficits and public debt are bad. Budget deficits are perceived to indicate poor economic management and to encourage inflation, international trading deficits and declining exchange rates. And when public debt rises to very high levels, lenders fear that government may default on its financial obligations and are reluctant to lend. Under this scenario, interest rates rise, governments lose capacity to employ deficit-based growth strategies and growth rates decline (Reinhart and Rogoff, 2010).

These are lively concerns in the debate about economic management in the United States and European Union as well as in Australia. However, there is much confusion about what budget deficits and debt mean. There is a widespread failure to distinguish between operating deficits and total budget deficits that include capital expenditure. Accordingly, in this chapter we start by defining these concepts and providing relevant Australian data. The second section discusses the various ways in which budget deficits can be funded. Sections three and four discuss the macroeconomic impacts and incidence of debt respectively. Finally, we discuss public policies for budget deficits and the public debt inclusive of international data.

Budget Deficits and Public Debt: Meaning and Measures

As we saw in Chapter 2, there are two principal measures of budget balance: the net operating balance and the net lending balance. The latter is also known as the fiscal balance and is the most common measure of budget surplus or deficit.

The actual and projected budgets of the Australian and NSW governments for 2016–17 and 2017–18 as reported in the respective 2016–17 budget papers are shown in Table 29.1.

Table 29.1 Operating statements of Australian and NSW governments ($ billions)

	Australian government		*NSW government*	
	2016-17	*2017-18*	*2016-17*	*2017-18*
Revenues				
Taxation revenue	386.2	415.4	30.4	31.6
Commonwealth grants	Na	na	30.8	31.1
Other grants and subsidies	Na	na	1.0	0.7
Sales of goods and services	8.9	11.5	8.5	9.0
Interest income	3.8	4.4	0.8	0.4
Dividend income	3.7	3.5	2.2	2.6
Fines, regulatory fees and other			4.4	4.4
Other	9.6	9.6		
Total current revenues	412.1	444.4	78.0	79.9
Gross operating expenses				
Wages and salaries	19.9	19.9	30.6	31.5
Superannuation	9.7	4.6	3.1	3.0
Depreciation	9.5	9.2	4.6	4.9
Payments for goods and services	91.3	102.5		
Other operating expenses	5.7	5.8	18.5	19.6
Total gross operating expenses	136.2	141.9	56.9	59.1
Superannuation interest expense	8.4	10.4	1.3	1.5
Interest expenses	17.6	18.5	2.1	2.1
Current transfers				
Current grants[a]	134.1	138.2	13.2	14.6
Subsidy expenses	12.7	12.1		
Direct personal benefits payments	129.9	130.5		
Total current transfers	276.7	280.8	13.2	14.6
Capital transfers (grants and write-downs)	11.9	12.6		
Total expenses	450.8	464.3	73.5	77.2
Net operating balance	-38.7	-19.8	4.5	2.7
Other economic flows[b]	5.9	-6.6	38.6	10.2
Full result: total change in net worth	-32.8	-26.4	43.1	12.9
Net acquisition of non-financial assets				
Purchases of non-financial assets	12.2	13.6	10.7	14.3
Plus change in inventories	0.2	0.2	0.0	0.0
Plus movements in non-financial assets	0.0	0.0	0.3	0.3
Less depreciation	9.5	9.2	-4.6	-4.9
Less sales of non-financial assets	0.8	4.0	-3.7	-5.3
Total net acquisition of non-financial assets	2.0	0.5	2.7	9.1
Net lending balance (budget surplus)	-40.7	-20.3	1.8	-6.4

(a) These are mainly grants to state and territory governments.

(b) Mainly changes in asset values included in 'Full result: total change in net worth' but not included in either net operating balance above or in net lending balance.

Sources: Australian Treasury, *Budget Strategy and Outlook 2016–17* Budget Paper No. 1, Statement 10, Table 1; NSW Government, *Budget 2016–17*, Table A1.1.

The table is structured in accordance with *Government Financial Statistics* (GFS) prepared by the Australian Bureau of Statistics, which all Australian governments are required to follow. The amounts reflect accruals rather than cash transactions. The Australian government figures include GST revenue, which it collects and passes on to the states.

Net operating balance
Current revenue less current expenses

The GFS **net operating balance** equals current revenue less current expenses. It is similar to a private sector profit and loss account. Revenue includes taxes, income from sales of goods and services, interest and dividend income, fees and fines. Current expenses are composed of operating expenses and transfers. Operating expenses include depreciation allowances but not expenditure on new capital assets. In an accrual budget, superannuation expenses are workers' entitlements, not what government may lay aside for superannuation. Capital transfers are grants to other bodies that are used for capital projects.

Net lending (borrowing) balance
Surplus (shortfall) in government revenue less total current and capital spending

The GFS **net lending balance** measures the excess of government revenue over total spending during a period (usually a year). The lending balance allows for capital as well as current transactions. The net lending balance equals the net operating balance minus net capital investment (NCI). NCI equals capital acquisitions[1] less capital depreciation and asset sales. The resulting surplus (or deficit) figure indicates the government's net lending (or borrowing) requirement in accrual terms.

As shown in Table 29.1, the Australian government ran a substantial deficit of $39 billion in the net operating budget in 2016-17 and planned for another large operating deficit in 2017–18. Net capital expenditure (the net acquisition of non-financial assets) added a small amount to create the higher overall budget deficit.

On the other hand, the NSW government ran an operating surplus in 2016-17 and planned another surplus in 2017-18. These results are supported by large grants from the Australian government that accounted for about 40 per cent of current revenues. Because state governments spend proportionately more on capital formation than does the Australian government, the overall budget balance is usually negative, indicating a borrowing requirement. In the case of NSW, the borrowing requirement was reduced by asset sales. This reduces the cash deficit. But sales do not increase public net worth.

The Australian government budgets are prepared to high professional and international standards and the bases for the budget are well explained in the annual *Budget Paper No. 1*. Nevertheless, budget figures should be viewed carefully, particularly with respect to off-budget liabilities and treatment of publicly owned enterprises. For example, the Australian Government has spent some $30 billion on the National Broadband Network (NBN) and total expenditure is expected to be around $50 billion. But the Government does not include these expenses in the annual budget because it regards them as investments in financial assets (akin to investments by public trading enterprises) that will be repaid along with a commercial return on the capital in due course. This is a very optimistic view of the financial prospects for the NBN Corporation.

Expenditure and financing including borrowing by public trading enterprises (PTEs) are not included in general government budgets. The total public sector borrowing requirement includes the net borrowing requirements of PTEs as well as all net borrowing by central, state and local governments.

Net public debt and net public worth

Net public debt
The value of government's stock of financial liabilities less financial assets

As we saw in Chapter 2 (see Box 2.3), there are three main measures of a government's overall financial position: net public debt, net financial liability and net public worth.

The **net public debt** (NPD) is the difference between the government's stock of financial liabilities and financial assets. In any year, the change in the NPD equals the net lending

[1] This excludes military equipment, which is recorded as a current expense in the budget financial statements.

balance. At any given time, NPD equals the sum of past budget deficits less any surpluses used to redeem the debt.

Because government may borrow at home or abroad, the financial liabilities may be owed to local or foreign citizens or firms. Also, government assets may be held in Australia or overseas. However, NPD is not the same as the foreign debt. The foreign debt is the sum of all public and private financial liabilities to foreigners less foreign financial assets held by Australians.

NPD is important because, together with interest rates, it indicates the call on future revenue flows to service that debt. However, it is a partial measure of a government's overall financial position. For example, NPD does not include accrued employee liabilities or any intangible liabilities associated with government guarantees. In some countries, unfunded pension guarantees exceed the net debt. In the United States, where the Social Security fund is run off-budget with contributions that are supposed to fund benefits, the actuarial shortfall allegedly runs into US$ trillions, although views vary on the exact amount. This shortfall is a political obligation of the US government, although future governments could reduce pension payments.

Net public worth (NPW) equals total publicly owned assets less total liabilities. It is the most comprehensive measure of a government's overall asset position.[2] NPW includes a government's non-financial assets, such as equity held in PTEs, as well as all its liabilities. The NPW measure recognises the difference between debt incurred to finance capital formation and debt incurred to fund consumption. When debt finances capital formation, NPD increases but NPW does not change. When debt is used to finance consumption, NPD increases and NPW falls.

Net public worth
Total publicly owned assets less total liabilities

The balance sheets of the Australian and NSW governments at 30 June 2017 are shown in Table 29.2 overleaf. At this date, the Commonwealth had an estimated negative NPW of –$292 billion, with gross liabilities of $809 billion exceeding assets of $507 billion. Employee entitlements, notably superannuation, accounted about one-third of the Commonwealth's liabilities. Between 2001 and 2008, Australian government budget surpluses transformed the NPD from $42 billion into −$22.0 billion (i.e. a net surplus). Since then, large budget deficits have increased the net public debt to $325 billion.

The position of the NSW state government is very different. It has a high and rising net public worth. NPW rose from $91 billion at 30 June 2001 to $225 billion at 30 June 2017. Over this same period the NPD of the NSW government was virtually unchanged at around $8 billion.

Official Australian statistics quoted in Box 2.3 (Chapter 2) indicate that as at 30 June 2017 the NPD of all Australian and state governments together amounted to only 22 per cent of GDP. The IMF figures quoted in Table 29.3 (page 523), which appear to be based on gross public debt rather than on net public debt, indicate that the Australian debt to GDP ratio is well below comparable ratios in most OECD countries.

Issues in interpretation

Several issues of interpretation arise with significant policy implications. We discuss below accruals versus cash accounting, the role of inflation, current versus capital accounting, off-budget items and the concept of the standardised deficit.

[2] Net public worth should be distinguished from net national worth. The latter includes private wealth. A few years ago, a World Bank report estimated that Australia is the richest country in the world in per capita terms with a wealth of $835 000 per capita due to its natural resource assets.

Table 29.2 Estimated balance sheets of Australian and NSW governments at 30 June 2017 ($billions)

	Commonwealth	*NSW*
Assets		
Financial assets		
Cash and deposits	3.4	5.6
Advances paid	54.1	1.0
Investments, loans and placements	181.9	
Other receivables	49.7	42.8
Equity investments [a]	92.2	116.9
Total financial assets	381.0	166.3
Total non-financial assets [b]	136.5	182.5
Total assets	517.5	348.9
Liabilities		
Interest bearing liabilities		
Deposits held	0.2	0.1
Advances received	0	0.8
Government securities	546.9	
Loans and other borrowing	17.3	33.0
Total interest-bearing liabilities	564.4	34.0
Provisions and payables		
Superannuation liability	175.0	53.7
Other employee entitlements and provisions	18.2	16.7
Other liabilities	50.1	19.6
Total provisions and payables	244.8	90.0
Total liabilities	809.2	123.9
Net public worth [c]	-291.7	224.9
Net public debt [d]	325.1	7.8

(a) Virtually all equity in public trading enterprises and public financial enterprises.
(b) Nearly all held in land and fixed assets.
(c) Net public worth equals total assets less total liabilities.
(d) Net public debt equals the sum of deposits held, advances received and borrowing less the sum of cash and deposits, advances paid and investments, loans and placements.

Sources: Australian Treasury, *Budget Strategy and Outlook 2016–17*, Budget Paper No. 1, Statement 10; NSW Government, *Budget Statement 2016–17*, Table A1.2.

Accrual and cash accounting. As noted, the figures in Table 29.1 are accrual numbers. These show revenue due and expenditure incurred but not actual cash payments. Australian governments are also required to provide cash flow statements.

In any period, the cash balance can be quite different from the accrual fiscal balance. This is important for the practical business of running government. Depending on the timing of cash revenues and payments, government may have to borrow for temporary working purposes over and above what it requires to fund the fiscal deficit measured in accrual terms.

Inflation. This also affects the real size of the budget deficit and government debt. The real deficit is the change in the real value of government debt over the year after adjusting for the effects of inflation and changes in real interest rates. Suppose that government has a net debt of $20 billion at the start of the year and the annual rate of inflation is 5 per cent. If there is no net borrowing and the nominal value of the debt stays constant, the real value of the debt falls by $1.0 billion in the year. This is equivalent to a budget surplus of $1.0 billion over the year.

Current and capital expenses. Traditionally government accounting did not recognise the difference between current and capital expenditures. Indeed, some governments (and many politicians) still do not recognise this distinction. For example, in the United States the central government and most state governments treat all expenditures as current expenditures.

Under the GFS system used by Australia, capital and current expenditures are distinguished. Capital formation is a capital expense. Use of capital is treated as a current expense (depreciation). If capital formation is financed by debt or by asset sales, there is little or no change in net public worth. When government borrows, it takes on a financial liability that broadly offsets the new capital asset. When government sells an asset, it exchanges one asset for another.[3]

Distinguishing between current and capital expenses allows us to focus on the net operating balance, which for many purposes is the best measure of the real economic deficit. In any period, the net operating balance plus revaluations of assets and liabilities equals the change in net public worth. On the other hand, the overall fiscal balance shows a government's funding requirement and indicates potential pressure on credit markets. In so far as government uses debt to finance capital expenditure, net public debt increases. This may matter if government has difficulty paying interest on the debt. But, providing the projects are economically efficient, there is no decline in net public worth. NPW is generally a better indicator of the economic strength of government than is NPD, which is a partial measure of government's overall financial position.

The main argument against separate capital budgeting is that current and capital expenses are not easily distinguished (Gruber, 2016). For example, the Australian budget counts expenditure on defence weapons platforms (DWP) as current expenses although the Australian Bureau of Statistics records DWP as expenditure on a non-financial (capital) asset on a fair market value basis. Likewise, expenditure on education and health services is often an investment in future output or wellbeing, but it is also treated as a current expense. On the other hand, expenditure on buildings, plant and equipment are treated as capital expenses. Clearly there are practical difficulties in distinguishing between capital and current expenses and this can lead to abuse. In 1993 the Clinton administration proposed that the major US food stamp program was an investment because it improved diets for the poor and therefore their productivity.

Off-budget items. Governments often incur (and indeed legislate) obligations and liabilities that are not recognised in the budget. For example, government may underestimate the liabilities accruing in retirement pensions for its employees or the costs of its guarantees or legal responsibilities. When the major Queensland floods occurred in January 2011 the Australian government introduced a special levy during the financial year to fund the unexpected compensation obligations. In the United States, the government has effectively guaranteed the debts of quasi-government mortgage giants Fannie Mae and Freddie Mac. More generally, budgets often do not allow for the performance of separate off-budget security funds like the Social Security Fund in the United States or the National Insurance Fund in the UK. Underfunding of the US Social Security Fund has meant that the real US budget deficit has been significantly underestimated in accrual terms. Also, as we have noted, PTEs are off-budget. In the UK, policy makers pay considerable attention to the total public sector borrowing requirement (PSBR).[4]

---

[3] Net public worth may rise if the capital formation is efficient or fall if it is inefficient.

[4] For discussion of the PSBR in the United Kingdom, see Bailey (2002, Chapter 11). In the United States the difference between all federal government expenditures and revenues is called the 'unified budget deficit' (Hyman, 1999, Chapter 12).

Cyclically adjusted budget outcome
The budget outcome that would occur with current tax arrangements at a specified normal level of activity

The cyclically adjusted budget deficit. In any year, the size of the deficit reflects both the structural (long-term) imbalance between revenues and expenditures and temporary factors. The most important temporary factors are fluctuations in economic activity associated with the business cycle. When economic activity is low, tax revenues are low and social security payments are high. Accordingly, the deficit is higher than it would be with more buoyant conditions. Other temporary factors may include one-off payments or timing factors, though timing is less important in accrual budgeting. For macroeconomic management, it is sometimes considered useful to estimate a cyclically adjusted budget outcome. This is the budget outcome that would occur with the current tax arrangements if the economy were at some specified normal level of activity. This may be a full employment economy but may also be a mean point in the business cycle that allows for an unemployment rate of perhaps 5 per cent. Any deficit at this mean point is deemed to be a structural deficit. The aim is to isolate any structural imbalance in the budget from cyclical effects. A temporary budget deficit is of less concern than a structural deficit. It follows that, when economic activity is low, the actual budget may be in deficit, but the cyclically adjusted budget may be in surplus.

Funding a Budget Deficit

As shown in Table 29.2, Commonwealth interest bearing liabilities totalled $564 billion at June 2017 and the 2016-17 budget papers forecast a further borrowing requirement of over $40 billion in the year. These liabilities are managed by the Australian Office of Financial Management (AOFM) which sits in the Australian Treasury. The AOFM issues Treasury Notes, Treasury Bonds and Treasury Indexed Bonds, manages the government's cash balances and invests in high quality financial assets. Treasury Notes are short-term borrowings typically over 90 days designed to deal with the government's within-year financial task. These notes may be required to meet cash requirements even when the government has a budget surplus. Treasury bonds are the main instrument used to meet the government's financing requirements.

The government can also fund its deficit by drawing down on its savings in the form of deposits held with the Reserve Bank or other state-owned funds such as the Building Australia Fund or the Education Investment Fund. Alternatively, it can borrow directly from the Reserve Bank of Australia (RBA) at the RBA-determined interest rate.

However, in the Australian system the RBA plays a central role in determining how the budget will be financed and the economic outcome. This reflects the independent role of the central bank in Australia. The RBA decides whether to finance and purchase Treasury Notes and Bonds on its own account or whether to sell these to the public. In effect, the RBA decides whether to allow the money supply in the economy to change (to monetise the deficit) or not to do so. If the RBA decides to allow the money supply to increase, it holds the Commonwealth securities on its own account or buys them back from the public. This allows the government to borrow without any impact on interest rates. On the other hand, if the RBA sells the notes or bonds to the public, domestic or foreign, the sale reduces the money supply generally available and increases interest rates. In this way the central bank has the capacity, via control over the money supply, to counteract partly at least any short-run fiscal policy that the Treasury may wish to put in place.

When the central bank is not independent of the government, the government can also fund a budget deficit by borrowing from the central banking system at any interest rate (including no interest) that the government determines. In effect, the central bank provides an overdraft facility at little or no charge out of which the government pays for goods and services and provides transfer payments. These cheques are paid into accounts at commercial banks, which in turn present them to the central bank for payment. The central bank then credits the commercial banks' accounts held with the value of these cheques. Although government does

not actually print the extra money, this is sometimes described as government printing of money.

In exceptional circumstances, this method of government funding has no economic cost. This occurs when an increase in the money supply induces an increase in output, which in turn increases the demand for money to match the supply, and there is no increase in prices. However, when an economy is close to full employment, an increase in the supply of money increases prices rather than output. The cost is the inflation tax (see Chapter 25). The tax is the fall in the real value of money balances, including bank deposits. This is a capital loss sustained by holders of money. Thus, governments in developed economies rarely finance deficits by borrowing from the banking system without paying appropriate interest because the potential to raise revenue without inflation is limited. Even moderate money financing of the order of 1–2 per cent of GDP has been associated with inflation over 50 per cent in many developing countries (Burgess and Stern, 1993).

Since the global financial crisis erupted in 2007–08, several governments, notably the United States and the UK, have adopted monetary policies called 'quantitative easing' that some commentators have described as 'printing money'. Quantitative easing involves the purchase of financial assets from banks and other private sector institutions, sometimes at above market prices, with new electronically created money. The scale of these operations was much greater than conventional open market operations designed to increase the money supply by relatively small amounts. However, the objective was similar, namely to inject money into the private sector when interest rates were very low rather than to fund government activities themselves.

Government can also fund a deficit by selling financial or productive assets. In a cash budget, asset sales reduce the nominal budget deficit and thus may reduce NPD. But the sale generally has no substantive economic impact. The sale of assets to the public has a similar economic effect to borrowing. In both cases the public provides revenue to government. In one case, it receives a financial asset in return. In the other case, it receives equity in a physical asset. But critically, NPW is the same whichever method of funding of capital formation is adopted.

Although asset sales may be shown in the budget as capital revenue, they are essentially a form of deficit finance. Suppose that government can finance additional expenditure by debt or by asset sales. Government would issue debt or equities respectively and the public acquires bonds or shares. However, the effects on net public worth and on markets are similar in both cases.[5] Similar interest rate rises and crowding out effects occur, with a similar reduction in total capital formation. The two financing methods also have similar impacts on future taxpayers. Borrowing increases the future taxes required to pay interest and debt. Asset sales increase future taxes by a similar amount because they reduce future government dividend income, the present value of which equals the asset sale value.[6]

Economic Effects of Budget Deficits

We consider below how borrowing to finance investment and consumption respectively may affect key macroeconomic variables, notably economic output.

[5] If the public has a preference for debt or equity, there may be very minor second-order wealth effects on spending.

[6] Unless productivity is higher in private ownership, in which case asset sales are justified because the present value of the assets is higher in private ownership than in public ownership (see Chapter 18).

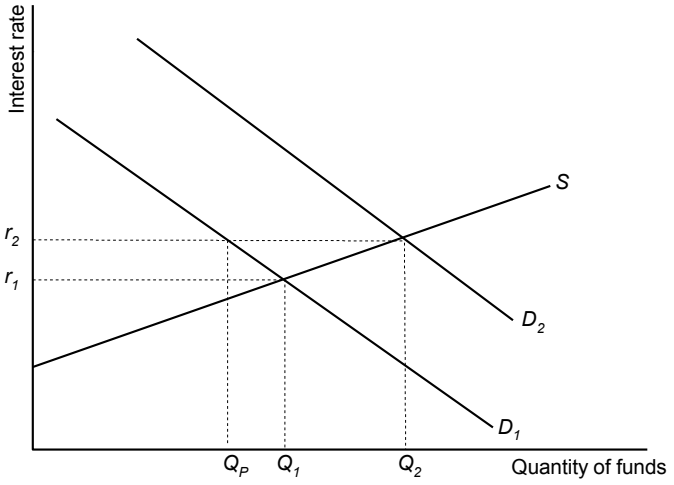


Figure 29.1 Demand and supply of capital

Borrowing to finance public investment. We examine first how borrowing to finance public investment may affect interest rates, private investment and output, assuming no change in the money base. Figure 29.1 shows the demand and supply for capital funds. The D_1 schedule shows private demand for capital and D_2 shows total demand inclusive of government demand. Given the supply of capital, government demand raises the rate of interest from r_1 to r_2 and the quantity of capital employed increases from Q_1 to Q_2. However, with the rise in interest rates, private investment falls from Q_1 to Q_P. Overall, total investment increases but there is some crowding out of private investment.

In this scenario, the extent to which private investment is crowded out depends on the demand and supply elasticities for capital. Crowding out is low if the private demand for capital is inelastic or if the supply of capital is elastic so that public borrowing has only a small impact on interest rates. The latter occurs in an open economy with an elastic international supply of capital. Also, the central bank may keep interest rates stable by holding more Treasury bonds and thus increasing the money supply. In effect, this increase in the money supply would shift the supply of capital curve to the right. Thus, there are a number of ways, mainly through the inflow of international capital or increases in the money supply, whereby public investment can increase total investment.

However, this is a simple partial equilibrium model. We need to consider both how increased public investment could affect household consumption and hence aggregate demand and how employment levels may affect the outcome.

Ricardian equivalence
Government actions do not change private equilibria

Government borrowing implies paying less tax now but more in the future. If individuals wish to retain their preferred inter-temporal consumption pattern, they must therefore save more today. Indeed, individuals may change their behaviour precisely to offset any actions of government! This view is also known as the **Ricardian equivalence**. Ricardo, a 19th century economist, argued that individuals consider the effects of the public debt on their own and their children's living standards and increase their bequests to their children when government deficits increase future indebtedness. With Ricardian equivalence, taxation and borrowing are equivalent ways of financing expenditures and have the same effect on private consumption over time. However, the Ricardian equivalence assumptions, which also include perfect capital markets and non-distortionary taxes, are not very realistic. In a study of households' saving responses to government deficits and a review of other such studies, Brittle (2010) concluded that household savings increase significantly with budget deficits,

offsetting perhaps 50 per cent of the deficit, but that responses are not fully Ricardian and that there remains a significant fiscal stimulus.

When households lend to government they may indeed reduce their consumption. But they can reduce consumption by less when lending than when taxed an equivalent amount because they can maintain consumption by borrowing against the security of government assets. Also, foreign capital may directly finance government borrowing or may finance individuals who want to maintain their consumption levels while lending to government. Providing that individuals do not save more now in anticipation of higher tax rates in future years, aggregate spending is likely to be higher if government finances investment by debt than by taxation. Accordingly, unless monetary policy is restrictive, financing capital expenditure by borrowing is likely to increase aggregate demand and be expansionary.

The amount of expansion depends on the level of employment (the resources available) in the economy and the openness of the economy. The greater are the resources available, the larger is the economic stimulus and the lower is inflation from any increase in aggregate demand. With full levels of employment, budget deficits increase the pressure on factor prices and the private demand for capital curve in Figure 29.1 shifts to the left. However, even with full employment, with an open economy additional resources may be imported without any immediate impact on local prices. Thus, if exchange rates do not change, budget deficits tend to be expansionary.

International transactions also affect exchange rates. The capital required to finance imported goods increases the claims of foreigners on future output. In so far as government borrowing increases the inflow of foreign capital, demand for the local currency increases and the exchange rate rises in the short run. However, when the principal and interest are repaid, foreign lenders may wish to repatriate their money and sell the local currency, thus lowering the exchange rate. A low exchange rate is expansionary for the local economy, but it also lowers real income by increasing the price of imported goods.

Finally, two further caveats should be noted about these observations on the effects of borrowing on key economic variables such as output, prices, savings and exchange rates. First, the effects are based on a series of partial equilibrium observations. For example, the conclusion that a budget deficit tends to reduce private investment assumes that the deficit does not change the level of aggregate expenditure and hence national output. If a budget deficit increases total output, this may in turn increase national saving and so offset any tendency for the deficit to reduce saving. Reliable macroeconomic predictions require (reliable) general equilibrium economic models.

Second, the observations assume broadly rational market responses. Sometimes financial markets react negatively to a budget deficit regardless of the reason for it because they think it indicates either government inefficiency or excessive expenditure that will produce inflation and a depreciating exchange rate. This may cause the price of equities to fall, interest premiums to rise, and private investment to fall. These effects of a budget deficit may have real economic consequences.

Borrowing to finance public consumption. From an accounting perspective, borrowing increases NPD by the amount of the loan regardless of whether the loan is financing public consumption or capital formation. But the impact on net public worth is quite different. Borrowing to finance consumption reduces NPW. Borrowing for investment maintains NPW (assuming the asset created provides equivalent services) and may even increase net public worth.

Turning to economic impacts, short-run macroeconomic outcomes are likely to be independent of whether borrowing finances consumption or capital formation. Although the composition of government expenditure would be different, total government expenditure would (or can) be the same. Because a budget deficit is likely to increase aggregate demand,

it is likely to stimulate economic output and employment at low levels of employment, but it may be inflationary at high levels of employment.

However, the use to which borrowing is put does affect longer term outcomes. When government borrows to finance public investment, total capital formation increases even if there is some crowding out of private investment. When government borrows to finance consumption, there is no net increase in capital formation and total capital investment may fall if there is any crowding out of private investment with higher interest rates. Thus borrowing to finance consumption reduces the capital to labour ratio, which in turn reduces labour productivity and future output and real wages. Present consumption is enhanced at the expense of future consumption.

The Burden of Deficits and the Public Debt

Who pays for budget deficits or, equivalently, for the net public debt? In a simple generational accounting model, future generations of taxpayers would appear to bear the burden of the NPD. However, actual tax incidence often differs from nominal incidence, and this applies also to the burden of the debt. We consider below the incidence of domestic borrowing to finance consumption and investment respectively and then consider briefly the implications of foreign borrowing.

Generational accounting
Measuring the effects of government fiscal policy on the members of each generation

Domestic debt funding of consumption. A traditional view popularised by Lerner (1948) is that financing consumption by internal debt rather than by taxation creates no burden for future generations! Future generations bear no net burden because, at any point in time, debt repayment simply transfers funds from taxpayers to bondholders. Therefore, the debt, whatever its size, does not affect the aggregate consumption of future generations. This is true in a closed economy.

If future generations as a whole are no worse off as a result of the debt, can government fund present consumption simply by raising debt at no cost to anyone? If they could, economists might even become popular! Alas, there is no free lunch. In a closed economy, three main groups of people are affected by deficit funding of public consumption:

- current taxpayers who gain increased consumption;
- bondholders who forgo current consumption but obtain compensation via increased consumption in the next period;
- future taxpayers who experience reduced consumption.

In this scheme, borrowing does not increase current consumption. Bondholders have rearranged their consumption over time and are compensated for forgoing current consumption. The beneficiaries of debt-funded consumption are current taxpayers and the burden for paying for this consumption lies with future taxpayers.

Overlapping generations model
A model in which individuals live in more than one period

An overlapping generations model clarifies the beneficiaries and losers from debt-funded consumption. In simple versions of the model, current and future taxpayers are called current and future generations. But some individuals are members of overlapping generations. Old members of the current generation are beneficiaries of debt-funded public consumption because they receive the consumption goods but do not survive to pay for it. However, young members of the current generation are also members of the next generation—they both benefit from debt-funded consumption and pay for it. On the other hand, young members of the next generation are taxpayers and obtain no benefits. Thus, based on the present values of lifetime consequences of government policies, older individuals benefit from debt funding at the expense of younger and yet-to-be-born individuals (see Rosen and Gayer, 2014).

There are two further reasons why debt-funded consumption may disadvantage future generations even in a closed economy. First, borrowing to finance public consumption may raise interest rates and reduce private investment, which could reduce future output and wages. Second, if NPD rises too high, government may have to pay a higher price for capital in the future.

Domestic debt funding of investment. On the other hand, the use of domestic debt to fund efficient investment creates no net cost to present or future generations. The present generation loses consumption because resources are diverted to capital formation. However, lenders who forgo current consumption by buying government bonds are compensated by increased consumption over the life of the bonds and when the bonds are redeemed. On the other hand, although future generations pay for the capital formation through increased taxes or user charges, they gain from increased goods and services. Given efficient financial markets and efficient investment, the gains should exceed the costs although some individuals in the next generation may be worse off from increased taxes. Of course, if investment is inefficient, borrowing can create significant costs for the next generation(s).

External debt. In many countries, including notably the United States and Australia, foreign investors hold a large part of the public debt. If, instead of taxation, foreign bondholders finance public consumption, there is less sacrifice of present consumption because fewer local bondholders are forgoing consumption. But there is increased cost, possibly a much-increased cost, to future generations because future taxpayers are paying foreign rather than local bondholders. There is also vulnerability to exchange risk if foreign bondholders wish to withdraw their funds.

On the other hand, if foreign loans finance capital formation, the impact depends on the productivity of the capital. If the return on the public investment is greater than the cost of the foreign funds, the combination of the investment with the foreign funds makes future generations better off.

Borrowing and Debt Policies

When, then, should governments use debt to finance their expenditure? And should governments have NPD or possibly NPW targets? These are key and topical public finance and political questions. In this section, we discuss short and medium-term deficit policies and then discuss further net debt and public worth policies.

Short-run macroeconomic management. Much of the debate about budget deficits over the last 50 years or so has revolved around the short-term macroeconomic effects of running deficits or surpluses. As discussed above, budget deficits (including operating deficits) tend to increase aggregate demand and are an important instrument of short-term employment and price-level objectives. Thus, operating deficits are generally desirable, and may be essential, when economic activity is depressed, while surpluses are desirable in boom times.

To some extent economies adjust via automatic stabilisers. Government budgets tend to be in surplus in boom times and in deficit in slumps. However discretionary fiscal (or monetary policy) is generally needed as well as automatic stabilisers impact slowly. Following the global financial crisis starting in late 2007 many governments temporarily adopted large budget deficits to counteract the extreme crisis of confidence and lack of lending in the private sector. This is generally thought to have avoided a very serious worldwide recession that would otherwise have occurred.

However, there are various constraints on the effectiveness of such fiscal policies. First, government cannot always act with the foresight and speed required to fine-tune the economy

with greater or smaller budget deficits (or surpluses). There are also lags between government spending and taxing decisions and their impacts. The Australian government's $40 billion stimulus package, the national jobs and building program for schools and public housing, introduced in February 2009 was planned to be substantially completed by the middle of 2010 but was not completed until 18 months later.

Second, stimulus programs may be both inefficient uses of resources and fail to deal with the root causes of the economic downturn. For example, they may not deal with excessive lending for consumption or low-quality assets. The real problems may be systemic, not temporary. Unless spending programs create useful assets, they may increase government's net debt while reducing NPW. This occurred in the global financial crisis in 2008. What started as a private market credit crisis has now morphed into a government credit crisis.

These difficulties with fiscal policy led to more reliance on monetary policy from around 2010 onwards as governments became concerned with running deficits. A key feature of monetary policy is the setting of interest rates. Another key feature in many countries is the independence of the central bank in determining interest rates. This reliance on low interest reduces, but does not eliminate, the role of fiscal policy in short-run macroeconomic management. Indeed, many authorities (OECD, 2016, and IMF 2017) believe, that in recent years, excessive reliance on monetary policy at the expense of fiscal policy has been a mistake (see also Abelson and Dalton, 2017).

Golden rule of public finance
Over the business cycle government borrowing should equal net capital investment

The golden rule of public finance. Turning to the medium term, the golden rule of public finance is that, over the business cycle, government borrowing should equal net capital investment. The rule is based on the distinction between current and capital expenditure. To ensure meaningful operation, revenue and expenses should be accounted for in accrual rather than in cash terms. The practical implication of the golden rule is that government should achieve a *zero net operating balance on average over the course of the business cycle*. The rule allows operating deficits in recession years, but requires surpluses in boom years. It disallows borrowing to fund consumption in the medium term.

The golden rule is based on equity (fairness) and efficiency considerations. The major equity issue is intergenerational equity. The beneficiary principle of equity requires that those who receive the benefits of services should pay for them. Borrowing to finance current consumption transfers the burden of payment for present benefits to future taxpayers (other than bondholders). On the other hand, borrowing to finance capital expenditures spreads the costs of capital expenditure over time in accordance with the benefits that are generated. In addition, in so far as future generations are likely to have a higher income than the present generation, they are better able to pay for the services.

From an efficiency perspective, avoiding a deficit on the current operating budget is an important financial and management discipline. As Shakespeare's King Henry IV laments:

> I can get no remedy against this consumption of the purse. Borrowing only lingers it out, but the disease is incurable (Shakespeare, *Henry IV*, Pt. 2).

Failure to achieve an operating balance reduces NPW and can raise the cost of finance for all government borrowing even in the short run if potential bondholders fear that government expenditure will cause inflation.

Of course, borrowing to finance capital expenditure presumes that projects are efficient, the social return on investment exceeds the cost of capital and that the projects are not simply displacing private projects. Consequently, NPW will increase and there is no net burden for future generations. The golden rule does not justify borrowing to finance white elephant projects that create a burden for future generations.

Thus, the golden rule is a medium-term policy that is consistent with short-run variations in the size of the budget deficit according to macroeconomic requirements and monetary policy. But this assumes that government starts from an acceptable financial position. The golden

rule may not be appropriate if government has too little NPW or too much NPD. Questions may also arise if the amount of debt held by foreign interests is viewed as excessive.

Public Debt and Net Public Worth Policies: Further Discussion

In practice, many financial institutions (including credit agencies) and governments are more stressed about debt than NPW. This is not entirely rational. Countries with high NPW can nearly always convert some non-financial assets into financial assets or use income from non-financial assets such as land, property and public trading enterprises to supplement their income from financial assets and so pay for any obligations arising from their debt position. However, discussions about government defaults are nearly always based on debt figures or, more specifically, the debt to GDP ratio. Accordingly, we first discuss the notion of a sustainable debt to GDP ratio.

To put the discussion into perspective, Table 29.3 shows reported *indicative gross* public debt to annual GDP ratios for selected countries. These figures must be treated cautiously because they are provided by national sources that are not always consistent in definition, for example with respect to government, or public sector, debt or to gross or net debt. The countries are divided in the table into high and low public debt to GDP ratios. For several high public debt countries, the debt/GDP ratios increased greatly due to government reactions to the onset of the global financial crisis.

Fiscal sustainability requires that the debt be readily serviceable. The product of debt and interest rates determines the interest payable to service the debt. In addition, it is usually necessary to pay down debt as assets mature. If debt is not repaid, the debt to GDP ratio rises as new capital is debt financed. Thus, suppose that a country has a debt to GDP ratio of 100 per cent with an average interest rate of 4 per cent and with the debt to be repaid on average over 20 years. The interest and debt payments would sum to 9.0 per cent of GDP (assuming no new borrowing). This is a high proportion of GDP to pay to bondholders, especially if they are external bondholders. As loans to external bondholders are redeemed the exchange rate may fall with a consequent impost on future generations. In any case, regardless of the source of the finance, such a high level of debt may necessitate increases in tax rates that would be inefficient and burdensome. Moreover, financial institutions and credit agencies question whether a government will have the fortitude to raise taxes or cut expenditure sufficiently, countries get down-rated and the rate at which government can borrow soars. Reinhart and Rogoff (2010b) estimated that debt/GDP ratios of less than 90 per cent have little effect on economic growth but that GDP growth rates typically fall by at least 1 per cent when the debt/GDP ratio rises above 90 per cent. However, this predated the major falls in interest rates and much larger debt/GDP ratios are now considered manageable (Abelson and Dalton, 2017).

Table 29.3 Gross public debt to GDP percentages for selected countries in 2017

High debt countries		*Low debt countries*	
Japan	233.9	Ireland	56.4
Greece	161.2	South Korea	40.0
Italy	120.1	Switzerland	37.2
Portugal	114.0	Australia	36.0
United States	109.6	Indonesia	30.0
United Kingdom	85.6	Russia	17.9
China	62.2	Russia	15.7

Source: IMF.

However, debt should not be considered separately from NPW. When NPW is significantly positive, government can finance future interest and debt payments in many ways other than by taxation, including by more borrowing to repay debts. On the other hand, although NPW is an important factor in a country's economic sustainability there does not seem to be any formula for setting a ratio of NPW to GDP. There is no strong reason why the state should own a prescribed percentage of the wealth of the nation.

In practice, national policies on budget deficits and debt to GDP ratios vary and, more importantly, are not always adhered to. In the United States, there is no set statutory set policy on deficits at the federal level, but every state in the union, except Vermont, is required by law to balance its budget each year. However, these are ex-ante expressions of will. There are few effective constraints on budgets ex-post running into deficits. As Rowley *et al*. (2002) observe, despite these legislative constraints, in 2001 state and local debt in the United States amounted to about 25 per cent of the national debt. Initially to protect the euro, the European Economic and Monetary Union required member countries to keep budget deficits below 3 per cent of GDP and public debt below 60 per cent of GDP. But these targets are well and truly history. In Australia, there are no legislated policies on budget deficits or debt at federal level.

On the other hand, few if any governments have a NPW target. Indeed, only in recent years have Australian governments made estimates of NPW. As we have seen, the Australian government (unlike the states) has a negative net worth. A negative NPW implies that past generations have consumed more than they paid for and are bequeathing liabilities to future generations. From 1996 to 2007 the Australian government aimed to achieve budget surpluses to rectify this and, in particular, to reduce its superannuation liabilities to its employees and prepare for its growing long-term obligations to public pensioners. However, from 2008 to 2011 this objective was subordinated to the aim of sustaining employment during the global financial crisis. Subsequently, the national government has continued to run annual operating and total budget deficits.

In conclusion, the golden rule of public finance is a good basis for fiscal policy. The rule requires that governments balance the **operating budge**t over the business cycle. This allows an operating deficit for short-term macroeconomic considerations when economies are significantly underemployed. Importantly, it also allows debt funding of capital investment.

However, the golden rule requires the discipline of operating surpluses in strong economic conditions. Operating surpluses may also be required if the debt to GDP ratio is so high that servicing the debt is not sustainable or if external debt is excessive and repayment could lead to a major fall in the exchange rate. Operating surpluses may also be required to ensure that large liabilities do not impose major tax burdens on future generations.

Summary

- The budget deficit measures the excess of government spending over revenues during a given period. The two main measures of the deficit are the net operating balance and the overall fiscal balance, usually measured in accrual terms.
- The net operating balance is current revenue less current expenses. The fiscal balance equals the net operating balance minus net capital investment. A negative fiscal balance indicates a financing requirement.
- Net public debt equals the financial liabilities of government less financial assets. Net public worth is total public assets less total liabilities. A change in net public worth reflects the net operating balance (along with any asset revaluations).
- To understand government's real financial position, allowances need to be made for the effects of inflation, current and capital expenses, off-budget items and the cyclically-adjusted budget deficit.
- The Australian government generally funds a deficit by issuing Treasury bonds or notes or by borrowing from the Reserve Bank at the bank-determined interest rate. The Reserve Bank decides whether to hold the Treasury bonds, which allows the money base to increase, or to sell the bonds to the public, which raises interest rates.
- Funding an operating deficit by borrowing instead of taxation generally increases current aggregate demand. This increases output when there is less than full employment. However, borrowing to finance consumption may reduce private investment and future output.
- The golden rule of public finance states that government should achieve a zero-operating balance on average over the course of the business cycle. Operating deficits may be warranted in recessions, but surpluses are required in boom periods.
- Operating surpluses may also be required if the level of government net debt is not sustainable, debts to foreigners are excessive or national savings are too low.
- When government borrows to finance capital formation, future generations gain improved services to offset the taxation required to service the debt. If projects are efficient, net public worth increases.
- Arguably in recent years, there has been excess reliance on monetary policy and not enough on fiscal policy.

Questions

1. If government could pay for services by borrowing from the central bank without paying interest and so avoid levying taxes, why should government not fund public services in this way?
2. There is a widespread view that reducing the budget deficit improves economic performance. What arguments might support this view? Do they have substance?
3. How would the sale of a public trading enterprise to the public affect the net operating balance, the net lending balance, net public debt and net public worth?
4. Suppose that, at the start of a year, the net public debt in Australia is A$500 billion, with three-quarters denominated in Australian dollars and a quarter in foreign currency. Suppose also that the rate of inflation in Australia in that year is 2 per cent and that the value of the Australian dollar rises by 3 per cent in the year. What are the implications for the real value of the debt at the end of the year and for government revenues?
5. Suppose that the Australian government agrees to underwrite a bid for the 2022 World Cup football tournament. Would this have any effect on the government's financial position?
6. What is generational accounting?
7. When, if ever, should government (i) finance consumption expenditures with debt or (ii) capital expenditures from taxes?
8. If the central bank has the power to decide whether to hold or sell government bonds and hence to increase or maintain the money base, does the budget deficit have any independent fiscal effect on the economy?
9. Does the generational incidence of debt depend on whether it is domestic or external debt? Should budget funding decisions depend on whether funding is provided from domestic or from external sources?
10. Nobel Laureate Paul Samuelson asked: 'Can it be truthfully said that internal borrowing shifts the war burden to future generations while taxing places it on the present generation?' and answered 'A thousand times no!'. Explain Samuelson's answer.

11. Reinhart and Rogoff (2010) found that GDP growth rates fell significantly when debt/GDP ratios exceed 90 per cent. Does this mean that high debt rates reduce economic growth?
12. What practical problems may arise with applying the golden rule of public finance?
13. Should governments have a net public debt target? If so, what target would you suggest?

Further Reading

Abelson, P. and T, Dalton (2017) 'Rebalancing Monetary and Fiscal Policies', Annual Conference of the Economic Society of Australia, Sydney (available at www.appliedeconomics.com.au)

Australian Treasury, annual, *Budget: Budget Strategy and Outlook*, Budget paper No. 1, AGPS, Canberra.

Feldstein, M.S. (2010) *Preventing a National Debt Explosion, Working Paper 16451*, National Bureau of Economic Research, Cambridge, MA.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed., Chapter 4, Worth Publishers, New York.

International Monetary Fund (2017) 'A Greater Role for Fiscal Policy', Chapter 1 in *Fiscal Monitor,* IMF, Washington.

OECD (2016) 'Using the Fiscal Levers to Escape the Low-Growth Trap', Chapter 2 in *Economic Outlook 2016,* OECD, Paris.

Reinhart, C.M. and Rogoff, K.S. (2010), 'Growth in a time of debt', *American Economic Review*, 100, 573–578.

Rosen, H.S. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapter 20, McGraw-Hill, New York.

Rowley, C.K., Shughart II, W.F. and Tollinson, R.D. (2002) 'Introduction', pp. xiii - xlviii in *The Economics of Budget Deficits*, Volume 1, eds: C.K. Rowley, W.F, Shughart II, and R.D. Tollinson, Edward Elgar, Cheltenham, UK.

Part 9

Taxation in Practice

Chapter

30

Taxation of Personal Income

Bachelors should be taxed heavily. It is not fair that some men should be happier than others.

Oscar Wilde

Basic Income Tax Concepts ♦ Personal Income Taxation in Australia ♦ General Issues in Personal Income Taxation ♦ Taxation of Labour Income ♦ Taxation of Income from Capital ♦ International Issues

Income is a commonly used word with many different meanings. In particular, legal and accounting definitions of income almost always differ from economic definitions. Because income tax is based on legislation, we need to know the legal definition of income (in any given country) as well as the economic definition. Accordingly this chapter starts with a discussion of the nature of taxable income and other basic income tax concepts.

In the second section we describe the main features of the Australian personal income tax system. The third section discusses general issues in personal income taxation. The fourth and fifth sections discuss more specific issues associated with the taxation of income from labour and capital. The last section discusses international taxation issues: the treatment of foreign income accruing to local residents and local income accruing to non-residents. While the chapter discusses mainly Australian tax practices, international examples are also cited. In any case, most of the issues discussed are common to most economies.

Basic Income Tax Concepts

There are four common elements to personal taxation—the tax base, the tax rate, the tax unit and the period. Here we focus on the tax base. Later sections discuss the other elements.

Economic income
The value of goods that can be consumed in any period without any change in net wealth

Economic income. Economists usually adopt the Haig–Simons definition of an individual's income as the value of goods that can be consumed in a given period without any change in net wealth.[1] This is equivalent to the sum of income from all sources inclusive of any changes in the value of assets held. Thus this definition includes all income from labour (including income in kind), in principle home-produced goods, income from capital including unrealised capital gains and imputed rents from ownership of assets (including owner-occupied homes), income from government and private bequests. It also includes income from all countries.

[1] Simons (1938, p. 50) defined personal income as the algebraic sum of (1) the market rights exercised in consumption and (2) the change in the value of the store of property rights between the beginning and end of the period. The origin of the concept is sometimes credited to a German legal scholar, Schanz (1896).

Some further points may be noted. First, this definition of an individual's income focuses on their access to consumption goods in any period and thus refers implicitly to an individual's net income (i.e. income after any costs associated with earning it are deducted). Second, the definition is based on accrued rather than realised income. Income includes unrealised increases in the value of assets. Third, with this definition, we are interested in real income (the amount of goods that can be consumed) rather than nominal income. Real income is measured in constant prices. If money income or asset values rise simply with general prices, real income has not changed.

Although this comprehensive concept of income is a useful benchmark against which to evaluate income for tax purposes, it has two main limitations. The first is a practical one. It is difficult to track and measure some components of income, for example unrealised changes in asset values, imputed rents and gifts. Such components are generally excluded from definitions of taxable income. However, some governments attempt to tax some forms of income in kind, for example provision of motor vehicles to employees, to ensure that employees do not escape tax by substituting goods for cash. The second issue is community acceptance of the components of income that should be taxed. Many societies are reluctant to tax services from a house (imputed rents) that have been purchased from income on which tax has been paid or to tax gifts provided from after-tax income.

Notions of what income should be taxed depend both on what is measurable and on social norms. Defining taxable income lies at the heart of debates about income taxation.

Taxable personal income. This can be defined as:

$$\text{Taxable personal income} = \text{assessable income} - \text{allowable deductions} \quad (30.1)$$

Taxable personal income
Legally assessed income less allowable deductions

Note that we are assuming here that the tax unit is the individual. This is the practice in most countries. However, as we will see, household composition can affect an individual's tax liability even when the tax unit is the individual.

Assessable income is the gross income of an individual that is subject to tax. This is defined in relevant tax legislation. In practice, the legislation is subject to ongoing interpretation by accountants, tax authorities, and ultimately the courts. Assessable income is almost always based on income from legal market activities and on realised rather than accrued income. But realised capital gains are often treated differently from other forms of income. And, income in some savings vehicles, such as superannuation (or retirement) funds, is often treated differently from other forms of income.

To determine taxable income, most governments allow individuals to deduct both work-related and some personal expenses from assessable income. Work-related expenses are expenditures required to earn income. They are inputs to business, not final consumption goods. Turning to personal expenses, many governments allow specified expenses to be deducted for social objectives. For example, the US government allows a range of deductions for medical expenses, mortgage interest, state and local taxes, and charitable contributions. Some Australian deductions are discussed below.

Personal income tax. This can be expressed as:

$$\text{Personal income tax} = (\text{taxable personal income} \times \text{average tax rate}) - \text{tax offsets} \quad (30.2)$$

Personal income tax
The product of taxable personal income and the relevant average tax rate less tax offsets

Tax offsets are deductions from tax liability as distinct from deductions from assessable income. Tax offsets are generally designed to provide targeted tax relief to people in need, for example individuals with dependents or with a high cost of living or for some other reason. For instance, a taxpayer might be allowed to reduce their tax liability by $\$X$ for each child under a certain age. A general feature of tax offsets is that each taxpayer receives the same dollar reduction in tax *independently* of their marginal tax rate. However, refunds are paid

only against tax liability and someone who pays little or no tax receives little or no benefit from the tax offset.

Another kind of tax offset is the tax credit that arises when another party has pre-paid tax on an individual's income. Employers may pay taxes on gross wages, financial institutions may pay tax on the gross interest payable to a lender and companies may pay taxes on profit attributable to shareholders. In these cases, individuals can reduce their personal tax liability by the amount of tax already paid.

Close substitutes to personal income tax. Many governments impose compulsory levies that are similar to personal income taxes. These include social security (national insurance) contributions, the superannuation guarantee levy in Australia and payroll taxes. A levy may be regarded as an income tax if (1) a change in the relevant levy rate would change the take-home pay of an employee and (2) there is no requited benefit.

Social security contributions
Charges levied on individuals or their employers to pay for social security benefits

Many OECD countries levy a **social security contribution** as a proportion of gross wages paid. These levies are payable for each *individual* worker whether employed or self-employed. Statutory responsibility for payment may lie with the employer or employee but, in most cases, employers deduct the levied amounts from earnings. Critically, the payments are based on individual earnings and a change in the social security contribution rate would generally result in a change in an employee's take-home pay. Turning to the second test, there is a requited benefit in that the social security schemes provide benefits in the case of unemployment, sickness or retirement that are usually related to contributions. However, the connection between benefits and contributions is often tenuous and the contributions are widely regarded as a form of tax (James and Nobes, 2002).

The superannuation guarantee levy in Australia is a little different in that there is a directly requited benefit. Since 1992 employers have been required to pay a minimum of 9 per cent (now 9.5 per cent) of an employee's gross wage into a superannuation fund of the employee's choice. Because there is a direct nexus between this contribution and the subsequent benefit, and because the payment is not made to government, the superannuation guarantee levy is not a tax.[2] It is a mandated private expenditure levied on individual earnings that is a close substitute for a tax. Also, a change in the levy rate would change the take-home pay of employees.

Payroll tax
A tax on the wages paid by an employer

A payroll tax is a levy on the *aggregate* labour payroll of firms, paid by the firm. The levy is a tax because the income accrues to consolidated revenue; there is no requited benefit. As we saw in Chapter 26, if the supply of labour is relatively inelastic a payroll tax is borne mainly by labour generally in the form of lower after-tax wages. However, because the tax is based on the aggregate payroll of a firm it has a negligible effect an individual's take-home pay when his or her personal gross income changes.

Personal Income Taxation in Australia

In Australia, the *Income Tax Assessment Act* (ITAA) determines provisions for income tax. The ITAA was first passed in 1936. Much of this Act still stands but it has been revised and expanded many times since then. It has also been supplemented, for example with the *Fringe Benefits Tax Assessment Act 1986*. Although the states have the statutory power to raise income tax, in practice they cannot do so without the agreement of the Commonwealth.

Australian personal income tax is based on both the residence of the taxpayer and the geographical source of the income. Australia asserts jurisdiction to tax residents of Australia on their worldwide income. However, Australia recognises the right of other countries to tax Australian residents for any income earned in or from their country. Under double-taxation

[2] There is a direct nexus (requited benefit) in accumulation retirement funds, but not in defined benefit schemes.

agreements, Australian residents are generally entitled to a tax credit on tax paid to another jurisdiction. For example, if an Australian resident earns $10 000 from another country and pays $2000 tax to that country and has a tax liability of $4000 on this income to the Australian government, she would usually receive a credit for the $2000 tax paid and pay the $2000 balance to the Australian government. Non-residents who earn income in Australia are also generally required to pay tax on this income to the Australian government.

Box 30.1 provides some summary data on Australian personal income tax collected. There is no separate social security levy. However, employers are required to contribute for each employee to a superannuation (retirement) fund. As discussed below, income contributed to superannuation funds is treated differently from normal income. Also, most Australian states levy a payroll tax of about 6 per cent on employee payrolls over about half a million dollars.

Assessable personal income

Assessable personal income is derived from labour, capital and government benefits.

Income from labour. This embraces monetary earnings for any services rendered, including wages and salaries, director's fees, allowances and bonuses including tips. Most lump sum payments made in lieu of income (e.g. lump sum payments for unused long service leave accrued after 1978) are also treated as income. However, redundancy payments are generally exempt. Income from labour also includes net income (gross income less expenses) from carrying on a business. This includes net income from sole trading, trusts and partnerships conducting businesses and other businesses.[3]

Individuals must also report the gross value of in-kind (fringe) benefits received from employers inclusive of the top marginal tax rate and the Medicare levy. Actually, employers pay the tax and levy. The reportable fringe benefit amount is not part of an individual's assessable income, but the net amount after tax is usually included in means-tested transfers.

Income from capital. The main forms of income from capital are interest payments from cash loans and bank accounts, dividends and capital gains from equities, and net rents and capital gains from property. It may be noted all these forms of income are nominal income uncorrected for inflation. Income from capital may accrue to individuals directly or through some investment vehicle such as partnerships or trusts including family trusts and managed funds. Few trusts or partnerships pay tax. Rather, they distribute all their income to various entities, which are then responsible for the relevant tax.

Box 30.1 Personal income tax in Australia in 2008-09

Personal income tax accounts for nearly 50 per cent of all Commonwealth tax revenue.

The Australian Taxation Office (2011) reported that 12.3 million taxpayers lodged personal income tax returns in 2008-09. Seventy-one per cent of these returns were submitted by tax agents. Taxpayers declared a total income of $539 billion. This was after net rental income losses of $6.5 billion.

Taxpayers claimed deductions of $31.7 billion including $16.4 billion in work-related expenses. In addition, 10.7 million individuals claimed tax offsets and credits totalling $24.7 billion, including $12.4 billion in refundable tax credits and offsets.

Overall, individuals had assessed taxable income of $554.2 billion and net tax payable of $115.7 billion plus $1.2 billion payable for higher education in loan repayments. These figures do not include taxes on incomes in superannuation funds ($5.7 billion) or fringe benefit taxes ($3.4 billion) that employers pay for goods-in-kind, primarily motor vehicles, provided to employees.

[3] This income is sometimes described as 'personal services income' and may include a return on capital invested in a business.

When a taxpayer has provided their tax file number to the relevant financial institution, the institution pays the gross interest with no tax withheld.

Dividends may be 'franked' or 'unfranked'. A franked dividend indicates that the company has paid tax on the income stream providing the dividends. This tax becomes a tax credit. Taxpayers who receive dividends report the gross amount (the dividends plus tax credit) as income. The tax liability is calculated on the gross amount, but the taxpayer receives the tax credit. The taxpayer pays extra tax if their marginal tax rate (MTR) is greater than the company tax rate and receives a rebate if their MTR is lower than the company tax rate.

The taxable income from property is the gross rent less all outgoings including maintenance and repairs, council rates, interest payments and depreciation of fixtures and fittings. Capital gains are taxed only when realised. As some of the return from housing is deferred until realisation, in some cases deductions may exceed gross rents, particularly when interest deductions are claimed. This is known as **negative gearing** because the interest on the property loan creates a negative net income return. Taxpayers can deduct this net loss from other taxable income. Suppose that a property investor records a net loss of $6000, has a MTR of 40 per cent and an unrealised capital gain of $10 000. The taxpayer reduces her tax liability by $2400 (= $6000 × 0.4). Before tax, the investor has a net gain of $10 000 - $6000 = $4000 over the year. After tax, her gain rises to $6400 = $10 000 – $6000 + $2400 over the year. Of course, an additional tax liability arises when the capital gain is realised.

Negative gearing
Occurs when the interest payment on a property loan creates a negative net income from property

Taxation of capital gains is complicated because many assets are exempt and because of the way that tax is estimated. For most purposes Australian tax applies to half of the nominal capital gains when certain assets are sold, including sales of shares, units in managed funds and land and buildings other than the primary home. Capital losses can be deducted from capital gains when the net gain is still positive. Otherwise losses can be carried forward until they can be offset against a gain. Capital gains tax does not apply to the family home, cars and assets for personal use or to sales of equities by retirement funds providing pensions.

In determining assessable income, the ITAA distinguishes between income that is taxable and capital payments that are not. For example, interest on capital is assessable income, but capital receipts such as a loan or loan repayment are not income. Periodical payments to replace lost income due to workers compensation are assessable income. On the other hand, a lump sum payment representing loss of earning capacity is deemed to be a capital payment and is not assessable.

Government benefits. Most income support payments are taxable. For example, the Newstart (unemployment) Allowance, the Age Pension and the Parenting Payment, as well as youth and sickness allowances are part of assessable income. However, the pensioner and beneficiary offsets remove most of the tax liabilities. Also, family assistance payments that are designed to meet specific costs, including the Family Tax Benefit Part A, Child Care Benefit and Child Care Rebate are not taxable.

In summary. Like most tax systems, the Australian income tax system is a hybrid system. Assessable income is quite comprehensive. It includes most forms of money income (wages, salaries and income-in-kind), most government transfers and some forms of income from capital (although tax treatments vary across savings vehicles). However, assessable income falls well short of the full Haig–Simons definition of income. It does not include imputed rents and capital gains from home ownership. Nominal capital gains are taxed at half the marginal rate of other forms of income and only when realised. Income earned in retirement funds before individuals retire is taxed at a concessionary (15 per cent) rate. Following retirement, the earnings of retirement funds from capital up to $1.6 million in value and pensions from the funds are tax free. Also, bequests from estates and other gifts are largely exempt from taxation. Thus, the tax base exempts various activities from tax and taxes some

forms of income from capital at lower rates, often at much lower rates, than income from labour.

Other features of Australian personal income tax

Allowable deductions. The two main kinds of allowable deductions from assessable income are work-related expenses and contributions to approved organisations. Taxpayers can also claim for the costs of managing their tax affairs.

Work-related expenses are approved expenses that contribute to earnings in an existing occupation. They include travel expenses on business, self-education expenses relating to income from current work, and various other expenses such as tools of trade, conferences, references, computers, telephone and home office expenses. They exclude journey to work expenses, self-education expenses relating to possible future income or child care expenses. Donations to eligible Australian charities, overseas aid agencies, school building funds and approved environmental and cultural organisations are also allowed as tax deductions, providing the taxpayer does not receive anything in return such as a raffle ticket or dinner.

Importantly, taxpayers can also reduce their taxable income by nominating that part of their earnings be directed into a retirement fund *before* assessable income is estimated. This is known as 'salary sacrificing'. The rules change frequently. Recently contributions have been tightened significantly. Currently individuals can salary sacrifice up to $25 000. If an individual of any age earns say $60,000 and elects to salary sacrifice $25 000, she pays income tax on only $35 000 and the pension fund pays a concessionary rate of tax (currently 15 per cent) on the income paid into the fund.

Tax rates. Table 30.1 shows the Australian tax rate structure for an Australian resident in 2017-18 along with average tax rates. The government changes these tax rates at its discretion. Australian tax rates are not indexed. That is, the income brackets do not rise automatically with inflation.

In addition, a Medicare Levy is raised at the rate of 2% on taxable incomes over $26 668. Although revenue from the levy is earmarked for health expenditure, it pays for only a small part of Medicare, and is essentially a tax on income that accrues to consolidated revenue.

Also, these rates do not include the superannuation guarantee charge. Employers must pay 9.5% of the ordinary time earnings of their employees into a complying superannuation fund or a retirement savings account. Employees here include all part-time and casual employees, who are aged over 18 and who are paid $450 or more before tax per month,

On the other hand, in 2017-18 the low-income tax offset (LITO) provides a small tax offset of up to $445 for all taxpayers with incomes of $37 000 or less. This amount is reduced by 1.5 cents for each dollar over $37 000.

Table 30.1 Australian income tax rates in 2017-18

Taxable income ($)	*Income tax*[a]	*Average tax rate (a)*
1 – 18,200	Nil	0%
18,201 – 37 000	19 cents for each $1 over $18 200	0 – 9.65%
37 001 – 87000	$3572 + 32.5 cents for each $1 over $37000	9.65 – 22.78%
87 001 – 180000	$19 822 + 37 cents for each $1 over $87000	22.78 – 30.13%
Over 180000	$54 232 + 45 cents for each $1 over $180 000	30.13% and upwards

(a) These average tax rates do not include the Medicare Levy, the superannuation guarantee charge or the low-income tax offset.

International comparisons of income tax rates. Comparisons of international income tax rates are complicated by the variety of income tax systems, tax bases and close substitutes. Some countries, for example Canada and the United States, have sub-national as well as national income tax systems. There is also a strong case for including social security schemes and the superannuation guarantee levy, and possibly even payroll taxes, in international comparisons. A further complexity is the variety of income bands and tax rates. One country may tax some income levels heavily and others lightly. Also, high-income tax rates may be offset by a narrow tax base. Thus, selective quotation of tax rates can be misleading. Another factor that complicates comparisons is the interaction of tax systems with welfare benefits. Individuals may face low nominal marginal tax rates but high implicit marginal tax rates as they lose benefits as incomes rise.

The Australian Treasury (2006) found that Australia was generally in the middle of OECD practice. For the OECD countries overall, contributions from personal income tax, social security and payroll taxes average 15 per cent of GDP. In Australia the comparable figure was 14 per cent, but this excluded the superannuation levy. In 2005–06, Australia's highest marginal tax rate (then 48.5 per cent) was eleventh highest of the 30 OECD countries examined. A marginal tax rate of just over 30 per cent on average weekly earnings was also close to the median OECD experience for such earnings inclusive of social security contributions. Again, this excluded the superannuation levy.

Tax offsets. Pattugalan and Ellis (2010) reported that at that time over 40 tax offsets existed in the Australian tax system. In addition to the LITO (then higher than now), these included:

1. The pension tax offset and the beneficiary tax offset to ensure that tax is not paid on certain transfer payments
2. Provisions for tax relief for personal circumstances. These include tax offsets for low-income taxpayers with dependent spouses or housekeepers, and for all taxpayers who live in 'remote' or 'special' areas (including Alice Springs, Darwin, Cairns, Airlie Beach and Lord Howe Island).
3. Tax offsets that provide incentives of various kinds, for example for mature age workers (over 55) to encourage workforce participation and for private health insurance.
4. Credits for a tax that has already been paid, for example credits for franked dividends and foreign tax.

Except for health insurance and tax credits, the offsets can only reduce tax paid to zero.

General Issues in Personal Income Taxation

We discuss below some issues with respect to the tax unit, the tax base, allowable deductions, the structure of tax rates, tax offsets, indexation and tax administration. Where previously discussed, the discussion below is brief.

The tax unit. The main general issues concerning the tax unit were discussed in Chapter 25. In Australia, as in most tax systems, taxation is based on the individual taxpayer. This is simpler and more transparent that taxation based on family units of different sizes. It also lowers the MTR for the second earner. Accordingly, the Henry Tax Review (2010) recommended that the tax unit remain the individual.

However, some social security benefits such as family payments, some tax offsets and the Medicare levy depend on family income, with individuals losing significant benefits as their partner's income rises. While such income grants are often regarded as part of the welfare system rather than the tax system, the distinction is somewhat arbitrary when income grants may be viewed as negative taxes. Viewing tax and welfare systems as an integrated system

has led some commentators such as Apps (2010) to describe the Australian tax–welfare system as based in effect on families rather than on individuals. Some people consider this is equitable. Apps (*ibid*) argues that this is inequitable because it creates a high effective MTR for the lower-income partner who pays tax and loses income grants as his or her income rises. It is also inefficient because second income earners generally face a high effective MTR and have a more elastic labour supply response.

The tax base. Two main issues arise with respect to the personal income tax base. One issue is the breadth of the tax base. A broad definition of taxable income is generally efficient because the wider the tax base, the lower can be marginal tax rates. It is also fair because all sources of income are treated equally. As we have seen, the Australian tax base is quite broad, with the exception of some forms of capital income. To increase the breadth and equity of the system, the Henry Tax Review (2010) proposed that all forms of work remuneration, including employer contributions to superannuation schemes and the main fringe benefits, should be included in the personal income tax base.

The second main issue is the quarantining of personal income in vehicles such as incorporated companies, trusts and superannuation funds. In 2008–09, 732 000 companies, 660 000 trusts, 419 000 partnerships and 332 000 self-managed superannuation funds submitted tax returns in Australia. These vehicles may allow individuals (often high-income individuals) to significantly reduce, or even escape, personal income tax. In 2018, most companies and superannuation funds pay tax rates of 30 per cent and 15 per cent respectively on income earned, compared with marginal tax rates of 45 per cent for high-income individuals. Individuals can also split income, usually but not only between family members, via trusts or partnerships and so utilise the lower marginal tax rates associated with lower incomes in the family. Covick (2004) and Pope (2005) argue that these arbitraging opportunities are a major breach of the principle of horizontal equity

Allowable deductions. The case for allowing work-related deductions is strong on efficiency and equity grounds. Firms and sole traders are allowed to deduct such expenses from gross income. Disallowing work-related expenses for employees would encourage excessive incorporation and be unfair. However, defining work expenses is not easy. For example, in Australia travel on business is allowed as a business expense, but commuting to work is not allowed. Other examples of grey areas include expenditure on child care, home computers and mobile phones. Some of these expenditures may be necessary to earn income, but also have a consumption element. Of course, similar issues apply in the corporate sector. In order to simplify arguments and bookkeeping, the Australian Tax Office (ATO) declared an automatic deduction (with no documentation) of $1000 from July 2012. But following a blow-out in work-related claims, the ATO now requires claims to be documented.

On the other hand, various issues arise with deductions for personal expenditures, such as private health care expenditures or interest payments for home mortgages (as allowed in the United States for example). First, the choice of type and amount of deductions is a political rather than a technical judgement. Many people might agree that some medical expenses should be subsidised, but precisely which expenses should be tax deductible involves a normative judgement. Second, deductions for personal expenditures change relative prices and so may distort expenditure patterns. Suppose that a taxpayer has a marginal tax rate of 33 per cent and spends $3000 on an allowable deduction, say a medical treatment. Their taxable income falls by $3000 and the tax liability by $1000. The after-tax cost of this expenditure is only $2000. The deductions may create deadweight losses (DWLs) as individuals consume goods that they would not purchase in the absence of a subsidy. Third, allowable deductions are not equitable. For any given deduction, income earners in higher marginal income tax brackets receive a larger reduction in tax liability. Alternatively, government could subsidise

the consumption of medical services by direct payments, which would be more transparent and could be more equitable.

Tax deductions for donations to government-recognised charities raise further issues. To exemplify these issues, suppose that someone with a MTR of 40 per cent contributes $1000 to a recognised charity. In effect, she gives $600 and the taxpayer contributes $400. This is efficient if the donor and government would have jointly contributed $1000 to this charity in the absence of the tax incentives. But, if they would have done so, the tax deduction is unnecessary. If they would not have done so, the tax deduction may be regarded as inefficient. Again, there are equity issues. Charities favoured by high-income earners (e.g. environmental causes) implicitly attract more government support for each private dollar contributed than do charitable institutions favoured by lower-income contributors (e.g. churches).

Tax structures

We discussed the optimal tax rate structure in Chapter 28. This included a discussion of the choice between a linear and a non-linear tax structure and the optimal MTRs for both systems. An important part of this discussion was the inclusion of income grants with taxes on earned income. Although the income grants were not themselves taxed, the withdrawal of the grants as earned income rose created an effective marginal tax rate on earned income. This is an important practical issue because sharp withdrawals of grants (or tapering of grants) with increases in income from other sources often creates a higher effective MTR than the official tax rates on personal income. We now note a couple of other points that are sometimes overlooked in the technical literature on optimal tax structures.

First, this literature tends to be based on a personal income tax system in which all income accrues directly to individuals. In practice, the intermediation of companies, trusts and other legal vehicles allows income splitting between family members or other persons. Where there are several MTRs, some individuals can split their income to minimise their MTRs and total income tax liability. This is inconsistent with vertical and horizontal equity. The proliferation of MTRs in Australia (and elsewhere) is also inefficient because many taxpayers choose their income and investment vehicles to maximise their after-tax income rather than their pre-tax income. The proliferation also creates high compliance costs. Reducing the number of personal income tax rates and aligning them more closely with rates in other legal vehicles, such as the corporate tax rate, would reduce these inequities and inefficiencies.

Second, and partly related to this, is the level of the tax threshold, which is the income level at which some tax starts to be paid. A high tax threshold significantly reduces the number of tax returns required and the associated compliance and administration costs. Also, it would simplify the tax–welfare benefit system. With a low threshold, large numbers of people concurrently pay tax and receive benefits. In line with this thinking, the Henry Tax Review (2010) recommended that the tax threshold should be raised radically from then $6000 to about $25 000 and that government grants not be taxed. On the other hand, a low threshold increases the revenue collected from middle- and high-income individuals and so allows for lower marginal tax rates for these income groups. It also discourages income splitting. Thus, there are some advantages in a high threshold for low-income earners and a zero threshold for high-income earners. The LITO is a modest attempt to achieve this kind of dual system.

Tax offsets. Tax offsets add additional complexity to the tax-transfer system. Thus, the Henry Tax Review (2010) recommended (1) that all structural tax offsets—LITO, senior Australians, pensioners and beneficiary tax offsets—should be incorporated into the personal income tax rate scale and (2) that all the many concessional tax offsets should be removed, rationalised or replaced by outlays as appropriate in each case.

Tax indexation
Adjusting tax thresholds and tax brackets for changes in the nominal value of money (inflation)

Indexation. When there is inflation, nominal incomes may rise with inflation but with no increase in purchasing power. When tax rates increase with the level of income and government does not adjust tax rates with inflation, individuals face higher marginal and average tax rates when their real income has not risen. This phenomenon, known as 'fiscal drag' or 'bracket creep', is often cited as a reason why taxes tend to rise as a proportion of GDP.

Unless income and asset values are adjusted for rising prices, inflation also produces a bias against income from capital. Interest earned on capital includes a real and a nominal component. Suppose that an investment of $100 is worth $107 at the end of a year and that inflation is 3 per cent. The return includes a real component of about 4 per cent and a nominal component of 3 per cent. Capital gains likewise contain real and nominal components. Clearly, governments can adjust tax rates for inflation and tax real income from capital rather than nominal income. This process is known as indexation. However, the Australian tax system, like many others, makes no such allowance for inflation. In the United States, the tax threshold, tax bracket widths and standard deductions are indexed, but there is no indexing of income from capital.

Administration costs. The personal income tax system in Australia is extraordinarily complex. The basic tax book of instructions runs to over 100 detailed A4 pages plus supplementary instructions dealing with special issues. The complexity of the tax system reflects in large part attempts to deal fairly with a very large number of special cases. But over 70 per cent of taxpayers employ tax agents and the personal and professional costs of compliance with the tax system are high.

Taxation of Labour Income

We have discussed various issues in the taxation of income from in previous chapters. Here we briefly pick up on some specific issues.

The major efficiency issue for taxation of labour income is the effect of tax rates on labour supply. This includes not only the effect of taxation on total labour hours, but the effects on acquisition of human capital, the nature of the work and the amount of effort put into work. High MTRs may reduce the return from training, encourage individuals to switch from high-value-added employment to lower value-adding activities with higher amenity and reduce the amount of effort per hour.[4]

Tax rates also affect government revenue. Although government revenue will generally rise with tax rates, at high tax rates revenue may fall with an increase in rates. This may occur not only because individuals substitute leisure for work but also because people find ways to earn income that reduce or escape payments of taxes.

As we saw in Chapter 27, the effect of personal income tax on labour supply depends on the overall labour supply elasticity that includes income and substitution effects. These effects often work in opposite directions. The income effect encourages most people to work more hours because they have a lower after-tax income and demand less leisure. The substitution effect of a tax on labour increases the price of market goods and services purchased from after-tax income relative to an unchanged price of leisure and home production which are tax free. This encourages people to do more home production and take more leisure because the relative price has fallen. Overall a change in income tax rates often has only a small effect on total labour supply, but there is a lot of variation in responses by different kinds of workers.

[4] These impacts depend on the substitution effect of higher tax rates dominating the income effect.

On the other hand, the DWL of income taxation depends only on substitution effects. As we saw in Chapter 27, assuming a linear compensated labour supply curve, the compensating variation measure of the DWL of taxation can be expressed as:

$$DWL = 0.5\, Q_1 w_1 \eta_{cs} t^2 \tag{30.3}$$

where Q_1 and w_1 are the initial amounts of labour and wages, η_{cs} is the elasticity of the compensated labour supply and t is the tax rate.

The implications may be illustrated by an example. Suppose that initial gross wage payments sum to \$100 million and the tax rate is increased from 30 to 35 per cent. Suppose also that labour hours are constant because the uncompensated labour supply curve is perfectly inelastic, but the compensated labour supply elasticity is 0.4 (reflecting the substitution effect). Tax revenue rises by \$5 million from \$30 million to \$35 million. Drawing on Equation 30.3, the DWL rises from \$1.8 million to \$2.45 million, which is an increase of \$0.65 million. The loss would be \$0.13 for every dollar of tax raised.

As reviewed in Chapter 27, these labour supply elasticities are quite representative. The ordinary labour supply elasticity of full-time males is close to zero and the compensated supply elasticity is typically about 0.3. On the other hand, the ordinary labour supply elasticity for part-time workers varies from about 0.5 to over 1.0 and the compensated supply elasticity may be some 0.3 points higher.

Given that high-income earners are usually full-time workers, high MTRs for high-income earners may have little effect on labour supply. But tax rate increases for part-time or secondary earners with high labour supply elasticities are likely to generate significant changes in work hours and deadweight losses. Accordingly, the Mirrlees Review (2011) recommended that individuals near retirement and mothers with young children, with presumably higher labour supply elasticities, should be taxed at lower marginal rates.

Taxation of Income from Capital

A tax on income from capital also has income and substitution effects. The income effect reduces both current and future consumption. This increases savings. The substitution effect shows up via an increase in the price of future consumption relative to present consumption, which encourages people to substitute present for future consumption and so reduce savings.

To assess the DWL of taxation for savers, we consider only the substitution effect holding real income constant. But, if tax rates reduce savings, the rate of interest may rise and this may create a DWL for firms employing capital. This is illustrated in Figure 30.1, where a tax on capital income shifts the supply of savings from S_1 to S_2, savings fall from Q_1 to Q_2 and the pre-tax interest rate rises from r_1 to r_2. Consequently, the marginal rate at which firms can transform present goods into future goods exceeds the marginal rate at which households are willing to exchange present for future consumption (MRT > MRS). The total DWL equals area *ABC*. Firms experience a DWL of *ABE*. Savers have a loss of area *BCE* if the supply curve is a compensated supply curve.

In econometric models, the quantity of saving (by households or nationally) is the dependent variable, and explanatory variables include the real rate of return on saving, disposable income and other potentially significant variables. If the coefficient on the rate of return is positive, it may be concluded that taxes on income from capital (which decrease the rate of return) reduce saving. A problem with these studies is that the expected real rate of return equals the nominal rate of return less the expected inflation rate, but the latter is not observed and must be based on assumptions of some kind.

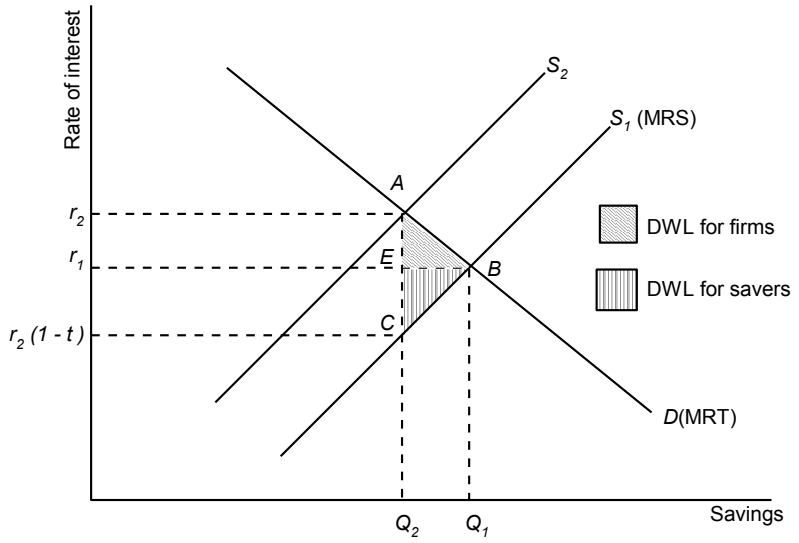


Figure 30.1 Deadweight loss of tax on income from capital

In a survey of the literature on savings, Bernheim (2002) concluded that the overall elasticity of savings with respect to interest rates is close to zero because the income and substitution effects cancel out. This implies that a tax on income from capital has little effect on borrowing costs or investment. However, a low overall elasticity of savings with respect to the after-tax interest rate may co-exist with a significant substitution effect of, say, 0.3. So there will be a DWL for savers as well as unequal treatment of inter-temporal consumption.

For and against uniform taxation of labour and capital income

Efficiency arguments for a uniform tax rate of taxation of labour and capital income are (1) that a broad definition of income with minimal tax base exemptions leads to lower MTRs and (2) that differential rates may distort the use of labour and capital inputs to production. Regarding equity, a dollar of income from either source has the same power over consumption. Also, the capacity of some individuals to choose whether to take income in the form of a payment for labour or capital would allow them to minimise their tax liability.

However, there are several complications. First, the principles of efficient taxation suggest that, to obtain a given tax revenue, DWL is minimised by relatively high taxes on factors of production in inelastic supply and low taxes on factors that are mobile and in more elastic supply. This usually implies higher tax rates on labour, land and natural resources and a low tax rate on other forms of capital. Second, taxing income from capital is effectively double taxation of savings, which distorts decisions on the timing of consumption and is unfair to savers.[5] Third, nominal returns to savings include an inflation component. Fourth, tax on internationally mobile capital reduces capital inflow in the long run. This reduces the capital–labour ratio, productivity and real wages. Thus, the tax is borne by local labour.

As we saw in Chapter 28, both the Mirrlees Review (MR, 2011) and the Henry Tax Review (HTR, 2010) recommended taxing most income from labour and capital at the same rate, with the MR favouring a more comprehensive definition of income. But MR then proposed other

[5] On the other hand many governments discriminate in favour of low labour earnings by providing earned income tax credits for low earnings but not for low income from capital (see Chapter 30).

ways to compensate for the double taxation and inflation issues associated with saving. MR suggested that only the net return to capital should be taxed after subtracting the risk-free opportunity cost of capital from the gross return. This would be a tax on super-profits (or equivalently, economic rents). HTR proposed that lifetime savings in owner-occupied housing should be free of tax, income in super funds should be taxed at 7.5 per cent and that to allow for inflation there should be a 40 per cent discount for interest, net residential rents and capital gains associated with personal savings. HTR also proposed that the corporate tax rate should be reduced gradually to 25 per cent. On the other hand, HTR recommended that there be an annual wealth tax on land, which is an immobile resource.

Differential capital income taxes

In Australia, tax rates on income from capital vary greatly by source of the income.

- At the high end of the spectrum, interest income and dividends are taxed at MTRs for individuals and include a tax on the nominal as well as the real components.
- On the other hand, only 50 per cent of capital gains on bonds or shares or most other marketable assets, including investment properties, are taxed.[6] Moreover these gains are taxed only when realised, not when accrued. Thus, asset owners gain tax-free interest income from their capital gain prior to sale.
- Moving further down the tax scale, all income accruing to individuals within retirement (superannuation) funds from whatever source is taxed at only 15 per cent and at 10 per cent for capital gains.[7] After retirement, no tax is paid on income earned on $1.6 million in assets or on any payments from retirement funds.
- Finally, homeowners pay no tax on imputed rents or on capital gains from the sale of their house. Nor do owners of consumer durables generally (including luxury boats and cars) pay tax on the imputed rents or capital gains. But housing is by far the most important source of imputed rents and capital gains.

There are various reasons for these differentials. Imputed rents are hard to measure and are taxed in very few countries (the Netherlands is an exception). And there has long been a community preference for favouring home ownership. Retirement funds are encouraged to reduce the long-term burden of retirement pensions on public finances. But, overall, *ad hoc* political decision making has resulted in an inconsistent approach to taxation of capital.

Capital income tax differentials create an inefficient allocation of capital because investors aim to maximise the after-tax return from capital rather than the gross return. This diverts capital from more productive uses to less productive ones, including from industry to housing. Also, some individuals may have better access to tax-preferred savings vehicles such as personal retirement funds or owner-occupied housing than others do, thus creating vertical and horizontal inequity. The HTR recommendation for a 40 per cent discount on returns on some higher taxed items (interest, residential rents and capital gains) would reduce these differentials.

International Issues

As we have noted, Australian residents are taxed on their worldwide income, but non-residents are taxed in Australia only on their local Australian income. This illustrates the two principles of international taxation: income can be taxed at the destination or at the source of

---

[6] Contrary to some popular opinion, 50 per cent tax on nominal capital gains is often higher than 100 per cent tax on real capital gains.

[7] This is a good deal for middle and high-income earners but makes little difference for low income earners.

International taxation: destination of income principle
Income is taxed where it is received

the income (or at both). Under the pure destination principle (also known as the residency or global principle), governments tax the worldwide income of their residents uniformly regardless of where the income has come from. The tax base here is net national income. Under the pure source of income principle (also known as the territorial system of taxation), governments tax all income earned in their territory regardless of the residence of the income recipient. In this case, the tax base is net domestic product. Residents of the home country are taxed on foreign income at source in the foreign country.

Two questions are of special interest: (1) What taxation system, if any, would achieve an efficient international allocation of resources? (2) What taxation system would maximise national income? Not surprisingly, if the answer to (2) is different to that for (1), in most cases (2) prevails. We focus mainly on allocation of capital, which is more mobile than labour, but talk briefly about taxation of labour income at the end.

International taxation: source of income principle
Income is taxed where it is earned

Efficient international taxation. Key features of any efficient economy are that investors choose investments that maximise the gross rate of return (efficient investment) and that savings are made by individuals who are most willing to forgo consumption (efficient savings). These principles of efficient investment and efficient saving apply to international economies as to national ones.

Efficient international investment can be achieved with a uniform source-based tax along with various destination (resident) taxes. When people are taxed at their residence on their total worldwide income and there are no differential source-of-income taxes, individuals will aim to invest in the territory and projects that provide the highest marginal gross return.

If source tax rates vary, to allow for equality of after-tax returns, the before-tax rates of return must vary and there is inefficient allocation of capital investment. Suppose that country *A* taxes income at source at 30 per cent and that country *B* taxes income at source at 20 per cent. International investors who seek to maximise their after-tax return may invest in *B* even when the gross return on investment is lower in *B* than in *A*. This inefficiency is avoided if all countries adopt a residence tax principle (and there are no source-based taxes).

However, if countries adopt differential residency tax rates, the after-tax return could differ between countries. In so far as tax rates affect the volume of saving, there would then be inefficient worldwide savings.

Efficient international savings can be achieved with a uniform residence tax and various source-based taxes in each country. From any investment non-residents would receive the same income as residents. Because residents everywhere invest to maximise returns after tax, after-tax returns from any investment are equal to all savers. Accordingly, international savings would be efficient, even if source tax rates vary. But international investment would not be efficient.

It follows that the most efficient international tax system would be a single uniform tax rate applied by all countries either at source or at residence, or at both (with no differential taxes). As in a closed economy, there would be a wedge between the marginal rate at which firms could transform resources into goods between periods and the marginal rate at which individuals wanted to substitute consumption of goods between periods. However, with a uniform source or residence tax, the allocation of resources between countries would be efficient and savings would be supplied by those who were most willing to save. These results are summarised in Table 30.2 overleaf.

If some countries tax at source and others at residence, but all countries adopt the same tax rate, efficiency of investment and savings can be maintained by tax credits. Residents would be fully credited for taxes paid at source.

However, problems arise when some countries tax at source and others at residence and tax rates differ. In this case efficient investment can still be achieved with full tax credits, inclusive of negative credits. To illustrate this, suppose that the income tax rate is 30 per cent

Table 30.2 Taxation conditions for efficient international investment and savings

Tax scenario		*Investment impact*	*Savings impact*
Residence tax	*Source tax*		
Uniform	None or uniform	Efficient	Efficient
Variable	None or uniform	Efficient	Not efficient
None or uniform	Uniform	Efficient	Efficient
None or uniform	Variable	Not efficient	Efficient
Variable	Variable	Not efficient	Not efficient

in country *A* and 20 per cent in country *B*. Individuals from *A* that invest in *B* would get a 20 per cent credit and pay an extra 10 per cent in country *A*. Residents of *A* would choose neutrally between investing in *A* and *B* and maximise the before-tax return. However, residents in *B* who invest in *A* would typically get only 20 per cent credit rather than 30 per cent and have an incentive to invest in *B* rather than in *A*. The allocation of world investment would be more efficient if country *B* provided an additional negative credit of 10 per cent to local residents who invested in country *A*. Capital would then flow to locations where the return is highest. However, because an individual in *B* would have more incentive to save than an individual in *A*, both the level and the allocation of world savings may be inefficient.

Fiscal externalities. Further problems arise with fiscal externalities. These arise because a change in the tax rate in one country may alter the value of the tax base in another country. Thus, a reduction of a tax rate in country *A* may cause capital or labour to leave country *B* and so reduce the tax base in *B*. Generally, this would reduce real income in *B* and therefore be a negative externality. To maintain its tax base, country *B* would have to reduce its tax rates. This would create in turn a negative externality for *A*. Although the tax bases may now be as they were before the tax changes, both countries may be applying lower income tax rates than they would wish.

Coordination of income tax rates, either formally or informally, would avoid or at least mitigate these negative externalities. Thus. cooperation may be Pareto improving.

However, countries may differ in their tax requirements and preferences, so harmonisation is not necessarily welfare enhancing for all countries. Moreover, tax coordination would be a form of cartel activity that could encourage an inefficiently high level of taxes. Tax competition encourages resources to flow to countries with lower tax structures (for any given level of public goods). There is no simple optimal policy for this issue of fiscal externalities, which also arises with international tax treatments of corporations and of consumption.

National policies. In practice, most countries apply both source and destination taxes. Also, although most countries provide tax credits up to local tax rates, few (if any) pay negative tax credits to attempt to level up uneven country tax rates. Typically, then, two main questions arise, in both cases assuming a country seeks to maximise its national income. First, how should a country treat the income of residents that has been taxed at source by a foreign country? Second, how should a country treat income earned locally by non-residents when it will be taxed in the home of the non-resident?

Regarding the foreign earnings of residents, the provision of tax credits up to the local tax rate enables residents to maximise their worldwide income after local tax. However, the provision of tax credits may not maximise the income of the destination country *inclusive* of the return to government. As discussed in the next chapter, local income is maximised by treating the foreign tax as a deduction from local taxable income rather than as a tax credit. Local residents would then direct their resources to countries where they can obtain the

highest return after payment of the foreign tax (but before payment of the local tax). This is not consistent with efficient international investment.

Turning to non-residents, in a textbook small open economy with a perfectly elastic supply of international labour and capital and no costs of movement, there would be little (if any) gain from taxing either their labour or capital income. The after-tax income required by this labour or the after-tax return required on capital is set by the international labour and capital markets respectively. In this case, local taxes are borne by the local economy. If an overseas worker requires an after-tax income of $60 000 and there is no tax, the cost to the local employer is $60 000. If the average tax rate is 25 per cent, the cost to the local employer rises to $80 000. If the employer is not willing to bear this cost the economy loses a worker whose productivity may exceed their wage cost.

However, this argument and conclusion assumes a perfect international labour market without any local economic rents. Many people come to work in Australia because they can achieve employment or income (or both) that they cannot obtain in other countries. Others come for personal or lifestyle reasons. Whenever such persons gain economic rent, their income can be taxed with no impact on the labour supply. Similar arguments apply to the cost and supply of international capital where there may also be economic rents, which is a story we take up in more detail in the next chapter.

Summary

- Personal income tax is a function of assessable personal income, allowable deductions, the tax rate structure, and tax offsets.
- Assessable personal income is based on the statutory definition of income in the relevant jurisdiction. Although statutory definitions usually include labour and capital income and some government benefits, they are invariably narrower than the economic (Haig–Simons) definition of income.
- Allowable deductions from assessable income usually include work-related expenses and some socially desired expenses such as health care expenses or contributions to charities.
- In Australia, taxable income includes most forms of monetary income including most government benefits, though not all forms of capital income. There are many tax deductions, a complex tax rate structure and differential treatments of capital income.
- Taxation issues include the choice of the tax unit, the width of the tax base, income splitting, tax rate structures, tax offsets, indexation, and differential treatment of different kinds of capital income.
- The taxation of labour income affects not only the labour supply (total hours worked) but also investment in education, occupation and location choices, and attitudes to effort and risk.
- Taxation of capital income probably has only a small impact on the amount of savings. However, differential tax rates may significantly affect the allocation of capital.
- International taxes produce further distortions. To achieve efficient international investment and savings, a uniform personal income tax rate would be applied by all countries either at source of income or at destination.
- However, national tax rates vary and most countries apply taxes at source and destination. When this is the case, national interest may imply different tax policies to internationally efficient policies.

Questions

1. It is sometimes argued that it is impossible for any tax system to achieve the following three principles simultaneously.
   - The income tax structure should incorporate rising marginal tax rates for individuals.
   - Families with equal incomes and other similar attributes should pay equal taxes.
   - The tax system should not distort family formation or labour supply.

   Is this true? Explain why or why not.
2. Are deductions for work expenses (i) definable and (ii) equitable? Give examples of possible anomalies.
3. What are the comparative merits or otherwise of (i) allowable deductions and (ii) tax offsets?
4. Is it desirable to have higher-income tax thresholds for low-income individuals and lower-income tax thresholds for medium- and high-income individuals?
5. What would indexation of personal income tax require? Is indexation desirable?
6. What are the differences between a capital gains tax on real gains and a tax on half the nominal gains? Which kind of tax would have lower deadweight costs?
7. Should income from labour and capital be taxed at the same marginal rates?
8. The Henry Tax Review proposed that income from interest, net residential rents and capital gains receive a 40 per cent discount before tax. What are the arguments for and against this recommendation?
9. What are the full labour supply effects of personal income taxation? And, how would you determine these effects? What would the deadweight losses be in each case?
10. From a national perspective what is the optimal tax treatment of (i) foreign earned income by residents and (ii) income earned by non-resident labour working in the host country? Does this conflict with efficient international investment or savings?

Further Reading

Auerbach, A.J. (2006) 'The future of capital income taxation', *Fiscal Studies*, 27, 399–420.

Australian Taxation Office, annual, *Taxation Statistics*.

Henry (Chair) Tax Review (2010) *Australia's Future Tax System*, Part Two, Volume One, pp. 3–94, 'Personal Income Tax',

Musgrave, R.A. and Musgrave, P.G. (1989) *Public Finance in Theory and Practice*, Chapters 19 and 20, McGraw-Hill, New York.

Pope, J. (2005) 'Reform of the personal income tax system in Australia', *Economic Papers*, 24, 316–331.

Rosen, H. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapters 17 and 18, McGraw-Hill, New York.

Warren, N. (2004) *Tax: Facts Fiction and Reform*, Chapters 8 and 9, Australian Tax Research Foundation, Research Study 41, Sydney.

Chapter

31

Corporate Taxation

Of course I am minimising my tax. And if anybody in this country doesn't minimise their tax, they want their heads read, because as a government, I can tell you you're not spending it that well that we should be donating extra.

Kerry Packer (businessman)

Corporate Taxation: Basic Concepts ♦ Corporate Taxation in Practice ♦ Incidence of Corporation Tax ♦ Efficiency Effects of Corporation Tax ♦ Taxation of International Corporate Income ♦ Concluding Observations

In this chapter we examine the taxation of company income. This is another form of taxation of income from capital. Typically, companies are owned by shareholders, who provide risk capital to the companies. Companies compensate shareholders out of profits. However, nearly all governments tax these profits before they are distributed to shareholders. This tax on profits is known as company income tax. Because companies are incorporated business entities, this is also often described as 'corporate taxation'.

Companies may be legally required to pay other taxes. For example, they may pay taxes on labour inputs in the form of payroll tax or social security contributions, taxes on land inputs in the form of land tax or rates, stamp duty on the purchase of property and taxes on the sale of commodities. They may also be required to pay royalties or a tax on excess profit in resource-based industries. We discussed the incidence of payroll and commodity taxes in Chapter 26 and the efficiency effects in Chapter 27. Land taxes, stamp duties and various taxes on resources are discussed as part of our examination of wealth taxes in the next chapter.

Companies are not the only form of commercial entity. Sole traders, partnerships, mutual funds and trusts also provide products with a view to making a return on investment. However, these entities must distribute all their profits notionally, if not actually, to members or persons of some kind and all profits are taxed fully in the hands of the recipients. This tax system, known as the full integration method, could be adopted for companies but it rarely is.

On the other hand, many business entities are designated as non-profit organisations. Although they may make a profit they must either retain it (as there are no shareholders) or distribute it to designated non-taxable activities. In either case they are not required to pay any tax on the surplus. The (non-tax) treatment of these entities is discussed in Chapter 32.

A further complication with corporate taxation is the variety of forms that taxation may take. These various forms affect the outcomes that may occur.

In this chapter we start by describing basic concepts in corporate taxation including the main kinds of corporate tax structure. We then describe some basic features of corporate tax systems in Australia and other OECD countries, including corporate tax rates. The third and

fourth sections discuss incidence and efficiency impacts of corporation tax. The penultimate section discusses tax treatment of international income. The final section presents some concluding observations including on the topical issue of the appropriate corporate tax rate.

Corporate Taxation: Basic Concepts

Corporate taxation, like personal income taxation, is based on statutory definitions of taxable income. Since accounting profits are often the basis of taxable company income, we start by defining accounting profit. We then examine how economic profit and taxable income may differ from accounting profit.

Accounting profit
The return on shareholder's capital (equity)

Accounting profit. Accounting profit is the return to shareholder-provided capital (equity). It is the surplus accruing to shareholders either in retained earnings or in distributed income. This surplus equals income less expenses, which are defined as follows:

Income = sales revenue + work in progress + change in inventories + realised capital gains

Expenses = costs of operating inputs + interest costs + depreciation + amortisation

Income and expenses are usually defined in accrual terms rather than by cash payments. Thus income includes the value of sales where goods have been provided but not paid for, the value of work in progress and the change in the value of stocks even without order for the stock. However, work in progress and stocks may be valued conservatively at cost rather than at possible market price, so that after allowing for costs there is no taxable income. Income also includes realised capital gains.[1]

Turning to expenses, input costs include employee compensation, purchases of energy and materials, rent and so on. Interest payments on borrowings are also deducted from income, but capital repayment is not. Nor are dividends deducted from income. From a taxation perspective this is a significant asymmetry in the treatment of capital employed.

Expenses also include depreciation on fixed capital such as buildings, plant and equipment. Depreciation is the loss of value of a physical asset during the tax period.[2] The value can be written down in various ways. The most common ways are the straight line and decreasing balance methods. The straight line method writes off an equal amount of the value of an asset each year over its expected life, which varies with the type of asset. The decreasing balance method applies a constant percentage rate of write-off each year (with higher absolute amounts in earlier years) until the last year when the residual value is written off.

In most cases, capital expenditures and gains are based on historic costs. There is often no allowance for changes in prices for capital assets. This means that real capital costs and depreciation expenses are understated (and real capital gains and real income are overstated).

Economic profit
Income in excess of the opportunity cost of all factors of production employed

Economic profit. Economic profit is a different concept from accounting profit. It is the surplus, if any, over and above the opportunity cost of all resources used. These costs include economic depreciation of plant and equipment and other capital assets employed based on current asset values and the opportunity cost of equity capital. Economic profit is sometimes known as economic rent or as pure (or super) profit. Thus, economic profit, where it exists, is generally far smaller than accounting profit. Economic profit would equal accounting profit only if a company funded all its financial requirements by debt or if it could account fully for the opportunity cost of equity in the accounts. In addition, all other costs including depreciation would be based on current economic values. These conditions are rarely met.

[1] Unrealised capital gains (or losses) may be included in balance sheets.

[2] Amortisation is a similar concept but applies to the writing down of intangible assets.

Taxable company income. This is generally closer to accounting profit than to economic profit. This is because there is rarely any allowance for the cost of equity capital. As we will see, both the Henry (2010) and Mirrlees (2011) tax reviews challenge this omission.

Taxable company income
Taxable income as determined by legislation

However, taxable company income or profit is based on statutes and is not necessarily based on accounting principles. Differences may arise in the treatment of income and expenses. For example, tax regimes may differ on the treatment of capital gains. On the expense side the accounting and tax treatment of depreciation, trading stock and losses are often quite different. For instance, taxation rules may allow a company to accelerate depreciation of assets or to deduct from income some multiple of expenditure on research and development. In estimating accounting profit, only the standard accounting allowance for depreciation or the actual amount spent on research and development would be allowed.

Attribution of international income. As with personal income, international business income may be taxed at source or at destination (residence). If income is taxed at both source and destination, there is double taxation unless the home country allows full tax credits for taxes paid at source.

In practice, most countries tax companies similarly to individuals. Resident companies are generally taxed on their worldwide business and investment income; non-resident companies are taxed on local sourced income. To offset double taxation, most countries allow credits for taxes paid in foreign countries. However, there are many variations and exceptions and the outcomes are quite untidy.

Corporate tax systems

A key issue in the design of corporation tax is the relationship to personal income tax. There are numerous possible permutations. Here we identify five main corporate tax systems.

1. **The classical corporation tax model.** Under this traditional model there are separate taxes on company income and on distributed profits in the hands of recipients. This model is employed in many countries, including the United States, Japan and several European countries. The model involves double taxation of income—profits are taxed when earned and again when paid out as dividends or received as capital gains. However, the impact could be small if either company or individual tax rates were low.
2. **Full imputation model.** There are again separate taxes on company income and on distributed profits, but in this case the recipients of dividends are granted full tax credits for any tax paid by the company. This model is employed in Australia and New Zealand, but in few other countries. In this model corporation tax is largely (though not entirely) a withholding tax.
3. **Partial imputation model.** In this case taxes are levied on company and recipient income, but the recipients receive only partial tax credits for tax paid by the company. Alternatively, dividends may be taxed at a lower rate than other personal income, as in Canada, the UK and some other European countries. However, the credits are a simple function of income received and do not reflect tax paid.
4. **A fully integrated tax system.** Under this system there is no company tax. Instead, all profits, whether distributed or not, are attributed to shareholders in relation to their holdings and taxed as shareholder income. This system applies to partnerships, trusts and managed funds. However, few countries have adopted this model because of practical difficulties with foreign shareholders and changes in shareholders over the tax year.

Tax imputation
Recipients of dividends receive full tax credits for profits tax paid by the company

Integrated tax system
All company profits, including undistributed profits, are attributed to shareholders and taxed as shareholder income

5. **A partially integrated tax system.** Here again there would be no company tax. Shareholders would be taxed on receipt of dividends and capital gains, but not on profits retained in the company. This model is noted here for completeness. Despite some practical advantages, it does not appear to have been adopted anywhere. This may reflect the potential for using companies as tax shelters.

These five tax models may create quite different outcomes. For example, the classical model involves double taxation of the return to equity capital. This may deter incorporation and equity finance. On the other hand, it may encourage retained earnings if capital gains attract less tax than dividends. Table 31.1 outlines some major possible impacts with respect to incorporation, use of debt or equity finance, the double taxation of savings and policies towards retained earnings.

Much of the technical literature focuses on the effects of the classical model. But many commentators (such as Meade, 1978) have questioned why companies should be treated differently from unincorporated businesses and taxed separately and, in effect, advocated the full integration model. We discuss these issues below.

Corporate Taxation in Practice

Company tax has many features. Here we discuss the tax base, the relationship between corporation and personal income tax, full and effective company tax rates and the treatment of international income. Most examples are Australian (see Box 31.1), but international comparisons are provided.

Table 31.1 Some implications of five company tax systems

	Form of company tax	*Impact on incorporation*	*Impact on form of finance*	*Single or double taxation*	*Impact on retained earnings*
1.	Classical double taxation: company tax + tax on dividends	Discriminates against incorporation	Discriminates against equity finance; bias towards debt	Double taxation of dividends	Favours retained earnings if capital gains taxed at low rate
2.	Full imputation model: company tax + taxes on dividends + full tax credits	Neutral with respect to incorporation	Neutral with respect to debt and equity for local capital but not for overseas capital	Largely avoids double taxation except for capital gains	May favour retained earnings or distribution depending on relative tax rates
3.	Partial imputation model: company tax + taxes on dividends + partial tax credits	Retains some bias against incorporation	Retains some bias against equity finance	Retains some double taxation, depending on size of tax credits	Favours retained earnings if capital gains taxed at low rate
4.	Fully integrated system: all profits including retained earnings taxed as if in hands of shareholders	Neutral with respect to incorporation	Neutral with respect to debt and equity	Fully avoids double taxation, but gets complicated with overseas shareholders	Neutral between retained earnings and distribution unless there is also a capital gains tax
5.	Partial integrated model: taxes on dividends and realised capital gains but not on retained profits	Favours incorporation	Biased in favour of equity due to deferred tax on capital gains	Avoids double taxation, but may create income tax shelters	Favours retained earnings due to tax deferment and if capital gains taxed at low rate

Box 31.1 Company income taxation in Australia

Australia's company income tax system has two main roles. It acts as a withholding tax on the income of Australian residents received from an Australian resident company. And it taxes the local source income of foreigners earned from an Australian company or the Australian branch of a foreign company.

In 2009–10, company income tax revenue in Australia was about $61 billion compared with personal income tax of $125 billion. As a proportion of total income tax, company income tax was around 30 per cent in the early 1970s, fell to about 15 per cent in the mid-1980s and has risen back to about 30 per cent. In 2009–10, company income tax accounted for 18 per cent of all tax revenue.

Factors that affect company income tax revenue are the number of companies, the size of company profits, and the company tax rate along with any tax concessions, for example for depreciation rates. There has been substantial growth in the number of companies and in company profits. Between 1990–91 and 2008–09, the number of companies in Australia more than doubled from 331 000 to 732 000. Over the same period company profits (gross operating surplus) rose from about 23 per cent of GDP to about 28 per cent.

On the other hand, the corporate tax rate was greatly reduced from a peak rate of 49 per cent in the late 1980s to 30 per cent in 2001–02, which remains the standard rate in 2017-18. Companies with a turnover rate of less than $25 million are taxed at 27.5 per cent of taxable profits in 2017-18.

A key feature of the Australian tax system is full dividend imputation (model 2 in Table 31.1) which was introduced in 1987. The Australian scheme eliminates double taxation of dividends. It also helps to prevent individuals deferring tax by taking some, or all, of their income through a company structure and not paying dividends, because the company must pay tax on profits whether dividends are declared or not. Individuals on marginal tax rates higher than the company tax rate may defer some of their tax liability but not all of it.

However, franking (tax) credits can be used only by Australian taxpayers. Foreigners (individuals or companies) holding Australian shares but paying tax in their own country obtain no benefit from franking credits. This makes shares in Australian companies more valuable to Australians than to foreigners and favours local ownership of Australian companies. It also resulted initially in a market for franking credits. Shortly before a franked dividend was due to be paid, an Australian taxpayer would buy the shares from a foreign owner while simultaneously agreeing to resell the shares after the dividend had been paid at a pre-fixed price. The Australian taxpayer would get the benefit of the franking credits while the foreign owner of the shares would share that benefit by repurchasing the shares at a lower price than they had been sold for. The Australian government deterred this trade by stipulating a minimum time that shares must be held before franking credits could be utilised.

Corporate tax burden. Company taxes comprise over 5 per cent of GDP in Australia compared with an OECD average of a little over 3 per cent. However, comparisons of total tax revenues must be treated cautiously for various reasons. Australia's imputation credit system means that a significant proportion of company tax revenue represents withholding tax payments in respect of individuals' income tax liabilities. Australian company tax revenue includes taxes on contributions to, and earnings in superannuation funds. It also includes a tax on economic profits of some petroleum projects which may be taxed via royalties or excise in other countries. Another factor is that Australia has a relatively high level of incorporation

Corporate tax base. Most OECD countries treat taxable company profits as the corporate income tax base. These profits include a normal return to equity plus any economic profit. The Australian Treasury (2006) reported that two (unnamed) countries trialled taxing just economic profit but did not continue and that Belgium has a version of this. Currently, Austria, Brazil and Italy have variants. Tax can be levied on economic profit by deducting from total profit a notional cost of equity capital. However, this is complicated because the cost of equity varies by company and the appropriate rate of return on capital depends on how losses are treated. Accordingly, few countries have adopted economic profit as a tax base.

The tax base is also affected by statutory rulings. Chief among them is the treatment of depreciation. Allowable short asset lives, high general rates of depreciation and concessional treatments for particular industries reduce the tax base. As reported by the Australian

Treasury (*ibid*.), the present (discounted) value of depreciation allowances in OECD countries is typically about 70 per cent of the purchase price of capital assets. The relevant figure for Australia is 66 per cent, which implies that Australia has a relatively high tax base. The treatment of research and development expenditures as current rather than capital expenses, industry concessions and allowances for small companies can also affect the tax base.

Company and personal income tax. Among OECD countries, the classical model remains the most common tax one (Warren, 2004). Almost as many OECD countries adopt a partial imputation model by making partial allowance in the personal income tax regimes for company taxes prepaid on dividends. Australia and New Zealand are the only countries with a full imputation tax system. Because of these differences, to make meaningful comparisons of the burden of company taxes on individuals in different countries, allowances need to be made for the double taxation that occurs in the classical and partial imputation tax systems.

The Australian imputation system works through use of franked dividends. When a company pays dividends to shareholders from profits on which the full Australian company tax rate of 30 per cent has been paid, the dividends are described as "fully franked".[3] The tax paid on the dividends is treated as though it had been paid by the shareholder. When no company tax has been paid the dividends are described as "unfranked". When a shareholder receives a dividend, she receives a statement that shows the gross dividend paid (broken into franked and unfranked parts) and the amount of tax that has been paid on the franked dividend (the imputed tax credit). On their individual tax return the shareholder declares as income the gross dividend plus the tax credit. This income creates a tax liability at the shareholder's marginal rate of tax.

Once the total tax liability has been calculated the shareholder's franking credits are deducted. If the shareholder's MTR is the same as the company tax rate the dividend is tax free to the shareholder, the tax having been prepaid by the company. If the shareholder's MTR is higher than the company tax there will be additional tax to pay at a rate equal to the shareholder's MTR less the company tax rate. If the shareholder's MTR is lower than the company tax rate she receives a tax rebate. In essence, dividends paid to shareholders are taxed once at each shareholder's marginal tax rate.

When a company retains after-tax profits rather than distributing them, there is no tax liability or credit for individual shareholders. Retention of profits should increase share values as net tangible assets have risen. If shareholders sell their shares they gain this increase in asset value less the capital gains tax which they pay on 50 per cent of the gains at their marginal tax rate (if the shares are held for over 12 months).

Statutory and effective company tax rates. Statutory tax rates have fallen in most OECD countries over the last 30 years. The unweighted OECD average corporate tax rate fell from 48 per cent in 1982 to 28 per cent in 2008 (Henry Tax Review, HTR, 2010). Over the same period, the statutory corporate tax rate in Australia fell from 47 per cent to 30 per cent, which remains the rate today except for small companies (see Box 31.1).

HTR also noted that the Australian corporate tax rate in 2009 was around 5 percentage points higher than the average for open small to medium OECD economies. Indeed, in this group Australia was the third highest behind Belgium (which has a narrower tax base) and Canada, which was moving to a lower rate.

More changes are now likely as, in 2018, the United States passed legislation to slash their corporate tax rate from 35 per cent to 20 per cent.

However, comparisons of statutory rates must be made cautiously for various reasons. First, tax bases vary with differences in the treatment of depreciation, foreign sourced income and

[3] No imputation credit is allowed for tax paid in other countries.

capital gains and with specific industry concessions. High tax rates may be offset by narrow tax bases. Second, some countries tax small companies at lower concessionary rates. Third, tax concessions may also materially affect actual tax rates. Markle and Shackleford (2009) estimated that, after allowance for all concessions, especially deductions of capital expenditure, the average effective tax rate for all Australian industries was only 22 per cent. Indeed, the estimated effective rate for mining was only 17 per cent.

These issues have led some analysts to attempt to estimate alternative effective company tax rates. A simple aggregate measure is total corporation tax revenue as a proportion of gross operating surplus (broadly company profits plus depreciation) as recorded in the national accounts. In Australia, this proportion has risen over the last 30 years, despite the fall in the corporate tax rate. However, this does not allow for tax credits available due to imputation.

Treatment of international income. As we have noted, most countries tax resident companies on their world-wide business and investment income and non-resident companies on local sourced income. However, there are significant differences between countries. Some countries (including the US and traditionally the UK and Japan) tax all foreign source income. Other countries (such as Australia, the Netherlands and Switzerland) provide exemptions for some foreign-sourced income depending on the type of income exempting active business income, the country the income is sourced from, the type of non-resident entity that earns the foreign income on behalf of the resident company and the degree to which it has been taxed. Also, most countries allow credits for taxes paid in foreign countries, but the credits depend on the type and place of the income and there may be special provision for losses (see Australian Treasury, 2006).

Australia taxes overseas earnings of Australian companies relatively lightly. The government taxes only passive foreign investment income earned by resident companies. It does not tax foreign active business income or capital gains from sales of active foreign businesses. Resident companies can also claim losses from tax-exempt income against domestic income. Australia allows foreign tax credits capped at Australian tax rates and quarantined to one active class of income and three passive classes. Excess credits can be carried forward.

Non-resident companies in Australia are taxed on their local income. To try to minimise evasion of local taxes, the government has an active oversight of transfer prices.

Incidence of Corporation Tax

In considering who really bears the corporation tax, three introductory observations may be made. First, all taxes are borne ultimately by individuals, not by companies. Of course, in the first instance shareholders bear the tax on company profits. In the absence of the corporation tax, companies could distribute gross rather than net profits. However, as we will see, depending on the tax system, the corporation tax may also be borne by all owners of capital, including owners of unincorporated entities. Further, in an open economy the tax may be borne by labour in the form of lower wages.

Second, corporation tax may be designed in many ways with different tax bases and imputation systems. We examine below the incidence of the tax in a full classical system and a full imputation system. A partial imputation system would contain some of the effects of a full imputation system. We also consider briefly the incidence of a tax on economic profits rather than on the return on equity capital.

Third, the incidence of any tax depends on the economic environment. This includes demand and supply elasticities, the competitiveness of the market environment and whether the economy is closed or open. Incidence may also vary in the short and long run as these market factors vary.

Incidence with a classical corporation tax system. A key feature of the classical tax system is that profits of incorporated entities are taxed twice whereas profits of unincorporated firms are taxed only once in the hands of the recipients. In a classic article Harberger (1962) used a simple general equilibrium model to show the incidence in this system.[4] He assumed two economic sectors, a corporate and a non-corporate sector. Each sector produced a different product under conditions of perfect competition in a closed economy. Importantly, because the economy was closed, the supply of labour and capital was fixed. A representative consumer was assumed to purchase output from both sectors.

Before the corporation tax is imposed, equilibrium requires that the gross rate of return on capital is the same in both sectors. After imposition of the tax, the after-tax return on capital in the corporate sector falls below the return on capital in the non-corporate sector. In the short run, with the capital in each sector fixed, a tax on company income reduces after-tax returns to shareholders who bear the full burden of the tax. The higher return in the non-corporate sector then causes capital to shift from the corporate to the non-corporate sector. This raises the after-tax return in the corporate sector and lowers the return in the non-corporate sector. Labour moves in the opposite direction from the non-corporate to the corporate sector.

When the system is again in equilibrium, output of the corporate sector will be lower and prices higher than in the pre-tax situation. On the other hand, the non-corporate sector will be larger and prices of that product lower. The after-tax return on capital will be equal in each sector and lower than before the tax was levied. In this model the tax burden falls generally on capital, but not just on shareholders in the corporate sector. The tax burden on owners of capital is greatest when firms in the corporate sector can easily substitute labour for capital but firms cannot easily substitute capital for labour in the non-corporate sector.

An important assumption of this model is the closed economy and, related to this, a fixed amount of capital. This explains why the rate of return on capital in the non-corporate sector falls to match the lower after-tax rate of return in the corporate sector.

In an open economy such as Australia with high mobility of capital, the long-run supply of capital is not fixed but is quite responsive (elastic) at the internationally competitive after-tax return on capital. In this case an increase in the corporation tax will initially reduce the after-tax rate of return in the corporate sector. Capital will exit the country (or not enter). In a perfect market, the decline in capital will cause the gross return on capital to rise until the after-tax rate of return is consistent with the internationally required after-tax rate of return. The fall in the capital/labour ratio will reduce output and labour earnings. Consumer prices may also increase but only in a limited way in an open economy.

Incidence with a full imputation tax system. With a full imputation system, shareholders in a corporation pay the same tax on distributed profits as do owners of capital in unincorporated entities. In both cases recipients of dividends or other payments from profits are taxed at their marginal tax rates. The corporation tax is a withholding tax. A change in the corporate tax rate has no effect on the final after-tax income of domestic shareholders. An increase in the tax rate would reduce the dividends paid out and increase the amount withheld in tax but increase the tax credits by an equivalent amount. In this case the income from companies is simply another form of income from capital and the incidence of the tax is similar to the incidence of any tax on income from capital. The tax is borne by the owners of the capital if the supply is completely inelastic but may be passed on to users of capital in the form of higher interest rates if the supply of capital is elastic.

The analysis for retained earnings is slightly different. An increase in the corporate tax rate will reduce retained earnings, but there is no offsetting tax credit to the shareholder. In the

[4] This model has a similar structure to the general equilibrium model described in Chapter 26.

short run, shareholders bear the burden of the increase in the tax rate. However, if shareholders are significantly disadvantaged, capital will flow to the local unincorporated sector or overseas, and the same kind of effects that we observed above will occur in this case, although on a reduced scale.

Incidence with taxation of economic profits. A tax on economic profits (or rents) is like a lump sum tax—it does not affect the use of resources. Remember that economic profit is a return over and above the opportunity costs of resources. This assumes that economic profit can be determined and that no one is responsible for it.

If economic profit is part of the return to an entrepreneur who is a risk-taking shareholder, then a tax on economic profit is a tax on entrepreneurs. The tax may also be passed on to the community at large if it discourages entrepreneurial behaviour. The Australian government changed the capital gains tax in 2000 from a tax on real capital gains to a 50 per cent tax on nominal gains because the venture capital industry claimed that the tax on real gains discouraged entrepreneurs and venture capital. On the other hand, if economic profit (or rent) is an act of God and not the responsibility of business managers or entrepreneurs, the burden of a tax on it will fall simply on the shareholders in enterprises that are blessed with the economic rent and have no passed-on impact.

Efficiency Effects of Corporation Tax

As we have observed, a tax on economic profit is perfectly efficient. If a company could purchase an asset which produced an expected rate of return of p, the company would purchase the asset if the cost of borrowing r was lower than p. Thus the condition for investing is $(p - r) > 0$. If there is a corporate tax rate of θ on net income, the after-tax profitability of the investment will be $(1-\theta)(p - r)$. The firm will invest if $(1-\theta)(p - r) > 0$. But if the project passes the before-tax condition, it also meets the after-tax condition. The company will undertake any investment where the rate of return is higher than the cost of capital, which implies an economic profit. This form of corporate tax would provide no disincentive to investment. In addition, this reduces the cost of equity finance and eliminates any bias towards debt finance in preference to equity

Citing these efficiency arguments, the Mirrlees Review (2011) of taxation recommended the introduction of an Allowance for Corporate Equity (ACE) within the UK corporate income tax. The ACE would be a deduction of the normal rate of return on capital representing the cost of using equity finance. This would be similar to deducting the cost of interest payments on debt finance. As the Mirrlees Review (Chapter 20) remarked, 'Exempting the normal rate of return on capital from corporate taxation fits well with our proposal to exempt the normal rate of return on capital invested in the business sector from personal taxation', which we discussed in Chapter 30.

This model would be a proxy tax on economic profit. It could under-estimate the cost of capital when there was any significant business risk and might not allow for the true costs of various non-capital inputs, especially entrepreneurship. HTR (2010) also considered that ACE had merits, as well as practical complications, and recommended that it should be considered in due course. HTR viewed it as a lower priority than reducing the corporate tax rate.

Another model with significant efficiency features is the full integration model. In this case the total taxable income of both corporations and unincorporated entities is taxed in the hands of the owners and there is no separate corporate tax. This treats all capital income on a level playing field, avoids differentials between corporate and personal tax rates and minimises economic distortions. There remain some general issues with respect to the taxation of income from capital which we discussed in the previous chapter, but not issues due specifically to the taxation of company income.

Nevertheless, as far as we are aware, no country has adopted a full integration model. Accordingly, we examine below the efficiency effects of the main model employed, namely the classical dual-tax model, and the Australian imputation model. We also discuss some issues in relation to the treatment of depreciation.

Impacts of the classical tax system and the full imputation model

The key feature of the classical system is the double taxation of company income. This may be mitigated by taxing dividends at a lower rate than other income of individuals, but an element of double taxation remains. Providers of equity capital are taxed first at company level and then on receipt of income, but providers of debt are taxed only once on receipt of income. In the full imputation system, providers of equity capital are taxed only once because they receive a tax credit for taxes paid by the company. We consider below the impacts of the tax arrangements on company structures, sources of finance, the amount of investment, and the production of goods and the distribution of profits (payout policies).

Corporate structures. Businesses incorporate for several reasons. These include legal limitations on personal liability for losses, access to more capital and liquidity and flexibility in capital arrangements. For small businesses, the protection of limited liability may be a key factor. However, owners are often required to provide personal guarantees for the company's debts, which negate any benefit of limited liability. For larger businesses, access to larger capital markets may substantially increase the value of the business or enable the business to expand with the use of risk capital that would not have been possible otherwise.

Nevertheless, the high cost of equity capital in the classical system may deter incorporation in marginal cases. Where this distorts preferred corporate structure, there is a deadweight loss (DWL). On the other hand, with full imputation there is no tax disincentive to incorporation. In most important matters the effective tax burden on owners of capital is similar in incorporated entities as in unincorporated entities.

Corporate finance. *Prima facie*, the high cost of equity in the classical system is likely to distort corporate financing away from equity and towards loans. Of course, companies will still employ equity capital for various reasons. One is that use of equity reduces corporate risk. Debt involves risk because interest on the debt must be met. A firm that cannot meet its interest payments faces bankruptcy or loss of shareholder and management control to the creditors. Dividends are paid only when there is a profit. Second, lenders are often unwilling to provide finance to firms that use little capital of their own. In any case, lenders will charge a higher rate of interest to firms that are highly geared (firms with a high level of debt to total assets). Nevertheless, the high cost of equity capital in the classical system may deter use of it in marginal cases and, again, this may create a deadweight loss.

On the other hand, there is no bias associated with equity finance in the full imputation system. To illustrate, suppose that a project requires $1000 in capital, that the cost of debt and equity is the same at 6 per cent and that the project is expected to produce a pre-tax operating profit before interest and tax of $100 a year. The company can deduct interest payments as a cost, but not deduct dividends. The corporate tax rate (t) is 30 per cent. Given full imputation, it is easy to show that financing arrangements do not affect the level of after-tax profits.

Consider first 100 per cent debt financing. The interest charge is $60 per annum. Pre-tax company profit would be $40, tax would be $12 and after-tax profit would be $28. Now consider 100 per cent equity funding. There would be no interest charge. Company profit would be $70 after tax. However, the use of company (or shareholder) equity of $1000 has a real cost—the amount of income forgone after tax. This equals ($1000 $\times$ 0.06) $(1 - t)$ = $42. Subtracting $42 from $70, we find that the after-tax profit is again $28. There is no preferential treatment of debt (or equity).

Investment and output. Following the Harberger analysis above, by raising the cost of capital, the classical corporation tax can affect output even in a closed economy. If the amount of capital is fixed, the tax will cause investment and output to shift from the corporate to the non-corporate sector. Efficiency requires that all firms face the same marginal costs. Because the corporate sector faces a higher cost of capital, the tax distorts the pattern of output towards less efficient operations at the margin. In an open economy, the supply of capital and consequently output will fall. These distortions also create a deadweight loss.

The distribution of profits. The double-taxation classical taxation system encourages firms to retain profits for two reasons. First, retention defers some tax payments. Second, in most tax regimes capital gains are taxed more lightly than dividends. Such a bias towards retention of profits distorts capital markets. Firms are encouraged to invest in their own business rather than allowing owners of capital to determine how capital is used most productively.

In the imputation system the incentives to retain or distribute profits depend principally on the interaction between company and personal income taxes.[5] When the corporation tax exceeds most individual MTRs, there is an incentive to distribute profits so that the recipients can receive tax credits and gain a refund from government for excess tax paid. When the corporation tax is less than individual MTRs, there is an incentive to retain profits in the company. The owners can use undistributed profits to build up wealth taxed at a relatively low tax rate. There is an additional incentive to retention when retained earnings can be converted into capital gains and taxed at a relatively low capital gains tax rate.

In Australia both these phenomena are observable. The weight of money in superannuation funds where income is taxed mainly at 15 per cent, which is half the current corporate tax rate, encourages distribution of profits. On the other hand, MTRs on personal income over 45 per cent encourage high-income company owners to retain savings in corporate vehicles that are taxed at a much lower rate.

Depreciation and investment

Company tax law has another impact on the cost of capital. The law generally allows firms to depreciate over time the historical cost of capital assets. Firms can subtract from income, before calculating their tax liability, an amount which is supposed to represent the loss of value of the capital asset due to production during the tax period. Depreciation allowances reduce the cost of capital through tax savings.

Because firms can depreciate the whole value of a capital asset, it may appear that this arrangement has no impact on investment. That is incorrect. The tax savings generated by depreciation allowances are received in the years after the asset is purchased. The value of a dollar depends on when it is received. A dollar received one year from now is worth $\$1/(1 + r)$ where r is the real rate of interest. A dollar received in 10 years is worth $\$1/(1 + r)^{10}$. If V is the present value of all tax savings from depreciation allowances on \$1 of capital expenditure, the effective non-deductible cost of capital is $(1 - V)K$, where K is the purchase cost of the asset. As we have seen, in Australia the average value of V (the discounted value of the purchase price of assets) is about 0.66. This implies that capital cannot be fully expensed and use of capital is discouraged. A further problem occurs when prices are rising because the depreciation allowances will not be sufficient to replace the capital used up. To replace the capital, firms need to include in their pricing structure an amount equal to the extra cost of replacing capital due to inflation. But while the firm may view the additional revenue as cost recovery, tax officials generally regard this as additional profit. To achieve full cost recovery and allow for inflation and tax, the firm would have to lift its price by $\pi/(1 - t)$ where π is the

[5] Under imputation, retained earnings are double taxed first as profits and then as capital gains.

rate of inflation and t is the company tax rate. If the firm cannot do that, there will be a disincentive to invest.

Government can offset these disincentives to invest in various ways.

- Accelerated depreciation allows firms to bring forward the tax savings.
- Investment allowances enable a firm to write off immediately part of the cost of a capital asset and then depreciate the full cost of the asset in the normal way.
- Tax credits allow a firm to claim a credit against income tax for part of the asset value.
- Depreciation schedules that allow investors to write off more than 100 per cent of asset value. In Australia in the 1980s investors could write off 150 per cent of the cost of producing films; in the 1990s firms could write off 150 per cent of money spent on research and development.

Each of these instruments increases the present value of the tax savings from depreciation and has been employed at various times in Australia. However, there are differences. For example, the financial benefit to the firm of an investment allowance depends on the rate of company tax, whereas the value of the tax credit is independent of the rate of tax. In both cases the firm can write off over 100 per cent of the purchase cost of the assets against tax. Depending on the size of the upfront benefit and the subsequent depreciation rate, either scheme could result in the present value of the deductions from income being greater than the cost of the asset.

Evidently, tax allowances may create positive incentives to invest. This may correct a market failure such as spill-over benefits that result in too little investment of some kind. However, such allowances may distort the allocation of capital and investments with a low gross rate of return may displace investments with higher returns. Also, the incentives can be expensive relative to the results achieved.

Further problems arise when a capital-intensive firm cannot benefit fully from depreciation allowances and tax credits because of low earnings. Examples include cyclical industries such as automobile production and industries with a long lead time between investment and profitable operation, such as construction of toll roads. Even if companies can carry losses forward to offset future tax liabilities, it may take a long time for discounted future benefits to offset the present costs. This creates a disincentive to undertake such investments.

In such a situation, a company may lease capital equipment from a firm with a more stable profit flow. The latter can take advantage of the depreciation allowances and tax credits and pass some of those benefits to the disadvantaged company through lower rental payments. Some tax regimes refuse tax deductibility for such leasing arrangements unless there are reasons other than the tax benefits for structuring the investment in such a way. Generally, the tax department allows such schemes if the company purchasing the equipment carries some risk and is not indemnified totally by the leasing company.

An alternative would be for a disadvantaged company, such as one in a cyclical industry, to merge with a company with a more stable, albeit lower, profit flow. Such tax-induced mergers make little economic sense because there are few synergies between the two companies in vastly different industries. The financial benefits are generated by the tax regime. The distortions do not occur if depreciation is based on economic income.

Taxation of International Corporate Income

We discussed efficiency principles for taxation of international personal income in Chapter 30. Similar principles apply to corporate income because companies are basically conduits for capital supplied and received by individuals. Rather than revisit how taxation affects efficient worldwide investment and saving, we discuss here the two main practical problems facing

governments, namely how should government tax (1) foreign-sourced income of a resident company and (2) local-sourced income of a non-resident company?

Foreign income of resident companies. Countries maximise their national income when the marginal after-tax return on capital invested in a foreign country equals the marginal gross return to capital invested at home. Suppose that a local firm can make a $10 million investment and annual profit of $1 million at home or abroad. Suppose also that the corporate tax rate is 30 per cent in the home country and 20 per cent in the foreign country. From a national perspective, investment by the firm at home yields a total return of $1.0 million, but investment in the foreign country yields a return of only $0.8 million. However, if the firm pays only the foreign tax on its foreign income it would have an incentive to invest overseas.

This problem is mitigated if the home country taxes the income of the local firm at the 30 per cent rate and gives a tax credit for foreign tax paid. But this is still not optimal for the home country. The home country maximises its national income by allowing a deduction for tax paid in the foreign country rather than a tax credit. In effect foreign taxes are treated like other costs borne in production. From the home country viewpoint, a firm's allocation of investment is efficient when:

$$(1 - t_h)\, p_h = (1 - t_h)\,(1 - t_f)\, p_f \qquad (31.1)$$

where t is again the tax rate, p is the pre-tax rate of return and subscripts h and f denote home and foreign countries respectively. The home country maximises its national income when the after-tax return to capital in the foreign country equals the gross return to capital at home.

Few countries treat foreign income in this way. However, Australia effectively treats foreign earnings in this way. Australia does not tax active foreign business earnings. However, these earnings (after foreign tax) are taxed in the hand of shareholders when distributed and they attract no imputation benefits (because local company tax has not been paid on these profits).

Some companies with foreign investments find the situation frustrating and have attempted to stream dividends. Australian shareholders would be paid dividends from profits earned in Australia on which Australian tax has been paid and are thus eligible for franking. Foreign shareholders would be paid from profits earned abroad which were not eligible for Australian franking credits. But this is possible only if there are two separate corporations, one owned by Australians owning the Australian assets and the other owned by foreigners owning the foreign assets. An Australian building materials group tried to organise its affairs in that manner in 1998. But it failed to satisfy either its Australian or its foreign shareholders and the scheme was abandoned. However, subsequently some companies have sought to deal with the non-neutral treatment of income from different sources by dual listing structures.

Local income of non-resident companies. In a small open economy, the equilibrium after-tax interest rate for the local economy is generally set by the world capital market. This rate is set independently of the local corporate tax rate. If the local economy is a price taker, there is a deadweight loss from taxing imported foreign capital. Suppose, as in Figure 31.1 overleaf, that the supply of foreign capital in the absence of any tax on it is perfectly elastic at price r and that the local corporate tax raises the price to $r\,(1 + t)$. Total domestic investment falls from Q_4 to Q_3. The DWL equals areas $(ABC) + (DEF)$. Area ABC is the loss due to the increased cost of local capital compared with foreign capital. Area DEF is the loss due to the fall in supply of foreign capital. When output prices are also fixed by world conditions, the higher price of capital cannot be passed on to consumers. The tax on the income of non-resident companies is then borne by labour or other fixed domestic factors.

By contrast a tax on labour would reduce the net wage rate but not distort marginal return to capital invested at home or abroad. As discussed in Chapter 30, in a well-informed world

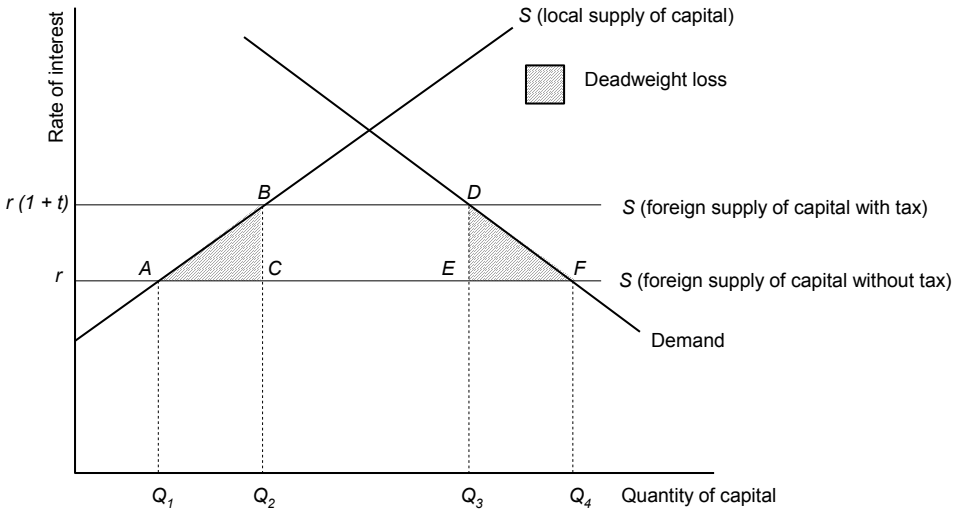


Figure 31.1 Effect of a local source tax on foreign capital

of investors, investment is more efficient with residence-based taxes of capital than with source-based taxes of capital. A labour income tax dominates a corporate income tax even for labour. It follows that foreign investors in a small open economy should not be taxed as domestic workers would bear the burden of the tax.

Nevertheless, most countries impose source-based taxes on returns to capital for several reasons. The first is the existence of sunk capital. Ex-post, fixed capital cannot readily escape. A country would not benefit from exempting this capital from a local-source tax. Second, many firms invest in a country for numerous reasons and with the intention of making real economic profits. A source tax on their accounting profits will not deter such investment. Third, source taxation benefits the local treasury at the expense of foreign treasuries when non-residents receive credits for local tax paid. Fourth, there are practical difficulties in distinguishing between profits accruing to local capital and to foreign capital.

Taxing multinationals. Finally, brief mention may be made of some practical difficulties in taxing multinational companies. The key problem is determining a realistic account of taxable income. Multinationals typically aim to maximise the proportion of their total profits in low-tax countries. They can do this in many ways. The most common methods are via artificial transfer prices and financing arrangements.

Transfer prices are charges for intra-company transactions, including charges for services and intellectual property rights. Most tax authorities require multinationals to use prices that would apply between two unrelated companies. This works reasonably well where markets are well established. However, when goods or services are unique or rarely traded, it is harder to establish fair prices. Also, the volume of sales involved would affect price as would the existence or absence of a long-term contract.

Second, because interest on borrowings is deductible before tax is calculated while dividends are not, a multinational investing in a foreign country can reduce its tax liability in that country by channelling some of the investment funds as loans, rather than share capital. Governments may respond by imposing a limit on the use of debt finance rather than equity capital or by limiting the amount of interest payments that can be deducted for tax calculations. However, this may discourage foreign investment. Thus, multinationals often have considerable discretion about how much tax they pay in different countries.

Concluding Observations

Most countries employ the classical, dual tax, system of corporate taxation, which taxes both company profits and dividends, albeit sometimes with reduced tax rates on dividends. This double taxation of equity capital increases the cost of capital to companies, biases capital financing towards debt and discourages incorporation. Arguments for it include revenue security, the idea that incorporation provides the privilege of limited liability for owners of businesses and the view (not necessarily true) that shareholders are relatively well-off. Although the distortions of double taxation appear to outweigh the benefits, only Australia and New Zealand have adopted full imputation and no leading economy has adopted a full integration strategy where all profits are taxed once as if in the hands of shareholders.

As we have seen, the Mirrlees Review (2011) recommended that companies operating in the UK should be allowed to deduct a normal rate of return on capital before tax, which would allow companies to treat equity in a similar way to debt. The tax base would approximate economic profit. This would neutralise most of the distortions noted above.

The Henry Tax Review (HTR, 2010) recommended that Australia move in the long run towards a cash-flow expenditure tax which would give an allowance for corporate equity. This expenditure tax as outlined would allow immediate 100 per cent deductions for capital expenditure which would significantly reduce the cost of capital investment. However, the Review discounted making such a radical structural change to the corporate tax system without considering the risks for revenue and the difficulties of making such a change unless other countries adopted a similar system.

HTR also considered the Australian corporate tax rate should be reduced from 30 per cent towards 25 per cent and that dividend imputation should be retained because of its neutrality towards financing and entity choices. The recommended lower corporate rate was based largely on the view that this would attract significantly more international capital. This has become more significant as the US Government legislated in 2018 to dramatically reduce the US corporate tax rate from 35 per cent to 20 per cent. The Review also suggested that there could be a shift to a partial integration system with company income tax reduced further but more limited relief offered to dividends. This appears to have been to offset the revenue loss from a lower corporate tax rate effective on foreign companies.

The current (2018) Australian Government has made a cut in the corporate tax rate to 25 per cent a major part of its political platform. The major reason given is the need to compete for international capital which does not have the benefits of dividend imputation and which is seen as highly mobile. A large part of capital for Australian business is foreign. It is argued that the lower rate will substantially increase private investment in Australia, which in turn will increase output, employment and, wages.

However, there are several counter-arguments. First, at the time of writing, the Government does not appear to have provided estimates of the likely additional foreign investment induced by the rate cut. Much new foreign investment is non-marginal and will occur at the 30 per cent rate. It may be more efficient to attract new foreign investment with deductions for capital expenditure. Second, given imputation, there is will be little effect on locally funded investment (unlike in the US which does not provide imputation). Third, there will be a significant loss of tax revenue from overseas sources, including tax revenue on existing investment, that will need to be replaced by local tax. It is not clear how this revenue shortfall will be filled. But any additional local tax will have some negative impact on aggregate demand and output. It may also have an efficiency cost (a DWL). Note, there may also be a loss of local tax revenue if high income individuals shelter income in corporate entities. Fourth, companies may reward shareholders (and managers) from the increased post-tax profits ahead of making additional investment. Thus, the outcome may be regressive.

As the Henry Tax Review observed, corporate tax reform is a complicated issue.

Summary

- Companies are taxed on profits as defined by legislation, which are quite different from economic profits and may also differ from accounting profits.
- There are various corporate tax systems. The classical system taxes both company income and dividends in the hands of recipients. With full imputation, recipients receive tax credits for any tax paid by the company. With partial imputation, recipients receive partial tax credits or pay tax at a concessional rate on dividend income.
- Double taxation was justified traditionally on the ground that company structure gives investors the valuable protection of limited liability. Also, it was assumed that most share capital is owned by wealthy individuals who bear the tax.
- Modern analysis suggests that the corporation tax is borne by all owners of capital (because after-tax returns are equalised), by users of capital who pay more for capital and by workers because the capital to labour ratio declines.
- The classic double-taxation system creates economic distortions. It raises the cost of capital and discourages investment, favours debt finance over equity, favours retained earnings over distributions and discriminates against incorporation. Given these disadvantages, it is surprising that Australia and New Zealand are the only countries to adopt full imputation of corporate taxes.
- One radical strategy would be full integration of company and personal tax with all profits (not just dividends) fully attributed to shareholders and no separate company tax. But there may be practical constraints. It could be harder to track profits and there could be excessive tax evasion and administration costs.
- Another major proposal to neutralise the double tax on equity is that companies be allowed to deduct an allowance for corporate equity (a normal rate of return on capital) so that they are taxed only on economic profits.
- International taxation also creates issues. Countries maximise their income by allowing foreign tax to be a deduction from taxable income rather than a tax credit. They also maximise local value added by not taxing mobile capital, but it is hard to distinguish between mobile and captive capital.
- The major current debate is about the size of the corporate tax rate rather than the structure. Advocates of a major cut in the corporate tax rate argue that it will greatly stimulate investment, employment and growth. Sceptics view it as more likely to be regressive with major benefits to affluent shareholders and little trickle-down effects to employees.

Questions

1. In what ways are taxable profits different from accounting and economic profits?
2. Why should corporate income be taxed rather than treated like trusts or partnerships where, for tax purposes, all income is fully distributed and taxed in the hands of the recipients?
3. Is there any double taxation of company income in the Australian system of corporate taxation?
4. What factors determine the incidence of company income taxation?
5. Is company tax a tax on entrepreneurship?
6. How may company tax affect the financial structure of firms?
7. Is company tax likely to affect the amount of capital formation? What are the arguments and what is the evidence?
8. Why does a change in the corporate tax rate have no effect on domestic shareholder income under a full imputation system?
9. What are the arguments for and against the Mirrlees Tax Review's proposal for an allowance of corporate equity?
10. What are the main issues governing a country's approach to international taxation?

Further Reading

Auerbach, A.J. (2005) *Who Bears the Corporate Tax? A Review of What we Know*, Working Paper 11686, National Bureau of Economic Research, Cambridge, MA.

Auerbach, A.J. (2006) 'The future of capital income taxation', *Fiscal Studies*, 27, 399–420.

Auerbach, A.J., Devereux, M.P. and Simpson, H. (2008) *Taxing Corporate Income*, Working Paper 14494, National Bureau of Economic Research, Cambridge, MA.

Devereux, M., Griffith, R. and Klemm, A. (2002) 'Corporate income tax reforms and international tax competition', *Economic Policy*, 17, 451–495.

Harberger, A.H. (1962) 'The Incidence of the Corporation Income Tax', *Journal of Political Economy*, 70, 215–240.

Henry Tax Review (2010) *Australia's Future Tax System, Part 2*, Vol. 1, 147–204.

Markle, K.S. and Shakleford, D.A. (2011) *Cross-Country Comparisons of Corporate Income Tax Rates*, Working Paper 16839, National Bureau of Economic Research, Cambridge, MA.

Mintz, J.M. (2011) 'An evaluation of the business tax recommendations of the Henry Review and the Australian Government response', pp. 161–182 in eds: C. Evans, R. Krever and P. Mellor, *Australia's Future Tax System: The Prospects after Henry*, Thomson Reuters, Sydney.

Mirrlees (Chair the Mirrlees Review) (2011) *Tax by Design*, Chapters 17 and 18, Institute of Fiscal Studies, London.

Rosen, H.S. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapter 19, McGraw-Hill, New York.

Smith, G. (2010) 'The Henry Review and taxing business income and rents', *Australian Economic Review*, 43, 422–428.

Chapter

32 Taxation of Consumption, Wealth and Not-for-Profit Organisations

Excise—A hateful tax levied on commodities.

Samuel Johnson

Taxation of Consumption ♦ Taxation of Wealth ♦ Australian Taxation of Land and Housing ♦ Tax Treatment of Not-for-Profit Organisations

In this chapter we turn to the taxation of the other two major tax bases, namely consumption and wealth. Both tax bases are important contributors to tax revenue, but generally less so than the income tax base. In both cases we examine the nature of the tax bases, applications in Australia and elsewhere and efficiency and equity implications of taxing these bases. We also discuss the tax treatment of not-for-profit organisations. In Australia, not-for-profit organisations provide substantial services and may indeed make large profits, but many also receive substantial tax concessions. We discuss the characteristics of these organisations and the economic implications of the various tax exemptions.

Taxation of Consumption

The term 'taxation of consumption' refers to the taxation of expenditure as distinct from the taxation of income. As an expenditure tax, it may be may be levied on intermediate goods (business inputs) or on final goods. As a practical matter, expenditure taxes are usually levied on producers rather than consumers, but we know from the study of tax incidence that economic incidence may be different from statutory incidence. In perfectly competitive markets, firms with perfectly elastic supply curves shift taxes forwards fully on to consumers. In imperfectly competitive markets, which are common, firms also shift most taxes forwards to consumers, but they may bear some part of the tax. When the latter occurs, a tax on expenditure becomes a tax on income rather than on consumption.

The main consumption tax in most OECD economies today is the value-added tax (VAT) described in Australia as the goods and services tax (GST), which is levied on producers. But there are many other actual or potential consumption taxes including: customs duties, excise taxes, retail sales taxes, a general expenditure tax and various other indirect taxes.

Customs duty
Tax on an internationally traded good

Customs duties are taxes on internationally traded goods. They are generally levied on selected imports, but sometimes on exports. Customs duties were designed historically to raise revenue. They are also used to protect local capital and labour (see Chapter 21) as well as for preferential trading purposes.

Excise taxes are taxes on consumption of selected goods. They are generally levied on producers because it is easier to collect taxes from a few firms than from millions of consumers. Historically, they were often imposed to match customs duties on imported goods. Hence, customs and excise departments would be combined in one agency. Today, excise duties are usually imposed on only a few selected goods, such as alcohol and tobacco products, partly to raise revenue on goods in inelastic demand and partly to discourage consumption of demerit goods (although these motives are not entirely consistent).

Excise tax
A tax on consumption of a selected good

Retail sales taxes are taxes levied on retail sales at point of purchase. Government may levy a retail tax on selected goods, possibly at different rates, or a general retail sales tax. Because there are fewer retailers than firms in the economy, a retail tax has lower administrative costs than a general VAT. But retail goods are not readily defined. Some goods may be used as business inputs or for consumption. Whether a good is a retail sale or not depends upon the purpose of the purchase which is hard to test.

Retail sales tax
Tax on retail sales

Value-added taxes are *ad valorem* taxes levied on the value added at each stage of production. Value added (VA) is the difference between a firm's sales and the purchased material inputs used in production:

$$VA = V_O - V_I = W + GOS \tag{32.1}$$

Value-added tax
A tax on value added at each point in production

where V_O is value of output (sales), V_I is payments for inputs, W = wages and GOS = gross operating surplus. Thus, VAT can be estimated in two ways: by subtracting the value of inputs from sales or by addition of wages and gross operating surplus. The subtraction method is more common. Generally, firms pay a tax on their sales (which is included in the price to purchasers) and obtain credits for taxes paid on inputs.

Although government can levy a single VAT on all goods and services, it usually exempts many goods and services. Because VAT is usually designed with a consumption base, it is generally collected on imports but not on exports. Exporters may also get credit for VAT paid on inputs.

Whether VAT is a tax on consumption or on income depends on the treatment of capital purchases. If capital purchases are treated like other purchases and firms obtain full allowance for any tax paid on capital items at the time of purchase, a general VAT falls wholly on consumption and has no effect on income. This form of VAT is known as the net consumption type of VAT and is the most common kind. A 10 per cent net consumption VAT is equivalent to a 10 per cent retail sales tax. If tax credits are not allowed for taxes paid on capital goods, investment is taxed as well as consumption and the VAT becomes in part a tax on income.

A cash flow tax. This tax concept is very similar to a VAT (or GST). As described by the Henry Tax Review (2010), firms would pay a tax on their net cash flow in any period, which would be equal to VA. Essentially a cash flow tax is GST without invoices. Firms would not have to levy a tax on sales or obtain credits for taxes paid on inputs. Wage payments would count as part of VA and not be a deduction for tax purposes. Capital payments would be included wholly as an expense. Despite its simplicity it does not seem to have been adopted anywhere perhaps because monitoring is harder and evasion easier with a cash flow tax than with a VAT (or GST) system which requires a paper trail of invoices.

Cash flow tax
A tax on the net cash flow of a firm

A general expenditure tax is a tax on a person's income less savings (i.e. on their consumption) over a period such as a year. An expenditure tax can be varied with the amount of expenditure and with the characteristics of a person's household and so be equitable. Despite its apparent simplicity, an expenditure tax has been employed only briefly in India

General expenditure cash
A tax on an individual's income less saving

and Sri Lanka. This may be because of the difficulty of measuring savings (or dissavings) each year or perhaps because it is thought unfair to exempt income from capital.

Other indirect taxes include wholesale taxes, turnover taxes, payments for licences and government user charges in excess of costs. Wholesale taxes are levied on wholesale sales. Although they are usually intended to be a tax on consumption, they may be a tax on income when the tax is selective and cannot be passed on in the domestic or export market.

A turnover tax is a tax on the gross value of sales (V_O) at each level of production. There is no deduction for inputs. This tax is rarely employed these days. It may involve multiple levels of tax on one final good. Also, the yield of a turnover tax depends arbitrarily on the number of stages of production.

Licences are taxes paid for the privilege of producing, owning or using a good or service. Australian examples include licences for provision of gaming services and for owning a motor vehicle. High government user charges relative to costs may also be viewed as an implicit form of excise tax.

Taxation of consumption in practice

As shown in Table 2.4 (page 28), as at 2017-18, revenue from consumption taxes accounted for about 27 per cent of total tax revenue in Australia. Of this nearly half came from the GST and just over a quarter from excise taxes. Taxes on imports, gambling, insurance and motor vehicles contributed most of the balance of consumption taxes.

In Australia, the GST is levied at a rate of 10 per cent on around 57 per cent of consumer expenditures. Thus, it is not a general tax. Expenditure on health, education and most basic food is 'GST-free' (zero rated). This means that there is no GST on sales and GST on inputs is refunded. Residential rents and financial services are 'GST exempt'. To be GST exempt means that GST is not levied on sales, but GST on inputs is not refunded.

Given these variations in GST treatments, the classification of commodities has significant implications (and complications) for tax liabilities. For example, GST-free food includes fruit and vegetables, meat, milk and cheese, tea and coffee, breakfast cereals and sugar, among other products. Food that attracts GST includes prepared food such as sandwiches, confectionery, savoury snacks, bakery products, ice-cream and biscuit goods. In the education and health sectors similar demarcation issues arise. For example, some courses in adult and community education colleges attract GST and others do not.

The excise tax in Australia has a narrow tax base. Nearly all the revenue is derived from taxes on beer, spirits, tobacco, petrol and other fuels, which are essentially taxes on drinking, smoking and driving. This excise revenue is in addition to GST revenue from these goods. Unlike the GST, the excise taxes are mainly unit taxes. The taxes on tobacco, beer and spirits are indexed to the consumer price index. The excise on petrol is changed from time to time, not indexed.

International comparisons. The Australian Treasury (2006) review of international taxes showed that revenues from consumption taxes averaged 9.4 per cent of GDP in the 10 largest OECD economies (OECD-10), compared with 9.0 per cent of GDP in Australia.

In the last two decades most OECD countries have shifted towards broad-based consumption taxes. Twenty-nine out of 30 OECD countries levy a VAT. The exception is the United States. But most US states levy a retail sales tax. As at 2006, the unweighted VAT rate in OECD countries was 17.4 per cent. In the European Union, a minimum standard rate of 15 per cent is prescribed. Thirteen OECD countries have no zero-rated goods. Consequently, revenues from VAT in OECD countries currently average 19 per cent of tax revenues compared with 13 per cent in Australia.

On the other hand, Australia has high excise taxes. Australia's reliance on these duties is the third highest in the OECD-10. This reflects mainly the high excise taxes on alcohol and tobacco. Also, Australia has the highest tax burden on use of motor vehicles in the OECD-10. However, Australia has a relatively low excise tax on fuel. As at January 2005, Australia had the fourth lowest excise tax on fuel in the OECD-30 and the third lowest unleaded petrol price (Australian Treasury, 2006).

Efficiency issues in commodity taxation

We have examined commodity taxes in several earlier chapters, notably in the discussions of externalities, efficient taxation and optimal taxation (Chapters 13, 27 and 28). As we have seen, the effects of a commodity tax depend on its characteristics, notably whether it is a broad-based tax or a specific excise tax. In this discussion we focus on the VAT (GST) system.

There are two main principles for efficient commodity taxation. First, to raise a target amount of revenue from selected commodities with least deadweight loss (DWL), there should be an equi-proportionate fall in the consumption of all goods. This implies setting differential tax rates so that the compensated demand for each commodity is reduced in the same proportion. This implies in turn that, other things being held constant, tax rates would be related inversely to the compensated own price elasticities of demand. However, other things are not constant because cross-price elasticities of demand and income effects (as well as administration costs) should be taken into account.

Second, broad tax bases with low tax rates generate a given amount of tax revenue at lower DWL than narrow tax bases with high tax rates. This is because across-the board increases have less impact on relative prices and the DWL rises with the square of the tax rate. Importantly, if revenue is collected via a general increase in prices of all commodities (rather than by excises on selected commodities) there is no distortion of relative prices. Thus, a broad-based VAT with a low tax rate generally has a lower DWL than do selective excise taxes. However, selective excise taxes may be justified to correct market failures (as a corrective tax) or as a tax on a demerit good. Thus, a tax on fuel consumption may be viewed as a corrective tax and a tax on tobacco consumption as a tax on a demerit good.

A VAT is also more efficient than a wholesale tax levied principally on manufacturers. As the share of services in the economy increases, a wholesale tax becomes increasingly narrowly based. Unless the tax rate is increased this tax yields a declining proportion of GDP as revenue. Also, levying taxes at a single (wholesale) point in the production chain generates arbitrary variations in the effective tax rate depending on the extent to which retail services add value to wholesale goods.

The VAT has smaller efficiency advantages compared with a retail sales tax. The VAT avoids the need to define retail sales or consumption goods, which is hard to do. Also, the double-entry accounting feature of VAT systems helps to reduce tax evasion. On the other hand, VAT systems are compliance intensive. There are far more tax points than with a retail sale tax.

In practice, most VAT systems, including the Australian GST system, have a general VAT rate alongside some VAT-exempt goods and along with some specific excise taxes. On efficiency grounds it is hard to justify narrowing the tax base with the exclusions. On the other hand, a more comprehensive system of differential tax rates would be administratively complicated and difficult to justify given the limitations of the inverse pricing rule theorem.

International tax principles. There are also advantages of a uniform VAT system across countries. Differential commodity tax rates between countries may distort international production or consumption just as differential income taxes do. Moreover, in this case there are no offsetting tax credits. For example, duty-free sales for exports within the European

Union (EU) distorted investment decisions, with major shopping malls built at airports and ferry terminals and operated on passenger ferries across the English Channel. To reduce such distortions, the EU abolished duty-free sales on international trips within the EU in 1999. However, price differences due to variations in the VAT and other costs remained. To reduce these differences, the EU adopted its minimum VAT rate of 15 per cent, though not full harmonisation.

Destination taxes
Goods are taxed when consumed at destination

Also, all EU countries adopt the **destination principle** of commodity taxation. Exports are zero rated for VAT purposes on the basis that the taxation rate at destination applies to all goods in that country. This means that consumers in each country can purchase goods from the cheapest source after allowing for transport costs. Therefore, international production is efficient. However, destination taxes may create consumption inefficiency because consumers have an incentive to purchase goods in countries with low VAT rates.

Source taxes
Goods are taxed when produced at source

This distortionary consumption effect could be avoided by **source (or origin) taxation**, whereby countries would levy taxes on all goods produced within their country, whether sold at home or exported. Consumers would then have no incentive to purchase locally produced goods outside their home country. However, differential source tax rates could distort the location of international production,

Such distortions of consumption and production can be overcome only by full tax harmonisation. All relevant countries, for example those in the EU, would agree on both a similar basis for taxation (destination or source) and on uniform tax rates. However, where countries have different expenditure preferences, revenue needs or different tax bases, they may prefer different tax rates. Tax harmonisation may prevent a country choosing its preferred tax rates.

Equity issues in commodity taxation

Commodity taxes are often regarded as inequitable because they are not based on the ability of individuals to pay tax. For any given taxed purchase, a low-income individual pays the same tax as a high-income individual. Accordingly, commodity taxation is often described as regressive. This means that commodity taxation takes a higher proportion of the income of a low-income earner than of a high-income earner. However, this occurs partly because high-income earners save a greater proportion of their income. Measured against expenditure, consumption taxes tend to be broadly proportional rather than regressive, although this depends on the commodities that are taxed.

In principle, it would be possible to design a progressive commodity tax structure (measured against income or expenditure) by levying differential tax rates. Goods that are a high proportion of the budgets of low-income individuals (sometimes described as necessities) would be taxed at a low, or zero, tax rate. Goods consumed mainly by high-income persons (sometimes described as luxuries) would be taxed at a high rate. Basic foods and clothes would attract a low tax rate; large boats and air travel would be taxed more highly.

However, there are strong arguments against such differential tax rates. First, income taxes are a more efficient and better targeted redistributive instrument. They can also be adjusted to deal with any inequities arising from other parts of the tax system, including GST-related inequities. Selective commodity taxes are a clumsy way to achieve distributional goals. Some high-income people would consume low-taxed goods and low-income people consume highly taxed goods. Second, such differential equity taxation could depart significantly from efficient differential tax rates and so create high DWLs. Third, attempts to obtain additional equity by exempting various goods from commodity taxation increase tax complexity and administrative costs.

Final observations on commodity taxation

As we saw in Chapter 28, the aim in designing a tax system is to create an optimal combination of taxes. If income tax deals adequately with distributional issues, consumption taxation should be efficient and not aim to achieve distributional objectives. However, in practice it may be hard to ignore the distributional implications of commodity taxes. The public may not be easily convinced that design of taxation for one tax base will offset tax incidence in another.

Our discussion of efficient commodity taxation above focused on impacts on consumption. Efficiency issues also arise with labour supply and savings. In a world without fiscal illusion, commodity and wage taxes have a similar effect on labour supply. Suppose Anne earns $100 an hour, is taxed $20 and takes home $80. She can buy $80 worth of goods. Now suppose that Anne can keep all her income, but that the government raises $20 in every $100 purchased (with a 25 per cent tax on $80 of consumption). Anne again consumes goods worth $80 before the consumption tax is added on. For any given level of revenue collection, a tax on consumption and tax on wages would have a similar deterrent effect on labour supply.

On the other hand, a commodity tax does not tax income from capital and so does not reduce investment or distort inter-temporal consumption choices. However, the net effect on investment may be small as taxation of income from capital appears to have only a small impact on savings.

There are also two practical advantages of consumption taxes. First, there are fewer problems with inflation than with income taxation. With *ad valorem* taxes, inflation is automatically allowed for. Second, it is arguably harder to avoid commodity taxes than income taxes (through trusts, offshore incomes and so on). Therefore, consumption taxes may make payment of taxes more equitable rather than less.

Taxation of Wealth

Wealth is a stock variable. An individual's wealth at any point in time equals the value of assets owned less any liabilities. Assets may include physical assets, most often land and housing, and financial assets such as shares, bonds or term deposits. The assets may be held directly or indirectly via managed funds (including superannuation). A general tax on wealth is a tax on the value of these assets less any liabilities. A selective tax may be imposed on the value of any of these assets and may not allow for any liabilities. As we will see, a tax on the capital value of an asset has a similar effect to a tax on the income from that asset (which is a flow variable).

As shown in Table 32.1, there are five main kinds of tax on wealth: a general wealth tax, taxes on land and housing and on natural resources, taxes on capital transactions and taxes on bequests and other gifts. We examine below Australian and international wealth taxes and the efficiency and equity implications.

Taxation of wealth in Australia and internationally

In 2017-18, Australian taxes on wealth were just over 10 per cent of total tax revenue. As a proportion of total tax revenue, Australia raises slightly more revenue from wealth taxes than most other OECD countries. Also, wealth taxes are raised in different ways. Compared with other OECD countries, Australia has high taxes on financial and capital transactions Australia raises an average amount from property taxes. However, there are no taxes on bequests of any kind.

Few countries raise a general tax on net wealth. Of the OECD-10, only Switzerland, Canada and Spain raise such a tax, and only Switzerland raises significant revenue from it. Thus, in not raising a general wealth tax, Australia is in the majority.

Table 32.1 Taxonomy of major wealth taxes

A periodic general wealth tax
A tax on an individual's net worth—their assets less their liabilities
Taxation of immovable real property
General taxation of land (possibly with some exemptions)
Taxation of land used for commercial purposes
Taxation of residential land
Taxation of residential housing
Land betterment tax (on unrealised gains)
Taxes on natural resources
Taxes on mineral and petroleum resources
Taxes on property transfers and financial transactions
Taxes (stamp duties) on transfer of property
Taxes (stamp duties) on transfer of equities
Taxes on bank transactions or on credit cards
Taxes on bequests and other gifts
Death duties or inheritance taxes
Taxation of gifts (transfers of wealth)

In Australia, as in most countries, taxes on real residential and commercial property make up around half of all tax income from wealth.[1] In almost all countries these taxes are levied by sub-national governments. These taxes do not include taxes on rental income from ownership of property or taxes on capital gains, which are treated as taxes on income from capital and which normally accrue to central government. At about 1.3 per cent of GDP, tax revenue from immovable property in Australia is close to the OECD average. However, property taxation in Australia differs from most other countries. In Australia, nearly all taxes on property are levied on "improved" land values. These land values include capital improvements to the land, such as provision of water and sewerage services, but not structures. In most other countries, property taxes are based on property values inclusive of building improvements.

Taxes on property transfers and financial transactions may include stamp duties on residential and non-residential conveyances and on mortgages, leases, shares and other products, sometimes on bank or credit card transactions. Australian reliance on these taxes (principally on property transfers) is the highest in the OECD and twice as high as the OECD average (Australian Treasury, 2006). Of the other OECD countries, only Spain and Ireland have similar reliance on these taxes.

On the other hand, Australia is one of only two OECD-10 countries, along with Canada, that does not impose any estate, inheritance or gift taxes.[2] Also, most other OECD countries impose both inheritance and gift taxes (though New Zealand has only a very small gift tax). However, most of these taxes are levied progressively in relation to the total value of the inheritance or gift, exempt family members or treat them concessionally and exempt donations to organisations such as charities that are deemed to act in the public interest. Because of these concessions, tax revenue from inheritance and gifts is generally very low and only around 0.3 per cent of GDP even at the high end of the range (in Japan, the Netherlands and the United States).

[1] In Australia, owner-occupied property and land used for primary production are usually exempt from property tax.

[2] Australian Treasury (2006) notes that Canada effectively imposes bequest taxes through deemed disposition in income tax legislation. In Australia, from July 2007 capital left in superannuation accounts after the death of the beneficiary is taxed at 15 per cent (except when distributed to spouses).

General effects of taxation of wealth

In principle there is no difference between the taxation of a stock of wealth and the taxation of the flow of income from it. Given appropriate tax rates, the taxation of wealth and of income from wealth is equivalent. Let the value of a stock of capital be K and the pre-tax return from capital be ρ, so that the gross income from capital is $K\rho$. There is a tax rate on the capital (t_k) that produces an equal amount of tax revenue to a tax rate on the income from capital (t_ρ):

$$Kt_k = K\rho\, t_\rho \tag{32.2}$$

Suppose that capital provides a 6 per cent rate of return and that t_ρ equals 30 per cent. This is equivalent to a tax rate on the stock of capital (or wealth) of 1.8 per cent.

Assuming no fiscal illusions, the taxation of capital would have similar effects to the taxation of income from capital. Given that savings are generally a positive function of the after-tax return on capital, a wealth tax reduces this return and hence reduces savings and capital formation, the capital–labour ratio and productivity. The tax also creates a DWL for individuals who are discouraged from saving (via the substitution effect). The effects on savings are doubled if both the capital stock and the income from it are taxed.

Equity issues are complex. Taxing wealth may be viewed as equitable if wealth is correlated with ability to pay or if it is unearned in the hands of the recipient (a gift of some kind). However, double taxation of capital and of the income from the capital may be viewed as inequitable. Indeed, if income is also taxed when earned, savings would be triply taxed.

However, there are some contrary points. First, some forms of capital provide untaxed, non-monetary benefits. In most OECD economies, households hold over half their capital in their homes and consumer durables. The imputed rent from these assets is rarely taxed. Therefore, taxing some forms of capital, such as property values, may capture some sources of real income that are not captured by income tax. Second, where property prices rise with the provision of local services, some property taxes may be viewed as a benefit tax for the provision of these services, which could make such taxes both efficient and fair. Third, the efficiency effects of a wealth tax depend on the mobility of capital. Taxes on immovable capital, such as land, have less DWL than taxes on mobile capital because the existing fixed capital cannot escape the tax.

In conclusion, there are efficiency and equity arguments for some taxation of capital stock, especially property which does not produce taxable income. The case for taxing wealth is weaker when income from the stock is already taxed. To say more than this, we need to examine the merits or otherwise of each main form of wealth tax.

A general wealth tax. There are strong reasons against a general wealth tax. It is generally easier to tax income from capital which is observable than personal net wealth which requires regular valuations. If the tax rate were to rise with wealth there would be much avoidance activity. In any case, when income from capital is taxed, a general wealth tax would represent double or even triple taxation, which would be inefficient and arguably unfair. Accordingly, few countries levy a general wealth tax. However, selective taxes on assets which are immobile, especially on assets that do not provide an explicit income stream, such as owner-occupied homes, are commonly adopted.

General wealth tax
A tax on the value of an individual's total assets less total liabilities

Taxation of immovable real property. Most of the literature on the taxation of property discusses two issues: taxes on the value of unimproved land and taxes on the value of property inclusive of structures. As it happens, neither of these taxes is levied in Australia, where all property taxes are levied on improved land values (improved land but excluding structures). We discuss this form of property taxation in the next main section. Here we discuss taxes on the value of unimproved land and on property values inclusive of structures.

Unimproved land value
The value of land without any capital improvement

Taxation of unimproved land. The idea of a tax on unimproved land values was popularised by Henry George (1879). Indeed, George proposed that the 'ground rent' from land be taxed 100 per cent. By ground rent, George meant the economic surplus from use of the land after payments for any other factors of production including any capital investment in the land. Thus, ground rent did not include any return to investment in the land. The unimproved capital value (UCV) of land is the capital valuation of the 'ground rents'. Because the supply of land is a gift of nature and fixed, a tax on ground rent is a tax on pure profit and has no effect on land supply. George argued that such a tax it is equitable because land is a gift of nature, not a result of capital or labour.

Despite the attractions of these arguments, there are three significant problems with a 100 per cent tax (or any very high tax rate) on ground rent or UCV. First, unimproved land values have an important role in allocating land to its most productive use, to housing, different types of agriculture and so on. If these land values are reduced to zero after tax, landowners have no incentive to use land in its most productive manner. Without a price mechanism, land would be used most inefficiently. Second, it is difficult to determine how much of the market price of land is pure land value and how much is due to capital improvements to the land. This could result in expensive administrative and judicial procedures. Third, introduction of a tax on pure land values would be retrospective in effect, as many landholders would have paid prices for the land that did not allow for such a tax.

Taxes on property values. Turning to the more common tax on property values, the effects depend on the scope of the tax. Notably, they depend on whether the tax is a small or a large part of the total tax on capital. If it is a small part, the effects can be analysed using partial equilibrium analysis. If it is a large part, a general equilibrium analysis is required. These two approaches (partial and general equilibrium) are sometimes described as the traditional and the new view of the effects of a property tax. This is rather misleading because the partial equilibrium approach still applies in many situations, notably when many local governments levy a property tax (as they do in the United States), and we want to determine the effect of a change in one or a few of these taxes. In such cases, most effects occur in the local area(s) making the tax change and effects elsewhere negligible.

The partial equilibrium approach assumes that a property tax works like an excise tax on the annual income derived from the property. But the tax has a different effect on the land and the housing components of property value because the local supply of land is fixed and the amount of capital spent on it is elastic. As shown in Figure 32.1a, the property tax on the home occupier shifts the demand for land down from D_1 to D_2. With a perfectly inelastic supply of land (S_L), the tax on the land component reduces the amount received by landowners and is borne entirely by landowners. Of course, the occupier of the property is often also the landowner. If future property rents are capitalised into lower initial land prices, the whole stream of property taxes is borne by the initial landowner.

In panel (b), the property tax reduces the demand for the structural component of the property from D_1 to D_2, but the supply of capital to build structures (S_B) is assumed to be perfectly elastic, at least in the medium run. If the return on local property is lower than the return on capital elsewhere, the capital will move to another location or sector and the supply of structures falls from B_1 to B_2. In this case the rents for the built component must rise until the after-tax return to capital is as high in the local housing sector as in other locations or sectors. Accordingly, after an adjustment period, the burden of the property tax is borne partly by the initial landowners and partly by renters in the form of increased rents.

Suppose now that there is a national property tax and that the capital employed in property is a substantial part of the capital employed in the economy. In a closed economy, capital will exit the housing market and be of such a magnitude to cause the after-tax return to capital to fall in all other sectors of the economy. In this general equilibrium analysis, the property tax

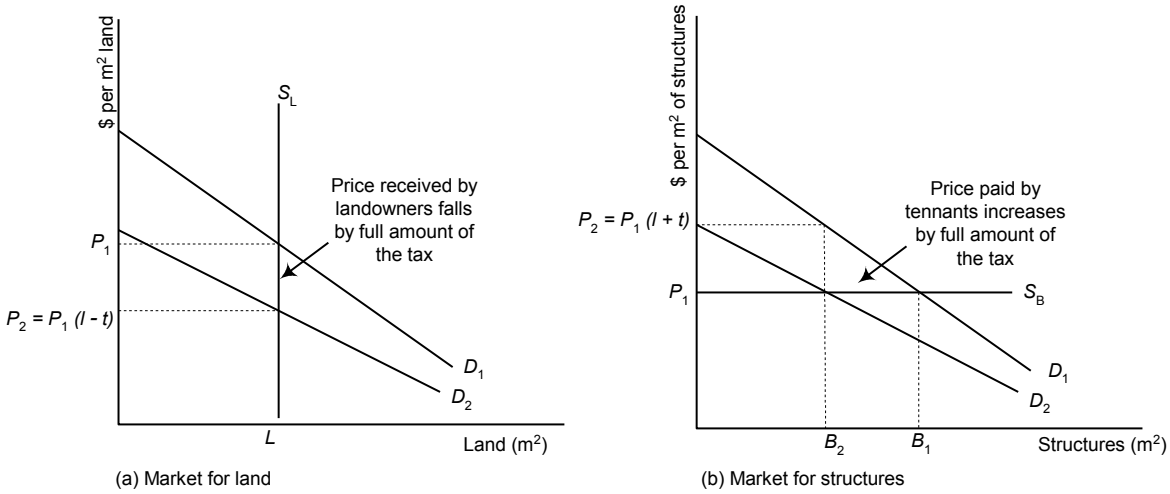


Figure 32.1 Incidence of local property tax on land and structures

is borne by all local owners of capital. Property rents rise, but by less than they would in the partial equilibrium model because the after-tax return to capital required in the housing sector has fallen.

However, in a small open economy the rate of return to capital is determined internationally. In this case the flow of capital out of the local property market will have only a small effect on the rate of return to capital in other parts of the local economy. Rents for structures in the property market will have to rise sufficiently to ensure that the after-tax rate of return in the property sector equals the internationally determined after-tax rate of return prevailing in other sectors.

Another view of property taxes is that they are user fees, not taxes, for the supply of local services to property owners. Studies in the United States by Oates (1969) and Weimer and Wolkoff (2001) found that residential property values are related positively to expenditure on local public services and inversely to municipal tax rates levied on property. Note that, when property values are high, local jurisdictions can have high tax revenues and expenditure along with low tax rates. The relationship between property values and local expenditure implies that local property taxes may be viewed as benefit taxes (i.e. as payments for services). In effect the demand for property schedules in Figure 32.1 would rise. This is also valid for Australia though to a lesser extent as fewer public services are provided locally and funded from local property rates.

These three views of property taxes (the partial and general equilibrium models and the user benefit hypothesis) may be valid in different circumstances. Where local changes in a property tax occur, the partial equilibrium approach is realistic. If national changes in the property tax occur in a closed economy, a general equilibrium model is realistic. And sometimes the property tax is a *de facto* user charge.

Turning to the efficiency effects of a property tax, the distinction between land and structures is again important. A general tax on unimproved land values has little DWL so long as it is low enough to have little or no effect on the total supply or use of land. On the other hand, selective land taxes may distort land ownership and use patterns. Also, a tax on structures is potentially inefficient because it may divert mobile capital away from housing towards lower valued goods.

Taxes on natural resources. In principle, a tax on a natural resource is like a tax on unimproved land. It would not affect the supply of the resource. In a world of certainty, where prices and costs are known, if the tax is less than the value of the resource, a tax on the resource value would have no effect on exploitation of the resource, as a normal return on capital investment can still be achieved (and there is no alternative use for the resource). This is illustrated in Figure 32.2a. Given the international price (P) for a mineral resource and an industry supply curve (S), the industry will produce Q_1 tonnes of output per annum. A tax on economic profit will reduce the return to mining (in Figure 32.2a the tax collects revenue shown by shaded area ABC), but not affect the level of output. Such a tax is described as a tax on economic rent or a pure resource tax.

Royalty tax
A tax on the quantity or value of mining or oil sales

On the other hand, a tax on the quantity or value of mining output (known as a **royalty tax**) creates a wedge between the sale price and the price collected by miners. As shown in panel (b), a royalty tax shifts the supply curve upwards by the amount of the tax and reduces output from Q_1 to Q_2. There is then a DWL equal to the shaded area ABC. This loss occurs because the addition to marginal cost may turn otherwise profitable marginal output into unprofitable output and cause firms to reduce output below its efficient level. For marginal mining operations, such royalties can have strong disincentive effects.

This standard analysis suggests that taxes on economic profits (rents) are superior to royalties. This may be the case, but matters are complicated in practice. Consider first the tax on economic rents or its equivalent, a tax on the value of the natural resource. A key problem is that neither economic rent nor the value of the resource is readily observed. Most mining companies are vertically integrated supply chain operations, and sometimes multi-product producers, with joint exploration, management and corporate expenses. Allocating joint costs to parts of the supply chain is often somewhat arbitrary.

Also, entrepreneurship needs to be rewarded as entrepreneurs often accept low incomes for years in the hope of large profits later. In addition, investors in mining require a risk premium and are unlikely to invest unless there is significant economic profit potential to offset the risks. Market valuations of resources usually reflect high discount rates. Moreover, when the total tax rate (on economic rents and corporate income) rises beyond about 50 per cent companies may seek ways to reduce tax, for example by increasing tax-deductible management payout, rather than minimising production costs, which itself is an economic waste.

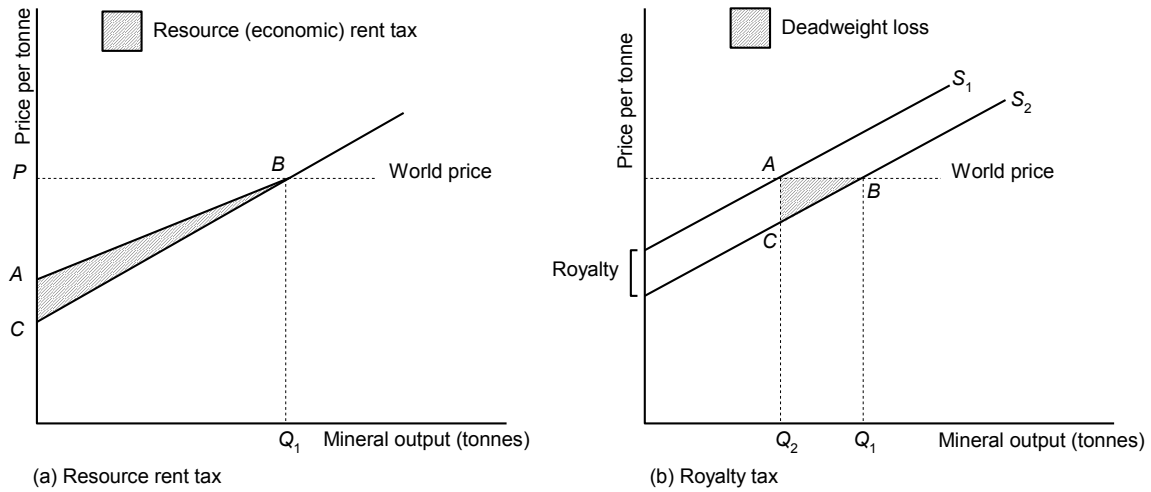


Figure 32.2 Comparison of resource (economic) rent tax and royalty tax

Moreover, the analysis above embodies two critical assumptions. The first is that the level of Australian output has no impact on the world price. This is not realistic for commodities like iron ore and coal. When the demand curve for a country's product is downward sloping, a country can gain by reducing output and increasing the world price (i.e. by exerting monopoly power and choosing quantity where marginal revenue equates to marginal cost). Secondly, we have implicitly assumed no foreign ownership of mining firms. If a mining operation is partly or wholly foreign owned, as many in Australia are, the royalties are not simply a transfer payment within Australia and some of the DWL is borne by foreign entities.

Australian experience exemplifies some of these issues. The Australian government receives around $1 billion annually from the petroleum resource rent tax (PRRT) and the states receive some $4 billion annually in royalties on mineral and petroleum production (mainly in Queensland and Western Australia). Under the PRRT introduced in 1987, the Australian government levies a 40 per cent tax on the economic profits of petroleum production *after* a full allowance for (debt and equity) capital costs and *before* corporate income tax. The state governments levy a tax on the production of minerals and petroleum either in the form of a unit tax on output or as a percentage of value realised on output before transport and processing.

The Henry Tax Review (HTR, 2010) proposed that a broadly similar 40 per cent tax should be applied to the economic profits of all mining enterprises in Australia. As a tax on pure profits (rents), this was intended to have no disincentive effect on exploration and investment in mining production. However, when prospects are uncertain investors weigh up whether the spectrum of possible returns after tax is sufficient to compensate for the possibility that they may lose their whole investment if the business is unsuccessful. To deal with this, the HTR proposed that the government would meet 40 per cent of losses (below a normal return on capital) and that this should make investors risk neutral.

In 2011, the (Labour) Australian government introduced a new Minerals Resource Rent Tax (MRRT) whereby excess profits were taxed at a nominal rate of 30 per cent, to apply only to iron ore and coal as from 1 July 2012. Small miners with less than $50 million profits would be exempt from the MRRT. The MRRT would be assessed on the value of commodity at its first saleable form (at the mine gate) less all costs to that point. However, the effective tax rate was 22.5 per cent because taxable excess profits will be allowed a 25 per cent extraction allowance in recognition of the contributions of the miner's expertise to profits at the mine gate. There were other complex valuation issues. The government also agreed to allow the mining companies to credit any increase in state government royalties against the MRRT. Thus, the MRRT did not displace inefficient royalty taxes. However, three years later the (Liberal) Australian government introduced legislation that repealed the MRRT.

Taxes on property transfers and financial transactions. These taxes can greatly increase the costs of trade and discourage trades. Consequently, productive assets may not be owned by those who can make best use of them and consumption assets may not be owned by those who value them most highly. For example, high stamp duties on property transfers deter households from moving into more suitable housing and result in an inefficient allocation of the housing stock. Abelson and Joyeux (2007) estimated that Australian stamp duties increase the costs of exchanging houses by 33 per cent, deter 7 per cent of housing exchanges and have a DWL of $375 million per annum.

Taxes on estates and gifts. These taxes may be justified on equity and efficiency grounds. Recipients of large untaxed bequests are strongly advantaged over other individuals. This is inconsistent with the principles of horizontal equity and equality of opportunity. This also suggests that an estate tax might be more equitable if it were taxed in the hands of the recipient rather than on the total size of the estate.

Arguably, the break-up of family holdings in a business may be regarded as efficient in that the assets are more likely to finish in the hands of those who can utilise the assets most productively. In the words of the wealthy American industrialist, Andrew Carnegie: 'the parent who leaves his son enormous wealth generally deadens the talents and energies of his son'. Rosen and Gayer (2014) cite research that supports this view. On the other hand, those who create wealth may be motivated by a desire to hand some of it on to their families. Therefore, an inheritance tax may discourage investment and savings.

However, a critical issue for estate taxes is tax avoidance. Avoidance may be achieved by designing a complex system of trust companies. A simpler method is to make bequests before death. To reduce this form of tax avoidance, death duties need to be integrated with taxes on transfers between living persons. In the United States estate and gift taxes are integrated through the Unified Transfer Tax. The gross estate includes gifts made during the decedent's lifetime. But there is a large lifetime exemption amount and spouses are exempt. These exemptions mean that death duties are paid on fewer than 2 per cent of estates in the United States. Administrative difficulties, as well as political constraints, are the major barriers to effective estate and gift taxes in most countries.

The Henry Tax Review (HTR, 2010) noted that most OECD countries impose estate taxes, either through taxes on the whole estate or on inheritances and argued that a bequest tax would be economically efficient. But the HTR also noted the administrative complexities (and political contentiousness) of inheritance taxes and suggested that the issue be discussed thoroughly and publicly before any action should be taken.

Australian Taxation of Land and Housing

Land and housing attracts many taxes because they are easy, immobile targets. In Australia, these taxes include state and local government land taxes (local council rates), stamp duties on property transfers, GST on sales of new dwellings and on renovations of existing ones, and capital gains taxes on rental properties and second homes. There are also charges on land and housing developers for the public provision of economic and social infrastructure—these are essentially user charges although some developers regard them as taxes.

Table 32.2 presents estimates of these tax revenues in absolute terms and in relation to the value of housing in 2004. Abelson and Joyeux (2007) estimated that the taxes amounted to nearly a quarter of annual private housing rents inclusive of imputed rents.

As noted above, land taxes in Australia are levied on the value of improved land and are therefore a tax on land and capital. Land taxes on owner-occupied housing are borne by the owner. However, land taxes on rental properties are borne mainly by housing tenants. Investors pass most of the land tax on in higher rents to tenants because capital is mobile and the after-tax return on housing equates approximately to the after-tax rate of return on alternative investments. In other words, the elasticity of rental housing supply is generally greater than the demand elasticity.

The incidence of developer charges and of GST for new houses is more controversial. The housing industry argues that both developer charges and the GST are passed on to consumers via higher house prices. However, new houses are a small part of the total housing market and their prices are set by the value they offer relative to established houses. If a developer has selected the highest possible sale price for a new property, he cannot increase this price because a new tax is imposed. The price of new houses will rise only if their supply is reduced. However, in many parts of Australia, the supply of new housing is determined by planning regulations and is not price sensitive. On the other hand, the price of raw land for housing generally exceeds the value of the land in alternative uses. In this situation, a tax on new housing reduces the price of raw land. It does not increase the price of new housing.

Table 32.2 Estimated value of taxes and subsidies for private housing in Australia in 2004

	Owner-occupied housing		*Private rental housing*		*All private housing*	
Subsidy/tax	*$bn*	*% annual housing value*	*$bn*	*% annual housing value*	*$bn*	*% annual housing value*
Taxes						
Land taxes (state governments)	na	na	1.2[a]	4.6	1.2	1.2
Land taxes (local government)	6.0	8.2	1.2[a]	4.6	7.2	7.3
Stamp duties	7.0	9.6	2.0	7.7	9.0	9.1
GST on land/new houses	3.0	4.1	0.5	0.5	3.5	3.5
GST on major renovations and additions	1.8	2.4	0.7	0.7	2.5	2.5
Total taxes	17.8	24.3	5.6	21.5	23.4	23.6
Subsidies						
Imputed rent tax concession	8.0	11.0	na	Na	8.0	8.1
No GST on imputed/actual rents	7.3	10.0	2.6	10.0	9.9	10.0
Capital gains tax concession	7.2	9.9	na	Na	7.2	7.3
Asymmetric tax treatment of losses and gains	na	na	1.6	6.2	1.6	1.6
First-home owner grant	1.2	1.6	na	Na	1.2	1.2
Private rent subsidies	na	na	1.8	6.9	1.8	1.8
Total subsidies	23.7	32.4	6.0	23.1	29.7	30.0
Subsidies – Taxes	5.9	8.1	0.4	1.6	6.3	6.4

(a) Estimated net tax for investors after allowing for lower income tax at a marginal tax rate of 40%.
Source: Abelson and Joyeux (2007).

Analysing the effects of taxes on housing is complicated because, trying to make housing more affordable, government also provides numerous subsidies to housing. In Australia there is no capital gains tax for owner-occupied housing. There are also grants to assist first-home buyers and rent subsidies for low-income households in private and public housing. As shown in Table 32.2, Abelson and Joyeux (2007) estimated that owner occupiers receive a gross housing subsidy equivalent to nearly a third of the annual rental value of their housing. Thus, they receive a net housing subsidy valued at just over 8 per cent of imputed rents. On the other hand, the estimated net subsidy for the investor/rental housing sector is about 1.6 per cent of annual housing rents. In Chapter 21 we considered the effects of the subsidy for homeowners on the price and consumption of housing and the associated DWL.

As we have seen, the taxation of pure land, a fixed resource, is efficient. However, land nearly always embodies capital expenditure and taxation of capital inputs to land may increase land and house prices. The Henry Tax Review (2010) pointed out that existing state-based land taxes are inefficient because they have many thresholds and exemptions and are not broadly based and that tax rates vary according to land use and landholding aggregation rules. The HTR proposed that a new land tax should apply to all land regardless of use based on the value per square metre of land above a certain value threshold that would exclude most agricultural land, and that the tax rate would be progressive. This tax would replace existing inefficient taxes including stamp duties on property transfers. However, the HTR proposal would not have a zero DWL because it would not be pure land tax (it would tax capital inputs) and it would have a distortionary tax threshold.

Tax Treatment of Not-for-Profit Organisations

In many countries, including Australia, not-for-profit organisations (NPOs) provide a wide range of services. Some of these are wealthy and receive large revenues and some run essentially commercial services. However, they are generally exempt from income tax and often receive other tax concessions. We describe below the nature of NPOs, their contribution to Australian GDP and the major tax concessions to them. We then discuss some economic and policy implications.

The nature of not-for-profit organisations. A NPO is an organisation that has no external equity interests. A NPO may have internal equity (net assets) and many NPOs do.[3] They may make and accumulate profits. However, a NPO must contribute any profit to the purpose for which the organisation was established. A NPO is not allowed to distribute profits to any of its members or to provide profit or financial gain to the individuals who establish or control it. NPOs are involved in a wide range of activities, including churches, church schools, cultural and environmental services, sporting clubs including many wealthy ones, community child care centres and neighbourhood associations. Most NPOs attract substantial tax concessions especially exemption from profit or income taxation.[4]

Charities constitute a high proportion of NPOs and attract the greatest tax concessions. An NPO can be designated as a charity if it has a dominant charitable purpose. However, legislation defines a charitable purpose very widely as provision of a 'public benefit'. An NPO provides a public benefit if it aims to achieve a common good, has 'practical utility' and is directed at a sufficient section of the general community. These are wide goal posts. Examples of charities include religious institutions, aged persons homes, homeless hostels, not-for-profit schools, organisations providing for people with disabilities, not-for-profit child care centres and societies that promote the fine arts (Australian Taxation Office, 2014).

The size of the not-for-profit sector. The Australian Bureau of Statistics (2009) reported that there were just over 41 000 NPOs in Australia at the end of June 2007. These NPOs employed 890 000 persons, including 370 000 full-time employees, and the rest part-time or casual employees, and contributed 3.4 per cent to Australian GDP.[5] Education and research accounted for 29 per cent of value added by NPOs, social services for 19 per cent, health for 18 per cent and culture and recreation for 12 per cent. These figures did not include not-for-profit public agencies, such as universities or public hospitals operated by religious orders, or mutual organisations (not-for-profits in the financial and insurance industry and trading co-operatives) that operate commercially and permit the distribution of assets to members when they are wound up.

Benefits of not-for-profit institutions. NPOs provide various benefits. Most importantly, they may supply services at lower cost than commercial firms can because they attract labour to work for them as volunteers or at low market rates. Second, members of the community may wish to join together to provide a local public good or service to their members that government is not providing and cannot readily provide. Third, some consumers may prefer to obtain services from NPOs rather than from private firms or government.

---

[3] 'Not-for-profit' organisations are often described simply as 'non-profit' organisation. The former term is a more accurate description.

[4] See Australian Taxation Office, 2014, *Tax Basics for Non-Profit Organisations*.

[5] If the value of free volunteer services is included, the estimated contribution of NPOs rose to 4.7 per cent of GDP.

Tax concessions for the not-for-profit sector. Tax arrangements for NPOs in Australia (Australian Taxation Office, 2018) are complex. Here we outline the major tax concessions.

Most NPOs in Australia are exempt from income tax. Almost the only requirement for tax exemption are that an organisation must be set up for an object other than making a profit and be based in Australia. As we have seen, they can and do often make very large returns on internal equity, often amounting to pure economic profit.

Unless claiming other tax exemptions, NPOs self-assess whether they are income tax exempt. If an NPO is income tax exempt, it is not required to pay income tax on profits from any services provided however incidental they may be to the primary purpose of the organisation. Therefore, profits from gaming machines or business ventures, returns from investments and realised capital gains are all income tax exempt. Of course, employee incomes are taxed and NPOs must deduct pay-as-you-go taxes from employee wages.

NPOs receive several other tax concessions. All NPOs are allowed a higher GST registration threshold than other businesses. Charities also obtain GST-free status for various sales and GST-exempt statues for various events (that are treated only as input taxed). Charities are exempt from the fringe benefit tax (FBT) on non-cash benefits to employees or eligible for a partial rebate on FBT. Many NPOs are exempt from state payroll and land taxes, and from municipal rates, or receive significant concessions on these taxes. Also, most NPOs receive donations that are income tax deductible. The Henry Tax Review (2010) estimated that the FBT and deductible gifts concessions each had a value of about $1.0 billion.

Efficiency implications of tax concessions for non-profit organisations. In principle, so long as NPOs aim to maximise their return on capital employed, exemption from income tax should not affect investment or operating decisions. If a charity is operating an unrelated business, it will maximise the contribution to charity by maximising the profits of this business.

However, the incentive to maximise profits is weakened when there is no explicit return on equity, albeit internal equity. Also, income tax exemption increases the after-tax profits retained by NPOs and encourages reinvestment in their own activities and businesses.

The lack of competitive neutrality also encourages inefficient outcomes. There are major competitive advantages in running businesses through an NPO and being GST exempt. NPO clubs that do not have to charge GST and can cross-subsidise meals and drinks from untaxed profits of gaming machines have a competitive advantage over hotels and public bars that must charge GST and pay income tax on profits. Instead of maximising profits by charging their members profit-maximising prices, NPOs may also provide significant indirect returns to members in the form of lower membership fees and service prices. On the other hand, the inability of NPOs to raise equity capital may constrain their capital expenditure, increase labour intensity and be relatively inefficient, and constrain their capacity to expand services.

Tax exemptions on inputs, notably FBT concessions, also distort resource allocation. These exemptions reduce the cost of NPO purchases and provide NPOs with a competitive advantage relative to standard commercial businesses. While these tax-driven cost advantages may be offset by the inefficient capital structures of NPOs, so that the overall size of the NPO sector and resources commanded by NPOs may be similar to what might occur in a neutrally competitive market, there is some inefficiency and a related deadweight loss as more resources are used by organisations with inefficient capital structures and operations.

Equity implications. The concessional tax treatment of NPOs also raises equity issues. The concessions recognise that many NPOs provide services to less well-off members of the community. Indeed, that is a major rationale for many NPOs.

On the other hand, the tax subsidies allow NPOs to provide services at lower prices to their members than commercial firms can provide to the public. Customers of NPOs can receive

cross-subsidised services at lower prices than are available from fully taxed commercial competitors. Church organisations often occupy prime municipal sites without contributing to the costs of municipal services. Employees of some NPOs can receive substantial fringe benefits free of tax. These examples appear to breach the principle of horizontal and vertical equity.

Conclusions. Some NPOs achieve efficiencies in supply that private firms or government cannot. They may also facilitate distributional objectives. However, there is no systemic set of principles underlying the NPO sector and many NPOs exist without clear efficiency or equity objectives. Many tax concessions for NPOs in Australia reflect tradition and political influence more than formal economic justification. Many tax concessions appear to be *ad hoc* decisions and may create significant deadweight losses and inequitable outcomes.

Summary

- This chapter examines the taxation of consumption and wealth and the tax treatment of not-for-profit organisations.
- Value-added taxes are the dominant form of consumption tax. Most countries also levy excise taxes on selected goods, some customs duties and various other indirect taxes.
- Given that income tax can target distributional objectives, consumption taxes should be designed primarily to raise revenue efficiently.
- The most efficient consumption tax system is a broad-based VAT (or cash-flow tax) with basically one tax rate and a few tax rate variations to correct market failures and to address special demerit goods.
- International production or consumption is distorted when countries adopt differential VAT rates. Distortions can be removed only if all countries adopt a similar tax principle (by source or destination) and a similar VAT rate.
- The taxation of wealth has similar effects on incentives and savings as a tax on income from capital except that it allows for the taxation of non-monetary returns to capital.
- If tax is levied on earnings, income from capital and wealth, there is triple taxation of savings.
- Taxes on improved land values or full property values are the main form of wealth tax. A tax on land is efficient in that land is not mobile. However, taxes on improvements are taxes on capital. Also, property tax needs to be designed not to make housing unaffordable.
- Taxes on natural resources are also efficient. But valuation of resources, with full allowance for the cost of discovery, exploitation and risk, may be difficult.
- Not-for-profit organisations provide some efficiency and equity benefits. However, the wide range of income and other tax concessions may create deadweight losses and some social inequities.

Questions

1. When is a commodity tax a tax on income rather than a tax on consumption? Give examples. Can the GST be a tax on income and not on consumption?
2. Can commodity taxes improve the allocation of resources?
3. Can and should commodity taxes be progressive?
4. Should the GST be (i) broadened and /or (ii) raised above 10 per cent in line with general international practice for the similar VAT tax systems?
5. What kinds of inefficiencies occur when countries fail to harmonise consumption taxes? Is tax harmonisation always desirable?
6. Are any kinds of wealth taxes desirable? If so, which ones and why?
7. Why are stamp duties on the purchase of housing and insurance considered especially inefficient?
8. What are the current problems with land taxes in Australia? What kind of land tax would be efficient and fair? What might be the disadvantages of land tax?
9. What are key features and merits of a resource rent tax versus a royalty tax? Is a royalty tax ever preferable to a resource rent tax? What issues arise with the Australian Minerals Rent Resource tax?
10. What are the main subsidies for not-for-profit organisations? Do they distort the allocation of resources? Explain your answers.

Further Reading

Abelson, P. and Joyeux, R. (2007) 'Price and efficiency effects of taxes and subsidies for Australian housing', *Economic Papers*, 26, 147–169.

Australian Treasury (2006) *International Comparison of Australia's Taxes*, Chapters 8 and 9, Treasury, Canberra.

Ergas, H., Harrison, M. and Pincus, J. (2010) 'Some economics of the mining sector', *Economic Papers*, 29, 369–383.

Freebairn, J. and Quiggin, J. (2010) 'Special taxation of the mining industry', *Economic Papers*, 29, 384–396.

Henry Tax Review (2010) *Australia's Future Tax System*, Part 2, Volume 1, 205–292, Australian Treasury, Canberra.

Kaplow, L. (2008) *The Theory of Taxation and Public Policy*, Chapter 8, Princeton University Press, New Jersey.

Productivity Commission (2010) *Contribution of the Not-for Profit Sector*, Research Report, Canberra.

Rosen, H.S. and Gayer, T. (2014) *Public Finance*, 10th ed., Chapter 21, McGraw-Hill, New York.

Part 10

Multi-Government Systems

Chapter

33

Multilevel Government

In great centralised nations the legislator is obliged to give a character of uniformity to the laws, which does not always suit the diversity of customs and districts.

De Toqueville

Assignment of Functions in a Multilevel System ♦ Optimal Size of Sub-Central Governments ♦ Taxation with Multilevel Government ♦ Intergovernmental Transfers ♦ Multilevel Government in Australia

In nearly all countries, government functions at several levels. Even in unitary countries, such as China, France and the UK, there are central, state or regional and local governments. Table 33.1 shows the numbers of governments by tier in two federal countries (Australia and the United States) and two unitary countries (France and the UK) in 2006.

There are significant differences between unitary and federal nations. In a unitary nation, the central government determines the powers and functions of all lower levels of government. It also allocates most public expenditure. In France and the UK, the central governments allocate about 80 per cent of all direct government expenditure—this is all government expenditure other than transfers between governmental units.[1] In a federation, the states that form the federation establish the powers and functions of the central government, usually in a constitution, and generally reserve significant powers for themselves. In the federations of Australia and the United States the central governments allocate about half of total direct government expenditure.

Table 33.1 Tiers of government (2006)

Country	*National*	*State and regions*	*Local*
Australia	1	6 states, 2 territories	673 (560)[a]
France	1	22 regions, 96 departments	36 679
United Kingdom	1	3 devolved state governments	367
United States	1	50 states	87 849

(a) Number of local authorities in 2011 after amalgamations since 2006 (source: Centre for Local Government, University of New England). There have been some further amalgamations in NSW since then.

(b) Source: Twomey and Withers (2007).

[1] The proportion of total direct government expenditures made by the central government is known as the centralisation ratio. It is a common measure of the extent to which a system is centralised.

Nevertheless, similar issues arise in unitary and federal countries. Which level of government can best conduct the various functions of government? What tax powers should be allocated to each level of governments? How should financial relationships between governments be managed?[2]

These issues are discussed in this chapter. We start by discussing how functions should be allocated to different levels of government and the implications for the size of sub-central governments. We then discuss the allocation of tax powers to different levels of government and the related issue of intergovernmental transfers. The last part of the chapter discusses some elements of the Australian multi-government system.

Assignment of Functions in a Multilevel System

As we have seen, the major economic functions of government are macroeconomic management, provision of social welfare, equity in distribution, provision of public goods, policies for other market failures and economic growth.[3] We discuss below the preferred assignment of these functions to each level of government.

Macroeconomic management. In the short run, national output, employment and prices depend principally on the level of aggregate demand relative to the capacity of the economy. Insufficient expenditure leads to unemployment. Excessive expenditure creates inflation. The macroeconomic task is to align aggregate expenditure with economic capacity. Using fiscal policy, government may adjust either its own expenditure or tax revenue to achieve the desired level of aggregate expenditure. Using monetary policy, government determines the money supply or its price (interest rates) to bring about the desired level of expenditure.

Clearly the central government or its agent, a central bank, should control monetary policy. Multiple currencies would be inefficient. Only one authority can control the supply of money or interest rates. Sub-central governments cannot be allowed in effect to print money to finance their services at the expense (via inflation and changes in the exchange rate of the currency) of citizens in other areas.

Central government should also be responsible for overall fiscal policy for several reasons. First, consistent national policy is desirable. If national and sub-central fiscal policies are inconsistent, the central government may and should take countervailing and overriding policies. Second, when resources are mobile, sub-central counter-cyclical policies are unnecessary. Third, most sub-central fiscal policies have limited local effect. Suppose that a regional government spends $100 million on a project, including $50 million on local factors of production of which 90 per cent would be otherwise employed in the absence of the project. The initial impact on local incomes equals ($100 million × 0.5 × 0.10) = $5 million. In addition, there may be a multiplier (M) effect, given approximately by:

$$M = 1/(1- MPCL) = 1/(MPT + MPS + MPI) \qquad (33.1)$$

where MPCL is the marginal propensity to consume local goods and MPT, MPS and MPI are the marginal propensities to pay tax, save and import to the local area for each local dollar earned. Allowing typical marginal propensities of say 0.3, 0.1 and 0.5 respectively, the regional multiplier is only 1.1 taking the total benefit to $5.5 million. Thus, sub-central pump priming is not very effective. These arguments explain why many central governments impose fiscal rules in the form of balanced (operating) budget requirements and constraints on borrowing on sub-central governments (see Sutherland *et al*., 2005).

---

[2] These are the classic issues in the literature known as 'fiscal federalism' (see Oates, 1999). As stressed here, these issues also apply in some degree in unitary countries.

[3] These functions derive from Musgrave's classic proposal that government activity should consist of three key functions: macroeconomic stabilisation, income redistribution and resource allocation (Musgrave, 1939 and 1959).

However, there is alternative view about fiscal responsibilities. If there are significant differences in local unemployment levels and labour mobility is low, regional government deficit-funded expenditure may be justified. As Gramlich (1987) observed, 'judicious use of budget changes by a sub-national government can reduce the variance in state (local) output around its long-run trend value'. If these budget changes materially affect aggregate demand, a central government (or bank) can neutralise the impact by altering interest rates. Thus, sub-central governments may be permitted to determine their levels of expenditure and tax without any significant loss of macroeconomic control.

Social welfare and distributional policies. Central government also has prime responsibility for provision of social welfare and distributional functions for two main reasons. First, only central government can ensure horizontal equity—that like individuals should be treated in a like way across the country. Sub-central governments cannot achieve this because they have different fiscal capacities. Second, substantial sub-central attempts to redistribute income may be counter-productive. A sub-central jurisdiction that taxes its affluent citizens heavily to redistribute extra income to its poorer citizens is likely to end up with few affluent citizens and more poor citizens (as the rich exit and unskilled workers immigrate). This occurred in New York in the early 1970s when the world's richest city nearly defaulted on its financial obligations as a result.

Notwithstanding the limited capacity of sub-central government (especially local government) to redistribute income, many communities wish to support the less able or isolated members of their community. Some support may have positive local externalities. Also, local communities are often better at perceiving local needs and more efficient at supplying them. Laubach (2005) points out that the delegation of substantial executive welfare responsibilities to the states in the United States (albeit financed mainly from Washington) has led to more innovative and effective design of welfare services and there has been little resulting emigration of resources from one state to another. Thus, sub-central governments generally do have a significant role in supporting poorer households subject to the constraint that redistribution does not distort business and household location.

Provision of public goods. Some public goods, such as defence and medical research, provide nationwide benefits. There can be only one level of provision and the efficient level would be determined by national willingness to pay the cost for an extra unit of the good.

Many other public goods, such as police, waste disposal services and recreational facilities, provide mainly regional or local benefits and can be provided in varying levels to different communities. Efficient provision of these goods requires that they satisfy local household preferences.

Typically, these preferences vary (there is demand heterogeneity). Suppose that two communities each have a population of 1000 and that 700 people in one group want a public good (a local park) with a cost of $\$X$ but that only 400 people in the other group want a local park at the same total cost. All others are opposed. If decisions are made by majority rule on a national basis, each group would have a local park; 1100 people would be pleased and 900 displeased. If separate decisions are made locally, one area would have a park and the other would not: 1300 people will be pleased and 700 displeased. Unless there are economies of scale, efficient quantities of local public goods should be provided regionally or locally.

Decentralisation theorem
Given differential preferences and no cost disadvantages, decentralized provision of local public goods increases welfare

This argument is formalised in the **decentralisation theorem**—if preferences for local public goods vary, and there are no economies of scale or externalities, decentralised provision increases welfare by equilibrating output to demand (Musso, 1998). Figure 33.1 shows a local public good produced at constant marginal cost. The demand curves D_A and D_B show aggregate demand for the local public good in two communities, A and B. The efficient quantities would be Q_A and Q_B respectively. If central government provided the same

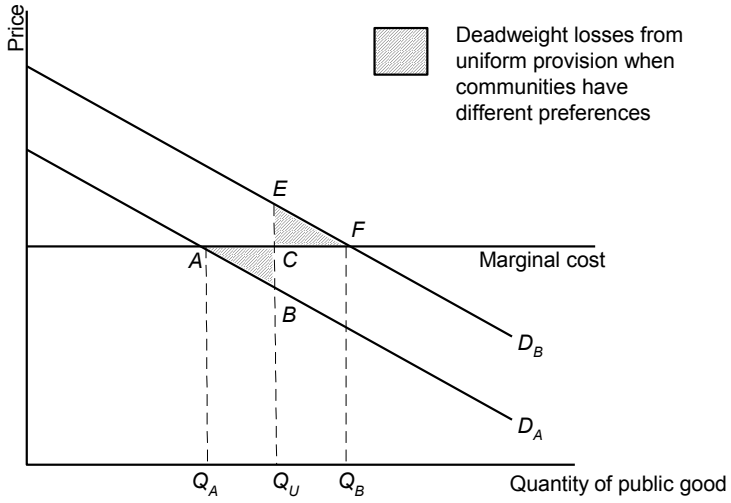


Figure 33.1 Deadweight loss due to uniform provision of a public good

output (Q_U) to the two communities, A would receive too much of the good and *B* too little. The welfare gains from decentralised provision would equal areas *ABC* and *CEF* respectively.

Other arguments for sub-central provision of local public goods are: sub-central government is usually better informed about local preferences; in small communities, costs are more transparent and easier to control; and competition between jurisdictions increases the efficiency of service delivery.

On the other hand, if there are economies of scale in service provision or spillovers between local governments, public goods may be provided more efficiently by large sub-central jurisdictions than by small ones. If economies of scale are extreme, there may be a case for national provision of the public good. Thus, the assignment of functions to levels of government involves principally balancing demand heterogeneity against economies of scale.

Other market failures. Public policies are also needed to deal with other market failures including imperfect competition, externalities and the environment, consumer and worker protection. Assigning policy responsibilities to levels of government follows similar principles to those just discussed. Where markets are national, responsibility would be national; where markets are regional or local, responsibility would be regional or local. Because many firms operate nationally across regional borders, responsibility for competition policy, trade practices and corporation law are usually a central government responsibility. One might also expect that consumer and worker protections would be national responsibilities, though in Australia they are often state responsibilities. On the other hand, many environmental issues are regional or local and would be regional or local responsibilities. However, climate change and national environmental heritage would be central government responsibilities.

Economic growth. Responsibilities for economic growth are again likely to be divided. In an open economy, where capital can move freely in and out, economic growth depends on the productivity of resources rather than on domestic savings. To promote growth, central government must encourage a competitive economy and free movement of capital and technology. Sub-central governments can influence growth by using resources efficiently,

avoiding excessive taxation, investing efficiently in infrastructure and human capital and protecting the productive environment. Indeed, sub-central competition can create economic growth. On the other hand, subsidies to uncompetitive local industry (including preferential local purchasing) or to foreign business seeking a local base divert resources from their most productive use and increase the cost of local purchases. Consequently, they are likely to reduce both national output and local income.

Summary. Central government has prime responsibility for macroeconomic policy, economic growth, welfare programs, regulation of national markets and international trade, and the supply of national public goods. Sub-central governments should be the main providers of all other public goods. This is sometimes described as '**the principle of subsidiarity**': subject to cost considerations, services should be supplied by the level of government that is closest to the users of the service. Sub-central government also has a role in providing some welfare services and in creating the infrastructure and conditions for economic growth

Subsidiarity principle
Public services should be supplied by the level of government closest to the users of the service unless this raises costs unduly

Unfortunately, application of these principles implies that two or more levels of government should be responsible for providing some services, especially where distributional issues and market failures arise. This occurs particularly in the provision of education and health services as well as with some welfare services. Dual provision of services may cause duplication, loss of accountability and increased cost.

Optimal Size of Sub-Central Governments

Given the functions of sub-central governments, what is their optimal size? This is sometimes modelled as a function of the optimal size for provision of local public goods. A common principle is that sub-central governments should be large enough to achieve economies of scale in service provision. However, ideally, the scale of facilities and community size would be optimised jointly.

Consider a hospital. For any given size of community, hospital size (measured say by number of beds) is efficient when the marginal cost of an extra hospital bed per member of the community equals the marginal benefit. A larger community requires a larger hospital. However, as hospital size increases, at some point the marginal cost of increased size exceeds the marginal benefit. Schedule *QQ* in Figure 33.2 shows the optimal hospital size for various community sizes.

On the other hand, for any given hospital size, there is an optimal community size. If a hospital has a fixed size, per capita cost falls with increased numbers in the community. However, service quality falls as crowding increases. For a given facility, the size of the community is optimal when the fall in the marginal quality of service with an extra member equals the marginal fall in per capita cost. In Figure 33.2 the *NN* curve shows the locus of optimal community sizes for various hospital sizes. At point *E*, both hospital and community size are optimal.

This model of community formation implies that optimal community size and service quantity may vary for each service, for health, police and fire services and so on. Assuming no joint costs or scope economies, there would be a large number of local authorities. This results even if individuals have similar preferences. However, separate local authorities for each service is not a practical cost-effective solution. Further, this is a simple model and we need to consider how other factors that might influence optimal jurisdiction size.

Benefits and costs of small jurisdictions

The benefits and costs of small jurisdictions mirror those for sub-central government generally. Small jurisdictions provide for diversity of preferences, are generally more responsive to local preferences and increase the opportunity for innovation and competition.

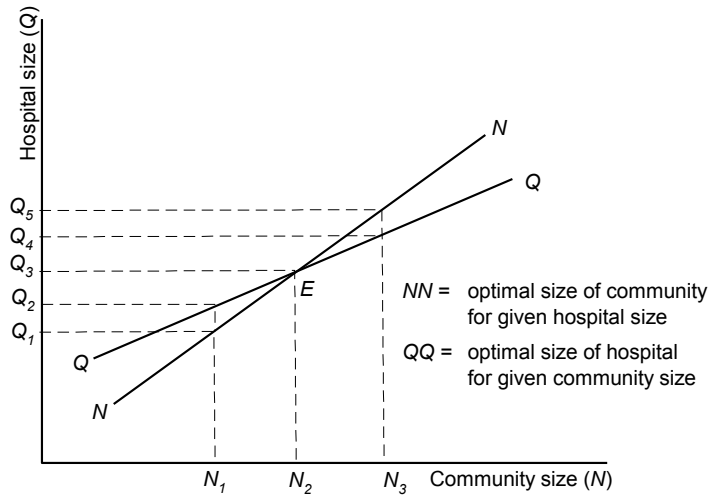


Figure 33.2 Optimal community size and service level

Members of a small community are more likely to be aware of how government spends their taxes and whether the administration is prudent and economical. Small jurisdictions may achieve cost savings by cooperation with other jurisdictions, for instance in provision of waste facilities, or by contracting out services, for example for rate collection. Box 33.1 provides evidence of the closer relationships between citizens and their representatives in small jurisdictions and the cost-efficiency of small jurisdictions.

However, small authorities may have disadvantages. Larger authorities may achieve economies of scale or scope in services and hence lower unit costs. This presumes that the technical advantages of size are not offset by bureaucratic costs and reduced accountability. And, in a dispersed population such as rural Australia, economies of scale in service provision may be achieved only at great inconvenience and travel cost for consumers. Spillovers are another possible disadvantage of small jurisdictions. Goods with positive spillovers to neighbouring communities, for example schools and libraries, may be under-supplied. On the other hand, local government may ignore externalities that harm adjacent communities, such as waste disposal into water systems.

Also, in Australia, it is often argued that state government administration is more effective when dealing with fewer local councils, especially in provision of infrastructure and land use planning.

Box 33.1 Benefits of small local government jurisdictions

In a study of local government in Sydney, Abelson (1981a) found that households in small local government areas received significantly more services from their elected representatives than did households in larger areas.

Abelson (1981b) found that expenditure per capita was higher in low-density areas and in areas with higher rates of population growth. However, there was no evidence of economies of scale. Expenditure per capita did not fall with the size of population.

In a more recent study of local councils in Sydney, Abelson (2015) found differences in expenditure per head were explained by differences in income per head, not by economies of scale.

Abelson (2016) concluded that local preferences are likely to be much better understood and served in smaller governing units and that there was no evidence that larger councils produce significant financial savings.

Evidently, optimal authority size depends on many factors. Diversity of preferences points towards smaller authorities. The potential for economies of scale point to larger ones. Ideally, the benefits and costs of alternative size authorities for each service would be quantified. However, in Australia there have been few studies of comparative service quality provided by different local jurisdictions.

Studies of the costs of local services are more common, but often have limitations. Consultants typically use a production function approach – they estimate the inputs needed to produce services, such as road maintenance, in authorities of various sizes and the associated costs of these inputs. This approach typically finds economies of scale. However, this approach overlooks the x-inefficiency behavioural costs of large units. Statistical studies of expenditure typically use multiple regression analysis to analyse whether per capita expenditure varies with size of jurisdiction. However, expenditure is the product of service levels and unit costs. High expenditure may reflect high service levels as distinct from high unit costs. Thus, it is important to include variables, such as per capita income, that may influence service levels separately from costs.

More discussion of these and related issues can be found in Abelson (2016).

Household choices and local government

In a classic article, Tiebout (1956) argued that a competitive system of local governments could provide an efficient outcome where households would choose the community that provides their desired package of services and taxes and would move if another community offers a preferred package. The Tiebout model (sometimes described as 'voting by the feet') implies that local public goods can be provided as efficiently as private goods in perfectly competitive markets. The equilibrium outcome is Pareto efficient—no one could be made better off by moving to another jurisdiction.

However, the Tiebout model is based on some critical assumptions.

- There are many competing local communities providing a mix of local goods and taxes that households want.
- Households are mobile and can relocate without cost to their preferred community.
- The cost per unit of public service is constant. If there are economies of scale smaller communities may not be able to offer an attractive local package.
- There are no neighbourhood spillovers.
- The costs of local public services are shared equally across all households. There are no cross-subsidies. If services are funded by a proportional property tax rate, the local community must be able to enact exclusionary zoning laws that prevent low-income households entering the community, which would lower the average property price and drive up the tax rate.

Some of these conditions occur in the United States where there are a large number of local jurisdictions and households are relatively mobile. Moreover, some outcomes predicted by the Tiebout model occur. In many US suburbs, households form broadly homogeneous social groups with similar preferences for local public goods. There is tax competition between local jurisdictions and limited local redistribution of income. Also, large differences in public services and tax prices are capitalised into property prices, which is a necessary condition for household equilibrium.

The Tiebout conditions apply less in Australia. There are only eight states and territories and household movement between them is expensive. Around 500 local authorities supply less important services and have little capacity to provide significantly differentiated packages of goods and taxes. The quality of local public services would be a minor factor in most household location choices.

Notwithstanding its limited actual relevance in Australia, does the Tiebout system provide a desirable policy model? If output can be produced at constant unit cost and there are no spillovers, the model provides a guide to efficient supply. A large number of competing local jurisdictions would provide efficient choices over local public goods.

On the other hand, the requirement of the model that a community can exclude households who would pay lower shares for public goods has significant equity implications. Communities would have an incentive to exclude low-income households. If local expenditure rises with population and local taxes are shared in relation to property values, the tax burden of existing residents rises when low-income households enter. Communities could prevent development of low-priced housing by restrictive zoning. The Tiebout model does not address these equity and efficiency issues.

Taxation with Multilevel Government

The fundamental problem with the allocation of taxes to different levels of government is that we have four main objectives but only three main tax bases, namely income, consumption and wealth. Following standard tax principles, the allocation of taxes to the different levels of government should satisfy four main objectives for good taxation:

1. Distributional objectives;
2. Efficiency—the allocation would minimise deadweight losses;
3. Fiscal adequacy—the allocation would provide adequate funding for each level of government;
4. Accountability.

It is difficult to satisfy all four objectives. It turns out that if tax bases are allocated to satisfy distributional and efficiency objectives, most taxes would be allocated to central government. Consequently, fiscal adequacy and accountability would not be achieved.

To achieve fiscal adequacy, the different levels of government may have to share a tax base. If they do not do so, inter-governmental transfers will be required. In this discussion we start from the presumption that governments should have separate tax bases and consider later some implications of shared tax bases.

Fitting taxes to economic functions

The core redistribution function of central government implies that central government should have access to the income tax. As discussed in Chapter 30, this is by far the most effective tax for income redistribution purposes. Taxes on commodities are typically regressive and a poor instrument of redistribution. Targeting luxury goods encourages commodity substitution. Taxes on wealth are usually too limited to achieve a general redistribution function.

Central government should also control corporation tax. This tax should be broadly aligned with personal tax rates to minimise substitution of corporate income for personal income. Centralisation of corporation tax economises on administration costs and minimises tax competition between sub-central governments that can lead to inefficient location of investment.

The macroeconomic functions of central government have slight implications for the allocation of taxes. The rise and fall of tax revenue over the business cycle provide natural stabilisers of the economy. And arguably, monetary policy is the main short-run instrument for macroeconomic policy. If active fiscal policy is used to manage short-run aggregate demand, it is generally easier and more effective to modify public expenditure than to make structural tax changes. Most tax changes, such as changes to income or consumption taxes,

have slow and uncertain macroeconomic impacts. Frequent changes create uncertainty and compliance problems.

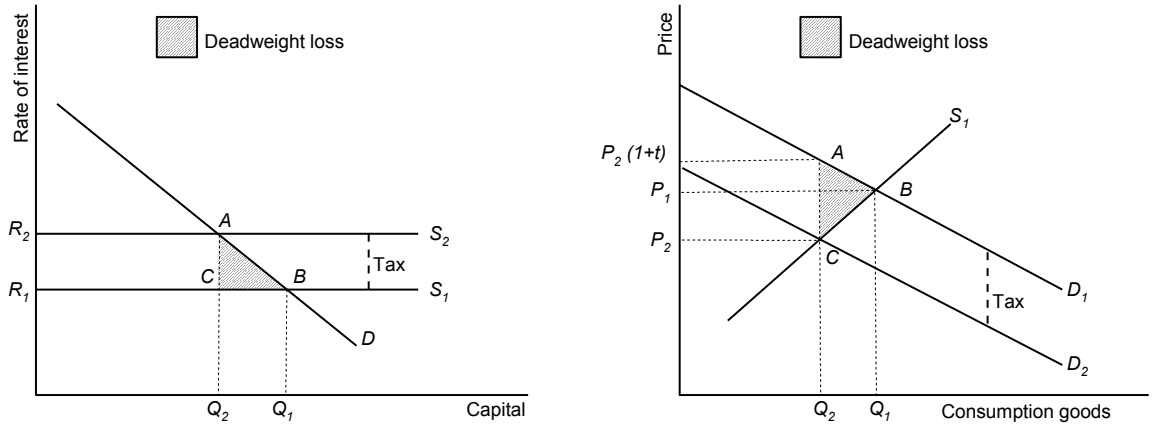
Turning to sub-central provision of local public goods, user charges are an efficient form of revenue for excludable services. However, they cannot be applied to non-excludable public goods and they often provide only a small part of total revenue needs. Given that income and corporation taxes are likely to be primarily central taxes, the issue is whether public goods can be funded by taxes on consumption or wealth?

Sub-central taxes and efficiency

There are two major problems with most sub-central taxes: the mobility of the tax base and spillovers. When tax bases are mobile, or when spillovers occur, decentralised taxes on production or consumption create more DWL than does central taxation. To minimise DWL, an efficient sub-central tax should not change behaviour. This means that it should be levied on relatively immobile factors of production or immobile consumption.

Figure 33.3 illustrates the DWL of sub-central taxes on capital and on consumption. Capital is highly mobile and supply elastic. Panel (a) shows the effect of a local tax on capital that shifts the supply schedule from S_1 to S_2. The demand for capital does not shift. The DWL is the triangle *ABC*. Panel (b) shows the effect of a local tax on consumption that shifts the demand schedule from D_1 to D_2. The supply curve is unchanged. The DWL is again triangle *ABC*. Local firms bear a significant cost as consumers switch purchases to other jurisdictions to avoid the local tax. These DWLs do not occur with a national corporate or consumption tax.

On the other hand, land is an efficient local tax base because land is immobile and the total supply is inelastic. Land tax cannot be escaped and is a reliable tax base. Therefore, a moderate general *ad valorem* tax on land has no DWL.[4] Land tax is easy to administer and helps to ensure fiscal adequacy for sub-central governments. Taxation of property inclusive of structures and land is also quite efficient. Existing structures are immobile. However, if structures are taxed, new capital may be invested in non-land assets or in other areas.



(a) Sub-national tax on mobile capital

(b) Sub-national tax on consumption goods

Figure 33.3 Local taxes on capital and on retail goods

[4] A unit tax could cause a landowner to abandon his land. Also a very high *ad valorem* tax could destroy the price mechanism for land and selective taxes on land can distort the allocation of land.

Tax competition. A common consequence of decentralised taxation is tax competition (a lowering of tax rates) between jurisdictions to attract resources. Tax competition encourages diversity of administrative practices, benchmarking of costs and efficient delivery of services.

However, tax competition can lead to inefficient outcomes. Suppose that two jurisdictions (*A* and *B*) levy a tax on capital, which can move freely between the areas to maximise the after-tax return. To increase local economic activity and its tax base, area *A* may reduce its tax rate. This reduces the tax base in *B*. If *B*'s tax revenue falls by more than expenditure due to the loss of industry, there is a negative externality. Most likely, when setting its tax rate area, *A* ignores this negative externality. Such tax competition may lead to under-taxation of a mobile tax base and under-provision of public goods.[5]

Tax exporting, shifting part of the tax burden on to firms or households in other areas, is another possible source of inefficiency in sub-central taxation. Tax exporting occurs when firms or households can deduct local taxes from taxable income for central government income tax as in the United States. In Australia, taxes on commercial properties are deductible income tax expenses. In effect, local services are funded partly by central government. Tax exporting enables local populations to obtain services for less than 100 per cent of their marginal cost and thus encourages provision of services beyond the efficient level.

Many studies have found that the supply of capital and labour and the location of consumption expenditure are sensitive to local taxes (Inman and Rubinfield, 1996). Chirinko and Wilson (2008) show that US state tax incentives have a significant impact on private investment but is a zero-sum game (there is little effect on total investment) when states reduce taxes in similar proportions. Day (1992) and Tretze *et al*. (1993) showed how labour shifts between regions in response to tax-related changes in Canada and the United States respectively. Wales (1968) showed how small consumption tax differentials influenced cross-border shopping. General equilibrium tax models have shown that these shifts can have significant DWLs (see Jones and Whalley, 1988, for Canada; Morgan *et al*., 1989 for the United States).

Australia has experienced the effects of tax competition. In 1977, to attract more retirees the Queensland government abolished state death duties. In the following years all other states followed suit. In the late 1990s, to attract more share transactions the Queensland government again led the way in reducing stamp duties on share transactions. NSW and Victoria followed within a year. However, tax competition is not necessarily inefficient if it leads to abolition of inefficient taxes or brings prices closer into line with marginal costs.

Vertical fiscal imbalance: fiscal adequacy and accountability

Evidently, central government is the most efficient taxer of the two fundamental tax bases (income and consumption). However, this creates **vertical fiscal imbalance** (VFI). VFI occurs when one level of government, almost always central government, raises more revenue than it needs for its own purposes and another level of government raises only part of what it needs.

Vertical fiscal imbalance
Occurs when different levels of government raise more or less revenue than they need

VFI encourages over-spending, especially by a central government which has excess income and faces weak fiscal constraint. Sub-central governments may also perceive the marginal cost of local services to be low because most revenue is supplied externally and supply excessive local public goods. In the words of James Guy, 'A principle which permits one government to raise money to be expended by another is unsound and is likely to induce a feeling of irresponsibility on the part of the expending authority'.[6]

[5] Similar inefficiencies can arise when jurisdictions use subsidies to attract industry from another jurisdiction.

[6] James Guy (Liberal, Wilmott, Tasmania) during the second reading of the 1946 States Grants legislation.

Another issue is the gradual take-over of services and policies by the central government and the erosion of local powers. Central government may place conditions on the use of grants to sub-central governments and the provision of local public goods that are inconsistent with local preferences.

There are contrary views. Williams (2005) argues that, if central grants to sub-central levels are fixed, sub-central government must raise the marginal tax required to pay for its services. Thus, sub-central government must determine the marginal cost and benefit of its service. Government would be influenced by voter preferences via competition between political parties.

Conclusions on efficiency

These general principles suggest that central government should tax personal income, mobile factors of production (notably capital) and consumption. Sub-central jurisdictions should tax bases with low inter-jurisdictional mobility, notably land and natural resources.

To minimise DWL, government should tax bases with a low sensitivity to changes in tax rates.[7] This implies that sub-central taxes should be based on residence of the worker or consumer rather than on the source of the income (the place of work or consumption).[8] Residences are less mobile than production and consumption. On the other hand, source-based taxes are often more practical than resident-based taxes.

The classic resident-based tax is the property tax, especially the land component. A tax on the capital component of property may cause capital to shift. Income tax is another potential resident-based tax. A sub-central income tax, with a locally chosen tax rate, can be 'piggy-backed' on to a central government income tax. A resident-based consumption tax would require the tax authority to have information on the total consumption of households, wherever purchases are made. In the absence of a general expenditure tax, this is impractical.

On the other hand, source-based taxes on local labour or consumption are practical and, for this reason, often adopted. These taxes are efficient in so far as the labour or consumption are not mobile. Usually local labour is less mobile than consumption. Thus, sub-central payroll taxes have less deadweight loss than sub-central consumption taxes.

However, if sub-central government cannot raise sufficient revenue to service local needs, the tax system does not provide fiscal adequacy. Fiscal adequacy can be addressed by transferring expenditure functions to central government or by allocating more tax powers (or a share of tax powers) to sub-central government. If neither strategy is practical, fiscal imbalance has to be resolved by inter-government transfers.

Equity issues

There are two major equity issues with sub-central taxes. First, sub-central taxes are often regressive—taxes rise with income but less than proportionately. Several studies have shown that local taxes in the United States are regressive (Inman and Rubinfield, 1996). To retain their tax bases, sub-central governments must be attractive to higher income households. Although U.S. and Canadian states have the power to tax income they rarely do so in a strongly progressive way. However, a fundamental tax principle is that tax systems should be judged as a whole. The regressive nature of state taxes is not a problem if the central government takes responsibility for redistribution through the personal income tax.

---

[7] The elasticity of the tax base is the ratio of the percentage change in the tax base attributable to a given percentage change in the tax rate applied to the base.

[8] Resident-based taxes are sometimes called destination-based taxes. Source-based taxes are sometimes described as origin-based taxes.

Secondly, sub-central governments have unequal capacity to provide public services. This is known as **horizontal fiscal imbalance** (HFI). HFI arises because of differences in expenditure needs or revenue-raising capacity (or both). High-income areas can provide higher quality public goods such as schools than low-income areas. Expenditure needs reflect differences in household needs and in the unit costs of services. However, where HFI is a barrier to an equitable supply of local public goods, the solution lies in inter-government transfers rather than in reallocation of tax bases.

Horizontal fiscal imbalance
Occurs when sub-central governments have unequal fiscal capacity to provide public services

Taxation and capitalisation. The argument that sub-central government creates horizontal inequities between households must be treated with caution. If property prices capitalise the value of public services, households receiving fewer services are compensated via lower house prices.

Suppose two communities, *A* and *B*, provide an equal amount of local public goods. *A* contains mainly high-income households and has a low tax rate on property (or on another tax base). *B* has mainly low-income households. To obtain the same tax revenue as *A*, community *B* must levy a higher tax rate on its tax base. If house prices were similar in the two communities, households would move from *B* to *A* and enjoy the lower local tax rate in *A*. Equilibrium would occur only when house prices were sufficiently higher in *A* than in *B* to offset the benefit of the lower tax rate. A similar argument applies if tax rates are the same in the two communities, but area *A* obtains more revenue from its tax rate and provides more public goods than *B*. Again, higher house prices in *A* will offset the greater supply of public goods. In equilibrium, differences in house prices fully offset differences in local taxes or services. There is then no horizontal inequity between similar households due to differences in the supply of local public goods.

This conclusion holds regardless of any attempts by a community to restrict the supply of housing or to exclude low-income households. Exclusion causes house prices to rise to a new equilibrium level. Existing landowners receive a windfall gain, but entrants to the community would pay the higher property price to gain the lower tax rate benefit (assuming no economies of scale in service provision). However, full capitalisation of local taxes and services assumes that households can move between communities without cost in response to differences in taxes and public spending. If there are constraints on household movements, differences in taxes and public services may not be fully capitalised into house prices.

Many studies have found evidence of capitalisation in the United States. Bloom et al. (1983) concluded that 'interjurisdictional property tax variations are between halfway and fully capitalised into house values'. Yinger et al. (1988) also found extensive capitalisation of differences in local property taxes. Other studies have found that property values are higher in school districts with higher per pupil expenditure and higher scores on achievement tests (e.g. Jud and Watts, 1981). Capitalisation doubtless occurs in Australia, but local government provides fewer services and has less impact on household location than in the United States.

Taxing the same tax base

VFI and its corollary, intergovernmental grants, can be avoided if different levels of government can levy taxes on the same tax base. In the United States, the states can choose almost any tax base other than national imports or exports or inter-state commerce. Thus, in the United States, as in Canada, the central government and the states share the income tax base, including personal and corporate taxes. And in the United States both the states and local government tax sales.

When different levels of government tax the same base, they usually have a common definition of the base but may vary the tax rate. The Canadian government requires the provinces to adopt a uniform definition of taxable income but allows the provinces to adopt a variety of tax credits and different tax rates. The uniform definition of the tax base minimises

administration, compliance and other costs that would arise from different definitions of the base. But different tax rates give the provinces flexibility in administering the income tax. As a proportion of the federal tax rate, the personal income tax rate is much higher in the poor state of Newfoundland than in the rich state of Alberta.

Theory suggests that when two levels of government share a tax base, both tax rates and total public expenditure are likely to rise, but these outcomes are not certain. If one level of government raises the tax rate on, say, an income tax base, the base declines with a fall in labour supply. To maintain its revenue, the other level of government must now raise its tax rate on this base. However, in determining its tax rate one level of government does not consider the effect on the tax base of other levels of government. It is therefore likely to raise the tax rate by more than if it were sole owner of the tax base. From a national perspective the tax base is over-used. On the other hand, if central government increases its expenditure when it raises the tax rate, this increase in expenditure may enable sub-central government to reduce its tax requirements and its tax rate.

Evidence on the effects of shared tax bases is limited and mixed. Besley and Rosen (1998) found that when the US government raised taxes on tobacco and petrol, the states also raised taxes. This suggests over-use of the tax base. On the other hand, in a study of 13 OECD countries Godspeed (2000) found that a one percentage point increase in national government income tax rates led to a fall of about 0.17 of a percentage point in sub-central income tax rates. This does not necessarily mean that the tax base is under-used. However, Dahlby and Wilson (1996) argue that sharing tax bases has produced significant tax distortions in Canada.

Sharing a tax base also creates difficulties for equity unless tax rates are harmonised. In 1962, to ensure equity and to control the progressivity of the rate structure the Canadian government required provinces to levy the personal income tax as a percentage of the basic federal income tax. However, over time the provinces were allowed to add tax credits for low-income earners and surcharges to the schedule and, since 2003, to set their own tax brackets and rates. Unless there is full capitalisation of provincial tax differences this could produce significant horizontal inequity across Canada.

Fiscal rules

Finally, it should be noted that some central governments also impose fiscal rules (applying to budgets overall) on sub-central governments. Drawing on a survey of OECD countries, Sutherland *et al*. (2005) found that many OECD governments adopt fiscal rules at the sub-central level to avoid or control unsustainable levels of public spending or debt. The most common fiscal rule is the budget balance requirement (usually for the operating budget). There may also be restrictions on borrowing. These rules are sometimes complemented by limits on tax rates. However, sub-central governments generally retain some tax autonomy. Rules-based controls are rarely applied to expenditure *per se*.

These fiscal rules are common in unitary countries. For example, in the UK, local authorities have very limited tax and expenditure autonomy (Sutherland *et al*., *ibid*. Box 2, p. 13). However, as we saw in Chapter 29, all the states in the United States (except Vermont) also have their own constitutional requirements to balance the operating budget.

Intergovernmental Transfers

Intergovernmental transfers take two main forms: revenue sharing or intergovernmental grants. Revenue sharing provides sub-central governments with a pre-determined share of central government revenue. For example, in Germany the constitution provides for the federal and state governments to share the yield of the major taxes. In a full revenue-sharing model, central government has no power to determine the proportion of revenue to be distributed to lower levels of government or how the distributed revenue is spent.

In practice, most central governments usually have discretion over the amount of revenue provided to sub-central governments. In Australia, Commonwealth legislation requires the Commonwealth government to pass all GST revenue in general grants to the states and territories. However, the Commonwealth determines the distribution of the revenue between the states and territories. Intergovernmental grants are of two main kinds.

General purpose grants. A recipient government can spend general purpose grants as it chooses. No conditions are attached to the grant. These grants are also described as untied, unconditional or block grants.

General purpose grant
A grant with no conditions attached

Specific purpose grants set conditions on how the grant may be spent. The conditions, including policy conditions, may be set out in considerable detail. These grants are also known as tied, conditional or categorical grants.

Specific purpose grants may be matching or non-matching grants. **Matching grants** require the recipient government to provide a contribution, often on a dollar for dollar basis, for the grant. **Non-matching grants** require no local contribution. We discuss below the aims and effects of these various kinds of grants.

Specific purpose grants
A grant with conditions attached

Matching grant
A grant which is conditionalon the recipient providing a specified contribution

General purpose grants

General purpose grants have two main aims: to compensate for vertical and horizontal imbalances (VFI and HFI). VFI refers to the general revenue shortfalls of sub-central governments. HFI reflects differences in the fiscal capacities of sub-central governments to provide to public goods. HFI is a function of local expenditure needs and revenue-raising capacity.

Expenditure needs reflect (1) the needs of households of various types and (2) the unit costs of public services. Some households, such as indigenous, non-English speaking or the frail aged, have high needs that may not be reflected solely in income levels. Unit costs may be relatively high because of the conditions of production, such as distance from sources of materials or high labour costs, or because of inefficiency and low productivity. Ideally fiscal equalisation would compensate for unavoidable higher unit costs but not for inefficiency.

Sub-central revenue-raising capacity depends primarily on the disposable (after income tax) income of resident households and businesses and to a limited extent on the capacity of sub-central government to tax non-resident businesses or households. Tax revenue is itself an unreliable indicator of tax capacity because revenue reflects tax effort as well as fiscal capacity. High revenue may reflect high taxes rather than high disposable household incomes.

Effects of general grants. A general purpose grant increases the income of the recipient community. It does not affect relative prices and has no substitution effects. A $10 million grant may be used to increase local public expenditure by $10 million or to reduce local tax revenue by $10 million.

Figure 33.4 overleaf illustrates possible effects of a general purpose grant. This shows consumption of public and private goods on the horizontal and vertical axis respectively. There are two budget lines. The line *AB* shows the initial community budget. The higher line *CD* shows the budget line with the general purpose grant. The convex curves represent community utility (indifference) curves. The initial equilibrium is at point *E*. After the grant, the community moves to point *F* on indifference curve I_2.

Now suppose that the community increases its income to the budget constraint *CD* by increasing export income. If an extra dollar of private income increases public spending by 10 cents rather than by 40 cents, additional income would result in a move from *E* to *G* and the community would achieve a higher level of welfare represented by curve I_3.)

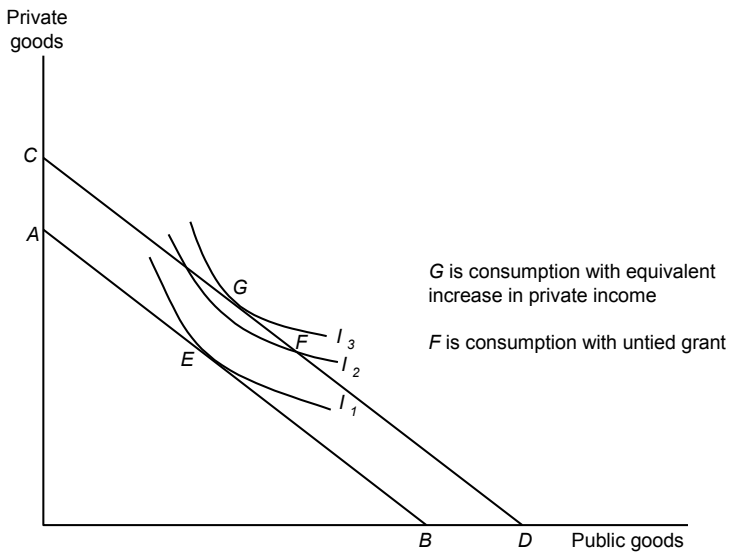


Figure 33.4 Effect of general purpose grants

It is often observed that public consumption increases more with an untied grant than with an independent increase in community income. US studies indicate that a dollar increase in general grants generates 40 cents of public spending while a dollar increase in private income increases public spending by only 10 cents (Rosen and Gayer, 2014, p.524). This is described as the 'flypaper effect'. Money tends to stick where it lands (in the hands of sub-central governments).

Dixon *et al*. (2002) estimated that the flypaper effect has a welfare cost of $150 million a year in Australia. Based on an average allocation of resources to public goods, the states most heavily subsidised by the Commonwealth (South Australia, Tasmania and the Northern Territory) make spending decisions out of line with household preferences. South Australia and Tasmania allocate an exceptionally high proportion of resources to public goods. The Northern Territory provides a normal proportion of public goods despite high unit costs.

The use of general purpose grants for equalisation purposes (to deal with HFI) creates another potential efficiency cost. Equalisation of access to public services requires that additional resources are allocated to high cost areas. This reduces the total quantum of services provided nationally. Moreover, if the grants encourage firms and households to locate or stay in high-cost areas because they do not bear the full costs associated with the location decision, there is an additional DWL due to the inefficient location decisions.

Specific purpose grants

Tied grants aim to ensure that sub-central governments provide the quantity and quality of public goods that they might not otherwise supply. This usually reflects the desire of central government to influence policy or service delivery. This may be viewed as reasonable where central government has prime responsibility for the service and the role of the sub-central government is principally to deliver the service.

The effects of tied grants depend partly on whether they are matching or non-matching grants. We consider first the effects of a non-matching tied grant. In Figure 33.5a the horizontal axis represents consumption of a specific public good and the vertical axis shows consumption of private goods and all other public goods. Receipt of a non-matching grant (*AC*) tied to the specific public good shifts the community's budget constraint from *AB* to

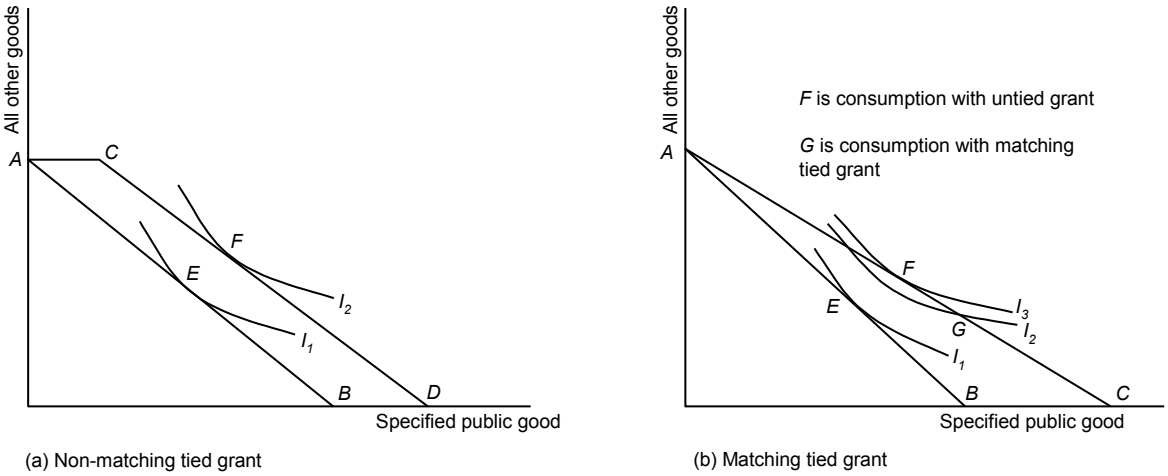


Figure 33.5 Effects of non-matching and matching tied grants

ACD. Importantly, if the non-matching grant is less than the community's desired expenditure on this public good, the grant is effectively an increase in general community income and would have the same effect as a general block grant.

Even if the tied grant exceeds initial expenditure on the local public good, the resultant expenditure may be the same as would occur with an untied grant of similar size. However, if many conditions are attached to the tied grant, they may change how the local public good is delivered.

Panel (b) illustrates the impacts of a matching tied grant. This grant changes the budget constraint from *AB* to *AC*. This changes relative prices and creates a substitution effect as well as an income effect. If a grant is matched on a dollar for dollar basis, the price of the local service is halved and the local community receives twice the output for each marginal dollar it spends on that service. A matching tied grant increases the incentive for the recipient to provide the specified service and is more likely to increase consumption of a specific local service than is a non-matching grant of a similar size. However, even in this case the recipient may increase its expenditure on other local services or reduce taxes, so that consumption of the specified service may shift from *E* to *F*, with little increase in local expenditure on it.

Tied grants often have DWL, especially matching tied grants. The greater the change in consumption from that which would occur with an untied grant of similar size, or from that which would occur with an equal increase in private income in the community, the greater the DWL of the tied grant.

Multilevel Government in Australia

In Australia the central (Commonwealth) government collects over 80 per cent of tax revenues, which gives it a dominant role in nearly all public expenditure and policy decisions. The Commonwealth is the only level of government to levy personal and corporate income tax and sales taxes (including the goods and services tax). Although the states have the constitutional power to levy personal income tax, they cannot do so without the agreement of the Commonwealth, which has the power to withhold other revenue from the states. In 1997, the High Court of Australia ruled that the states do not have the constitutional right to levy sales taxes.

The Commonwealth provides national public goods, such as defence and national economic infrastructure, and social security benefits. Under section 96 of the Constitution, the Commonwealth can also grant financial assistance to any state 'on such terms and conditions as the Parliament thinks fit'. In 2006, the High Court of Australia ruled that the Commonwealth can control not only the activities of corporations but also their relationships with their employees or contractors, thus giving the Commonwealth power over industrial relations.

However, the states and territories are the main suppliers of public services to the community, including law and order, primary and secondary education, most health services, utilities and transport services. Local government, which is a creation of the states, provides mainly local services, such as local roads, parks, libraries and waste services, along with local planning and building controls.

Because of its control over the major tax bases the Commonwealth government raises far more revenue than it spends on its own programs. Although the Commonwealth collects over 80 per cent of all tax revenue, it is responsible for only half of government direct expenditure. By contrast, state and territory governments raise about 15 per cent of tax revenue but account for some 45 per cent of government direct expenditure. This is a high degree of VFI, although not entirely exceptional (Twomey and Withers, 2007).

Consequently, intergovernmental grants have a large role in Australian public finance. Table 33.2 shows the Commonwealth's budgeted grants to the states and territories in 2011–12. These totalled $93.9 billion, equal to 6.8 per cent of GDP. Of this total, about half represented GST revenue, which is distributed as general purpose grants. The distribution is based on the principle of horizontal fiscal equalisation. The Commonwealth Grants Commission (CGC) estimates the per capita expenditure required to provide an average standard of government services in each state and territory and the differences in the capacities of the jurisdictions to raise revenue. This currently has traditionally resulted in a substantial redistribution of GST revenue from Western Australia, Queensland, NSW and Victoria to the other states and territories.

The Commonwealth distributes the other 50 per cent in special purpose (SP) and national partnership payments. The main SP payments traditionally related to national health care, schools, skills and workforce development, disability services and affordable housing. Each set of SP payments is based on a national agreement that sets out the objectives, outputs and outcomes, performance indicators and roles of the Commonwealth and states. National partnership payments are financial contributions to the states to deliver specific projects, often with conditions and a requirement for matching funds.

Some issues in the Australian federal system

The allocation of public functions in Australia is broadly as would be expected, with central government primarily responsible for the macroeconomic and distribution functions of government and state and local governments responsible for the delivery of many public services. However, the exceptionally high level of VFI, due to the uneven distribution of taxation powers between central and sub-central governments, creates several problems.

Table 33.2 Commonwealth grants to the states and territories, 2011-12

Payments	*$ billion*	*% of total*
General purpose grants (GST revenue)	46.5	49.5
Special Purpose Payments	26.3	28.0
National Partnership Payments	21.1	22.4
Other general revenue assistance	1.1	1.2
Total	93.9	100.0

Source: Treasurer (2011) *Federal Financial Relations, 2011-12*, Budget Paper No 3.

One issue is a lack of fiscal accountability, principally of central government which has excess funds to distribute and total discretion over their allocation. These allocations often do not match the preferences of the recipient states and distort the way in which the states would otherwise direct resources. Also, in so far as states receive free money, they may lack expenditure discipline.

Second, large amounts of state bureaucratic resources are devoted to trying to capture the central government's excess revenues.

Third, of the limited set of state taxes, several are inefficient or inequitable. The taxes include payroll, land tax, stamp duties on exchange of assets (mainly property), motor vehicle taxes and taxes on both insurance and on gambling (risk averters and risk lovers both get taxed!). As Abelson and Joyeux (2007) show, stamp duties impose a significant DWL on property transactions. On the other hand, taxes on gambling are efficient in that the demand for gambling is inelastic, but highly inequitable. Land taxes and payroll taxes could be efficient if they were administered at a low rate across a broad base. However, they are often administered in a discriminatory and non-neutral manner with large areas of exclusion or reduced tax rates.

A fourth issue is the Commonwealth's distribution of the block grants to the states and territories. This is based on the recommendations of the Commonwealth Grants Commission (CGC). The CGC estimates the revenue that would allow each state and territory to provide an average level of public services as a function of expenditure needs and revenue capacity. Expenditure needs reflect differences in the needs for services (e.g. a high Indigenous need) and the unit costs of services due, for example, to differences in input costs or remote service provision. The CGC estimates revenue capacity principally as a function of the value of the state's tax base along with its mineral revenues.

There are concerns with objectives and process. A welfare objective of equalisation would normally apply to an individual's total welfare that depends on private income, costs of living and amenities as well as access to public goods. Equalising one element of welfare may reduce overall equality of welfare. Albouy (2010) showed that the large Canadian federal transfers to the Atlantic and Prairie states (other than Alberta) subsidise populations that have equal earnings ability to the populations in the states that are losing revenue. The inter-province wage differentials compensate for cost of living and amenity differentials. The transfers do not increase equity. Rather, they go to low productive areas which are subsequently over-populated.

The CGC process is complex and costly with reports running into thousands of pages.[9] More substantively, it is not clear how effectively the CGC distinguishes between high unit costs of services (which might be subsidised on equity grounds although this is an inefficient use of public resources) and inefficiency in delivery of services (which should not be subsidised). There are also problems with the CGC's calculation of revenue capacity, which is a function of the value of the tax base in each state rather than of disposable household income. The value of the tax base is an unreliable indicator of a community's capacity to raise revenue because it is not necessarily correlated with household disposable income (Abelson, 2011). NSW and Victoria have high value tax bases because property prices are high, but for many households in these states high house prices represent a cost and a reduction in real household income. The CGC methodology results in a large subsidy to ACT households who are the most affluent in the country and have no obvious disadvantages of service provision.[10]

Fifth, the specific purpose and national payment Commonwealth grants often create an overlap of responsibilities and inefficient duplication of services. In education the

[9] The observations here are based on practices up to 2012. At this time, we have not reviewed recent practices.

[10] The major reasons for this result are the lack of access to the payroll tax and alleged high unit costs of administering a small jurisdiction.

Commonwealth subsidises private secondary schools, university students and some tertiary vocational training. The states are responsible for public secondary schools and most tertiary vocational training and universities are run under state statutes. In the health sector, the central government is responsible for medical services, subsidised medicines and nursing homes. The states are responsible for community health, hospitals and independent housing for the frail elderly. This can lead to inefficient cost shifting.

Finally, it may be observed that Australia has a large number of local government authorities. Local governments fund their services mainly by a tax on land values along with user charges and grants from the Commonwealth and state governments. The state governments frequently threaten to reduce the number of local governments, which they have the power to do, arguing that small local governments are inefficient. However, most residents of most local areas strongly oppose any loss of autonomy. In the words of Montesquieu, 'In a small republic, the public good is more strongly felt, better known and closer to each citizen'. This appears to be a common view in Australia and is doubtless the case in many other countries.

Summary

- Nearly all countries must determine the size and number of governments, the allocation of functions and taxes to various levels of government and financial relationships between governments.
- Analysis of these issues draws on many core economic concepts, including public goods and externalities, the role of household preferences, the role of competition and the impacts of transfers on behaviour.
- Central government has prime responsibility for macroeconomic policy, provision of social welfare services, income distribution, regulation of national markets and national public goods. The major role of sub-central government is provision of regional and local public goods.
- Sub-central jurisdictions provide for diversity of preferences, increase competition and innovation and may be more cost conscious. However, inefficiencies may arise due to lack to scale or spillover effects.
- The Tiebout model suggests that with sufficient competition between local jurisdictions, households can move to obtain their preferred service and tax package. However, this is unlikely when choices are restricted and movement costs are high.
- To achieve distributional objectives, central government should tax income. Because of the mobility of consumption, there is also a strong case for central government to tax consumption.
- Sub-central governments should tax bases that have low mobility, notably land and natural resources. However, if sub-central government cannot raise sufficient revenue to service local needs, the tax system does not meet the criterion of fiscal adequacy.
- Sharing a tax base would achieve fiscal adequacy. However, common tax bases tend to be over-taxed.
- When one level of government collects more revenue than it needs, and another level collects less, the imbalance is corrected by intergovernmental transfers. This can be achieved by revenue sharing. Tied grants can have major distortionary effects. Even untied grants can be distortionary via the flypaper effect.
- The Australian federal system has high vertical fiscal imbalance. This gives the central government steadily increasing power over policy and expenditure decisions at all levels. The redistributive process through the CGC has some significant inequitable and inefficient features.

Questions

1. What role, if any, do sub-central governments have in income distribution?
2. What are the advantages and disadvantages of sharing an income tax base?
3. When is tax competition desirable or undesirable? Should sub-central governments compete to attract mobile equity capital?
4. How would you attempt to determine whether the size of local jurisdictions affects:
   i. the costs of services?
   ii. the quality of services?
5. How would you test whether differential tax rates of local jurisdictions are capitalised?
6. How would you determine the revenue needs of poorer sub-central governments?
7. When, if ever, should central government to attach conditions to grants to other levels of government?
8. Who should tax petroleum and mineral rents in a federation? And how should these rents be taxed?
9. Why may horizontal fiscal equalisation increase inequality of welfare?
10. What inefficiencies may result from horizontal fiscal equalisation?

Further Reading

Albouy, D. (2010) *Evaluating the Efficiency and Equity of Federal Fiscal Equalisation*, Working Paper 16144, National Bureau of Economic Research, New York.

Gruber, J. (2016) *Public Finance and Public Policy*, 5th ed. Chapter 10, Worth Publishers, New York.

Musso, J.A. (1998) 'Fiscal federalism as a framework for government reform', pp. 347–395 in *Handbook of Public Finance*, F. Thompson and M.T. Green (eds), Marcel Dekker, New York.

Oates, W.E. (1999) 'An essay in fiscal federalism', *Journal of Economic Literature*, 37, 1120–1149.

Rosen, H. and Gayer, T. (2014) *Public Finance,* 10th ed, Chapter 22, McGraw-Hill, New York.

Sutherland, D., Price, R. and Joumard, I. (2005) *Fiscal Rules for Sub-Central Governments: Design and Impact*, Economics Department Working Paper, No. 465, OECD, Paris.

Warren, N. (2010) 'The Henry Review, State Taxation and the Federation', *Australian Economic Review*, 43, 409–421.

Williams, R. (2005) 'Fiscal federalism: aims, instruments and outcomes', *Australian Economic Review*, 38, 351–369.

Chapter

34

Globalisation and Government

Whoever is the lord of Malacca has his hand on the throat of Venice.

Tomé Pires, 16th century Portuguese trader and diplomat

The Nature of Globalisation ♦ Globalisation, Trade, Output and Prices ♦ Distributional Consequences of Globalisation ♦ Globalisation and Government Revenue ♦ Globalisation and Government Expenditure ♦ Globalisation and Government Policies

Globalisation
The worldwide integration of markets

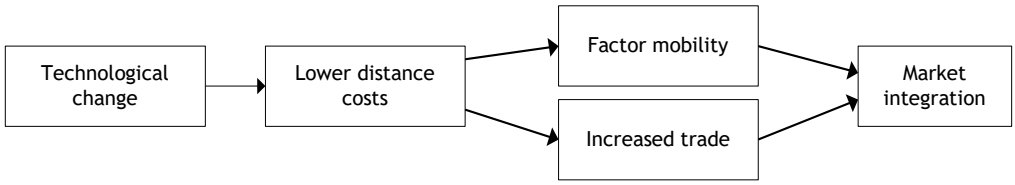
Globalisation is highly controversial. Proponents argue that the international flow of capital, labour and goods increases output and incomes in most countries. And they point to the extraordinary growth in world GDP over the last 30 years due in no small part to the growth in international movements of capital, labour and goods and the integration of China into the world economy. In the words of Benjamin Franklin, 'no nation was ever ruined by trade'.

On the other hand, critics blame globalisation for a variety of economic and political ills. They hold globalisation responsible in developed countries for major increases in inequality, falling real wages of many workers, and stresses on housing affordability, personal security and welfare services due to high immigration rates. Anger at globalisation led to the nationalist Brexit vote in the UK in 2016, the "America-First" election of President Trump in the US in 2017 and the rising tide of populist parties in European elections (and to a minor extent in Australia) in recent years. In developing economies, globalisation may be perceived as 21st century economic imperialism.

In this chapter, we start by describing the nature of globalisation. Section 2 discusses how globalisation affects trade, output and prices. Section 3 reviews the impact of globalisation on income distribution. Sections 4 and 5 discuss how globalisation influences government revenue and expenditure. The last part briefly discusses government policies.

The Nature of Globalisation

In economic analysis, **globalisation** means the worldwide integration of markets. In the end state, there would be one world market for both factors of production and goods rather than national or regional markets. Markets integrate as the costs of distance fall. Capital, labour and technology become more mobile and international trade in factors of production and goods increases. These events are driven primarily by technological change, notably recently by the worldwide digitisation of marekts. Thus, the causation runs:



However, globalisation is not simply a function of technical change. Government policies also drive globalisation. Policies on labour movements, capital flows, and trade in goods and services influence the pace and nature of globalisation. Box 34.1 describes how government policies have played an important role in the rise of globalisation.

Dimensions of globalisation

Globalisation has five main economic features. The first is international trade. Since the Second World War, international trade in goods and services has grown twice as fast as world output—at about 6 per cent per annum compared with a 3 per cent growth in world output. This implies that imports and exports as a proportion of output have increased substantially in most countries.

Second, financial markets are increasingly integrated. In the 1980s and 1990s many governments relaxed controls on capital flows. Cross-border flows of capital increased over twice as fast as cross-border commodity flows and over four times as fast as world output.

Box 34.1 Globalisation in historical perspective

There was trade between the Middle East and China over 2000 years ago. International trade expanded steadily over the centuries with the development of sailing ships and the discovery of new lands, the industrial revolution in the 18th century, and the development of the iron steamship and the telegraph in the 19th century.

Traders required finance to bridge the gap between the purchase of goods and their sale, often many months later. Trading banks generally provided this short-term finance. On the other hand, long-term capital flows were used to clear and plant land, dig mines, build factories, construct railways and develop other infrastructure. Merchant banks and finance houses provided most of this long-term finance in the form of bonds, debentures and shares. By 1914, international trade as a proportion of world output was almost as high as it is today.

The period from 1920 to the mid-1950s was one of de-globalisation. International trade and flows of finance fell especially during the Great Depression of the 1930s and the associated protectionist policies. After the Second World War, most European countries restricted outflows of capital to protect their fixed exchange rates and diminished foreign exchange reserves.

In the 1960s, many newly independent countries turned to inward-looking, import-substituting economic strategies. Two of the world's largest countries, China and the Soviet Union, adopted command and control socialist systems. These systems were under considerable strain when the information and communication technology revolution arrived in the 1980s.

Government policies have also had a major influence on globalisation in recent years. In the early 1980s the world was divided into three main economic systems: the developed and emerging market economies; government-led developing economies, including India, that were inward looking, import substituting and foreign investment hostile; and the socialist economies of China, the Soviet Union and a few others.

Over the last 30 years many countries (especially China, India and Eastern European countries) have made dramatic policy changes, notably by liberalising exchange rates and trade and capital flows. In the early 1980s, fewer than two billion people lived in broadly open market economies. To-day nearly all the world's population of seven billion people live in at least semi-open market economies, with extraordinary digitisation of information services. But some of these technological and policy changes are now strongly questioned.

These capital flows include foreign direct investment (FDI) and financial (portfolio) investment. FDI is direct investment in physical capital. It is often associated with transfer of technology. Portfolio investment may be equity investment (purchase of assets) or debt finance. Portfolio investment indirectly supports local capital formation and enhances local capital supply. Both forms of investment increase foreign ownership of domestic assets. Portfolio investment is generally more volatile.

Third, production systems are increasingly internationalised as the technologies of transport, logistics and communications advance and firms operate in more than one country. Multinational firms may run vertically integrated operations or multi-domestic strategies. In a vertically integrated operation, some parts of the firm produce inputs for other parts based on their competitive advantage. Intra-firm trade accounts for over a quarter of world trade. Firms can also organise integrated operations by outsourcing and joining into contractual relationships with other firms. With a multi-domestic strategy, the parent firm wholly or partly owns and controls affiliates in several markets, but each affiliate serves mainly its local or regional market.

The fourth major economic feature of globalisation is permanent and temporary labour migration. In most OECD countries the proportion of foreign born population increased by between 1 and 4 per cent between 1995 and 2005 (Hanson, 2008). However in about half the OECD countries the foreign-born population was still under 10 per cent. Worldwide, only 3 per cent of all people lived outside their country of birth. By comparison, in Australia, foreign-born persons rose from 27.4 per cent in 2001 to 30.2 per cent in 2016.

The fifth feature is the use of international standards and law as the basis for conducting business. Sovereign states have signed numerous treaties that provide shared standards and rules. Institutions such as the International Monetary Fund (IMF), the Bank for International Settlements, the OECD, the World Trade Organization (WTO) and other UN organisations develop compacts that become agreements between most countries of the globe. The precursor to the WTO was the General Agreement on Tariffs and Trade (GATT). After the Second World War, in eight GATT rounds tariffs fell from an average of almost 50 per cent to less than 5 per cent.[1] As signatories to the WTO, some 155 countries[2] have now joined into a detailed code of conduct governing virtually every aspect of international commerce, including not only tariffs and quotas but also health standards in international trade and many complex issues of international services.

Measures of globalisation

As shown in Table 34.1, measures of the amount of integration in the major product and factor markets are available, notably data on trade flows, financial flows, direct portfolio investment and payments to non-nationals as a percentage of GDP. These reflect the first four features of market integration noted above. It is harder to obtain simple measures of transaction restraints such as non-tariff barriers to trade, capital controls or risk premiums that reflect political instability.

Drawing on these measures of integration, Table 34.1 provides a qualitative assessment of the extent of globalisation in the markets for goods, financial and physical capital, and labour. H, M and L indicate high, medium and low levels of globalisation respectively. Financial markets are the most integrated. However, interest rate parity is not fully achieved when there is a high degree of political risk. There also remain significant non-tariff barriers in goods markets, capital controls on FDI in many countries and widespread controls on cross-border labour movements. In general, the markets of developed economies are more highly integrated with each other than they are with the markets of less developed countries.

[1] Schulze and Ursprung (1999).
[2] The number increases regularly.

Table 34.1 Measures of globalisation: high, medium and low globalisation

Market	*Flow data*	*G*	*Stock data*	*G*	*Price and wage equality*	*G*	*Actual barriers*	*G*
Goods and services	Trade % of GDP	M	Na		Law of one price	L	Tariffs Non-tariff barriers	M L
Financial capital	Portfolio investment as % of GDP	H	Foreign ownership as % of total stock of assets	M	Interest rate parity (IRP)	M	Capital controls	M
Physical capital	FDI investment as % of GDP	M	FDI as % of capital stock	M	Real IRP	M	Capital controls	M
Labour	Income payments to non-nationals as % of GDP	M	Non-nationals as % of workforce	M	Reduced international wage differentials	L	Migration controls	L

Source: Based on Schulze and Ursprung (1999). (*G* = globalisation H, M, L = high, medium and low.)

Ideally, we would like to have a single measure of a country's globalisation. The simplest single measure is probably international trade as a proportion of GDP. However, that does not account for the other features of globalisation. The KOF Index of Economic Globalization is an example of an attempt to estimate a comprehensive measure of economic globalisation (http://globalisation.kof.ethz.ch). This gives 50 per cent weight to flows of trade, FDI, portfolio investment and income payments to non-nationals as a percentage of GDP (with a weight of about a quarter to each) and a weight of 50 per cent to various restrictions such as hidden import barriers, mean traffic rate, and capital account restrictions. This approach seems to have two limitations. There does not seem to be a clear rationale for the weights allocated to the different features of globalisation and there appears to be double counting in estimating both volumes of trade and capital flows and restrictions on these flows. Nevertheless, the results are plausible. In the latest (2014 data) ranking, Singapore, Luxembourg and Ireland rank 1, 2 and 3 in the Economic Globalization Index, Australia is 42, while the ex-Soviet republics of Turkmenistan and Uzbekistan rank 204 and 205 and the Virgin Islands and West Bank/Gaza bring up the rear at numbers 206 and 207.

Globalisation, Trade, Output and Prices

A fundamental theorem of economics is the theorem of comparative advantage: all parties to trade can gain. We start by outlining this theorem.[3] We then briefly describe some other effects of globalisation on factor markets.

Comparative advantage
Exists when the opportunity cost of production of some good is lower in one country than in another

Comparative advantage and the benefits of trade

The theorem of comparative advantage states that two parties (firms or countries) can gain from trade whenever one party has a comparative advantage in the production of some good(s). A country has a comparative advantage in producing a good if the opportunity cost in terms of other goods forgone is lower than in another country. A country nearly always has a comparative advantage in producing some good even though it may be less efficient in absolute terms at producing both (or all) goods that both parties produce.

Theorem of comparative advantage
Any two parties (e.g. countries) can gain from trade whenever one party has a comparative advantage in the production of some good

Suppose that workers in Australia and China produce two goods: wheat and textiles. As shown in Table 34.2 overleaf, an Australian worker can produce 10 kg of wheat or 10 metres

[3] The effects of international trade are discussed in most microeconomic texts, for example Taylor, Moosa and Cowling (2000). The discussion here is brief.

Table 34.2 Comparative advantage and gains from trade

	Output per labour day		*Opportunity cost of production*	
Commodity	*Australia*	*China*	*Australia*	*China*
Wheat (kgs)	10	2	1.0 cloth metres	2.5 cloth metres
Cloth (metres)	10	5	1.0 wheat kg	0.4 wheat kg

of cloth in a day. The (opportunity) cost of a kg of wheat is 1 metre of cloth. Conversely, the cost of a metre of cloth is 1 kg of wheat. A Chinese worker can produce 2 kg of wheat or 5 metres of cloth in a day. The (opportunity) cost of a kg of wheat is 2.5 metres of cloth, and the cost of a metre of cloth is 0.4 kg of wheat. These figures imply that Australia has a comparative advantage in producing wheat because the amount of cloth given up per kg of wheat is lower than in China. China has a comparative advantage in producing cloth because the opportunity cost of a metre of cloth is only 0.4 kg of wheat instead of 1.0 kg.

Although, in this hypothetical case, Australian workers have an absolute advantage in producing wheat and cloth, both countries can gain from trade. Suppose that there are two workers in each country and that one produces wheat and the other cloth. Table 34.2 shows total daily output. Now suppose that the workers adopt the Australian exchange rate of one metre of cloth for one kg of wheat. The two Chinese workers would produce 10 metres of cloth and trade cloth for wheat from Australia. If they sell 4 metres of cloth, they would consume 6 metres of cloth and 4 kg of wheat compared with 5 metres of cloth and 2 kg of wheat without trading. In this case, the Australians would produce 14 kg of wheat and 6 metres of cloth, but they would consume the same as without trade because the exchange rate does not change. Now retain these new levels of production but suppose that the exchange rate is 2 metres of cloth for 1 kg of wheat. With trading, the Chinese would consume 6 metres of cloth and 2 kg of wheat, while the Australians would consume 18 metres of cloth and 10 kg of wheat. Both parties consume more cloth than before, but no less wheat. *Trade increases output when countries produce according to their comparative advantage*.

However, two caveats are needed. First, it assumes that workers can switch readily between producing wheat and cloth. Those who cannot will become unemployed (or underemployed). Second, trade is beneficial only if the differences in productivity offset freight costs. Suppose that a Chinese worker spends 0.6 of a day in travel time to execute the trade. In this case the Chinese workers could produce 7.0 metres of cloth and no wheat. With the most favourable exchange rate for them (1:1), they could trade and finish with 5 metres of cloth and 2 kg of wheat, and the Australians would be no worse off. Trade is beneficial only when travel time falls to less than 0.6 of a day.

The effects of a fall in transport costs. These are shown in Figure 34.1. Suppose firms in two countries, X and Y, produce computers in a competitive market. Firms in X produce computers at an ex-factory cost of A and the costs of supplying computers inclusive of transport costs to different locations in X and Y are shown by line ACD. The equivalent cost schedule for firms in Y is shown by line ECB. Firms in X supply at a competitive price up to Q_1. Thus, firms in X have a share of the market in Y. Now suppose that transport costs fall and that firms in X can supply customers at prices along the AF line. Although the ECB schedule would also rotate downward (this is not drawn), firms in X now supply all the computers in Y.

Given a competitive market, the gain from the fall in transport costs is equal to area $ACEF$. Further, all gains will be passed on to consumers, including those in Y. Firms in Y simply switch to producing other goods. If markets are competitive, there are gains from trade in both countries. However, if the market is not competitive, firms in X may charge above marginal cost and gain some rent from increased market share while firms in Y may lose some producer surplus. Both consumers and firms in Y may lose as transport costs fall.

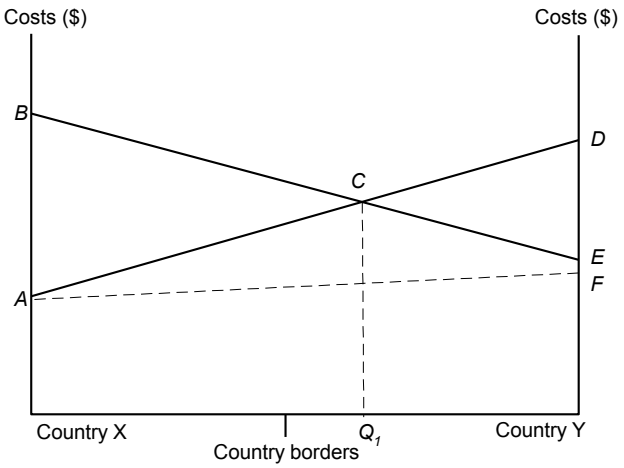


Figure 34.1 Benefits of lower transport costs and increased trade

Lower transport costs generally increase the number of firms that can supply any given market. This reduces prices and produces net benefits in that market. In a competitive market, consumers gain and producers adapt.

Larger markets also enable firms to achieve economies of scale and to reduce the cost per unit of output. Even with no difference in the efficiency of production between countries, countries can gain by specialising, adopting economies of scale and trading rather than by self-sufficiency policies (see Taylor *et al*., 2000). However, countries may not gain if they cannot adapt their production, especially if labour or capital is under-employed, or if they become subject to the monopoly power of foreign suppliers.

Globalisation and factor markets

We now turn to globalisation impacts on factor markets, starting with capital markets. International financial markets move capital from savers to borrowers and foreign exchange from nations with balance of payments surpluses to nations with current account deficits. Competitive markets reduce the gap between the interest rate paid to lenders and the rate charged to borrowers and the spread between the buying and selling price of foreign exchange. Both spreads create a deadweight loss (DWL) due to missed transactions. Globalisation allows financial institutions to compete in the markets of any country. This provides productivity gains, enables capital to be employed in the most productive opportunities and reduces DWLs from missed transactions.

Given the large earnings differentials for workers of similar skills in different countries, economic theory suggests that migration of workers from low to high productivity countries increases global output and income. Hanson (2008) reported that these global gains from migration are largely captured by the migrants, which they share with their families through remittances. This may be at the expense of some local incomes, but it appears not enough to produce negative country or world economic growth.

Another aspect of globalisation is the international integration of production systems. This facilitates the international flow of technology and capital, raises local productivity and allows goods to be produced in countries that could not otherwise produce these goods. It is generally efficient as multinational firms seek to produce their goods at least cost. Also, local

workers in the multinational firms generally choose to work in these firms because they perceive the real wage is higher than in other local alternatives.

However, the benefits of globalisation depend on the ability of markets to adjust. Following the integration of Eastern and Western European economies, between 1995 and 2005 employment in Germany fell by 1.4 million, mostly in manufacturing. Sinn (2007) observed that: 'The workers went into the welfare state, into state-financed unemployment. The dirty industrial jobs were abolished but nothing was created in exchange.' This resulted from rigid wages combined with generous wage replacement incomes. When wages are high, the return to capital falls and capital leaves the country while unemployment increases. But given Germany's economic performance in recent years, this verdict may have been premature.

On the other hand, it seems that labour markets in the industrial heartland of the United States, especially in the steel and aluminium industries, have not been able to adjust to international competition.

The evidence

Evidence supports the thesis that globalisation (integrated world market) increases output and reduces prices. Over the 20 or so years between the mid-1980's and the mid-2000's there occurred an extraordinary increase in the integration of world markets driven by technology, especially in communications, and by political changes in the Soviet Union, Eastern Europe, China and India. The rate of world economic growth over this period was significantly higher than in any previous decade of the 20th century. Consumers of numerous products, including food, clothing, steel, motor vehicles, communications and education services, benefited from lower transport and distance costs and increased international competition.

In Chapter 5, we found that open economies tend to achieve higher rates of growth than closed economies. Edwards (1998) provided further evidence with a detailed econometric study of the relationship between openness and growth for 93 countries. He represented economic growth by estimated total factor productivity growth and tested for nine indicators of openness. Edwards concluded that more open countries have achieved faster productivity growth and that this result is robust with respect to openness indicators, estimation technique, time period and functional form. The World Bank (2002) reported that the more globalised developing countries increased their per capita growth rates from one per cent per annum in the 1960s, to 3 per cent in the 1970s, 4 per cent in the 1980s and 5 per cent in the 1990s, as global markets became more integrated.

However, globalisation also contributed to the nearly worldwide depression between 2008 and 2011. Financial institutions and economies around the world were badly hit, first between 2007 and 2009 by the crises in the US housing market and banks, and then between 2009 and 2011 by the government debt crisis in some European countries that affected banks that had lent to them. Countries with more financial autonomy suffered less than those linked closer to global financial markets, but even they could not escape the effects of the worldwide slowdown in economic activity.

Distributional Consequences of Globalisation

While theory and evidence suggest that countries as a whole generally gain from globalisation, some groups within countries may still lose. So, who gains and who loses from globalisation?

We consider here how globalisation affects skilled and unskilled labour and owners of capital and land. Of course, some individuals may incur losses as workers but gain as consumers. In this analysis, skilled labour is assumed to be relatively abundant in developed economies and unskilled labour to be abundant in less developed economies.

The Heckscher–Ohlin international trade model (the basis for most modelling of international trade) indicates that, if capital and technology are completely mobile and there are no transport costs, trade equalises factor and product prices across countries. This is known as the **theorem of factor price equalisation**. In a closed economy, goods produced by factors in plentiful supply sell at a low price. Goods produced by scarce factors command a high price. A country exports low-cost goods and imports goods produced locally at relatively high cost. Trade increases earnings of local factors in plentiful supply and reduces earnings of factors in scarce supply. Developed economies export goods produced by skilled labour and import goods produced by unskilled labour. Poor countries do the reverse. Thus, the wages of skilled workers in developed economies and of unskilled workers in less developed countries rise. Conversely, the wages of unskilled workers in developed economies and of skilled workers in less developed economies fall. This process continues until the ratio of skilled to unskilled wages is the same in developed and less developed countries. This suggests that trade increases wage dispersion within rich countries but reduces it in poor countries.

Theorem of factor price equalisation
With unrestricted trade and no transport costs, trade equalises factor and product prices across trading partners

However, the theorem of factor price convergence holds fully only if the relevant countries can produce all goods, imported goods are close substitutes for locally produced goods, there are few non-traded goods and labour is mobile between sectors. Returns to factors do not converge when there are significant sectors of non-traded goods or where labour does not move into internationally traded goods sectors.

Also, globalisation does not eliminate all transport costs, other barriers to factor and product movements or political risks. Although capital is mobile, workers in developed economies work with substantially more physical capital than do workers in poor countries. They are therefore more productive and continue to earn higher wages than comparably skilled workers in poor countries. Due to transport costs, price differentials exist for traded goods and many goods are not traded. This also allows wage differentials between countries for comparably skilled workers.

Turning to capital, with free capital movement, the marginal return to capital rises as capital moves to the most productive opportunities. The rise in the rate of return depends on the elasticity of world savings. The more that savings respond to higher rates of return, the lower the increase in the marginal return to capital.

On the other hand, land is not mobile. Trade and capital mobility increase the return to productive land, which receives more capital and obtains a larger trade area for its products and reduce the return on less productive land.

Consumers generally gain from globalisation. Globalisation reduces prices by lowering transport costs, increasing competition and lower-unit costs of production. Consumers of imported goods gain. Australian consumers have benefited greatly from substantial falls in the relative prices of imported manufactured goods over the last 30 years. However, some consumers may lose when exports increase and local prices rise. When developing economies increase food exports to developed economies, consumers in the former pay higher local food prices. In Australia, fish prices increase with the rise in fish exports to Asia.

Some critical issues. In developed economies, less skilled workers are major potential losers from globalisation. In Australia, they compete increasingly with low-wage workers in Asia. Where Australian workers work with more capital, their productivity is higher and they receive higher wages. However, if the ratio of Australian earnings to Asian earnings exceeds the ratio of Australian to Asian productivity, less skilled Australian workers are likely to experience lower wages or less employment, or both.

There are some compensating factors. The rise in income of high-skilled workers increases the demand for services, many of which are labour intensive and non-traded. Second, low-skilled workers gain (as do others) from lower import prices, especially for manufactured goods. Third, higher economic growth increases tax revenues and public services.

Within developing countries, some workers may not gain from globalisation because they do not participate in traded markets. However, the criticism that multinational firms exploit unskilled labour, including children, in poor countries seems less secure. Certainly, unskilled workers in poor countries receive much lower wages than their counterparts in rich countries. This does not mean that globalisation has worsened their conditions. Child labour reflects poverty, not trade. Edmonds (2001) showed that opening up the Vietnamese economy led to a large increase in household income and a sharp drop in child labour. The wages paid by foreign affiliates of multinational firms to poor-country workers are typically double the local manufacturing wage.[4] As Stiglitz, generally a strong critic of globalisation, observed (2002, p. 5): 'People in the West may regard low paying jobs at Nike as exploitation, but for many people in the developing world, working in a factory is a far better option than staying down on the farm and growing rice'.[5]

However, critics of globalisation, including Stiglitz (*ibid*.), argue that the process of globalisation is profoundly influenced by the power relations between states. The rich countries make the rules of international market engagement and reap the benefits. The most egregious example is the protection of farm producers in the European Union and the United States. The US protects other industries such as steel producers. It also upholds restrictive covenants on the intellectual property rights of its corporations regardless of the impact on poor countries. On the other hand, it breaks the rules, for example when Bayer (a German multinational company) was required to reduce the price of ciprofloxacin to combat anthrax threats in the US. The US government has been less accommodating with patents for AIDS drugs, which poor countries want relaxed so that they can obtain the drugs at much lower prices.

Stiglitz (*ibid*.) and other critics of globalisation also cite the dominance of international institutions such as the International Monetary Fund (IMF) and the World Bank by rich countries. In their view, both institutions have imposed harsh and politically unrealistic governance conditions on poor countries in return for loans provided and these conditions have created economic recession and political instability rather than economic growth.

Globalisation and income distribution: some evidence. In this section we review the effects of globalisation on worldwide poverty and inequality and on inequalities within countries.

Global poverty has fallen substantially over the last 30 years since markets became much more integrated. Drawing on World Bank data, Harrison (2006) reported that the percentage of people living below the extreme poverty line of $1 per day (in 1985 prices) had halved since the early 1990s, although the number still exceeded one billion. The fall was largely due to declines in poverty in South and East Asia, including China, the Middle East and North Africa. There was little decline in poverty in Central Asia, Latin America or sub-Saharan Africa, where nearly half the population continued to live below the severe poverty line. Dollar and Kraay (2001) also showed that the poor had shared in gains in national income growth, with the poorest 20 per cent in most countries obtaining the same percentage gain as the country did. There have also been significant gains in life expectancy and education in low-income countries over the last 30 years. The most serious levels of poverty exist in countries that are least integrated with the global economy.

As we saw in Chapter 5, incomes between countries also appear to have converged, especially for countries that are relatively open, which should reduce global inequalities. Lindert and Williamson (2001) found that globalisation had reduced income gaps between nations. However, in a review of literature on global income inequality, Anand and Segal

[4] *The Economist*, 29 September 2001.

[5] Another argument is that multinationals exploit their workers by paying them less than the value of their marginal product. This may occur with imperfect competition in the labour market.

(2008) found that the worldwide Gini coefficient (GC) was between 0.630 and 0.686 in the 1990s and that there is no consensus whether this has risen or fallen over the last 10 years. These various findings with respect to global inequality are sensitive to both the time period and the way that countries are included. Because of the size of China and its income growth, inequality measures weighted by population usually indicate that global inequality has fallen. Unweighted inequality measures often show the reverse.

However, as noted in Chapter 20, many studies have found that globalisation is associated with rising inequality *within* countries since 1980 (Harrison, 2006; OECD, 2008 and 2009; Aristei *et al*., 2009). As reported by OECD (2011), the GC for disposable household income in OECD countries rose from 0.29 in the late 1980s to nearly 0.32 in 2010. It rose for 17 of the 22 OECD countries for which long-term data were available. In nearly all OECD countries, the wages of the 10 per cent highest-paid workers have risen faster than the wages of the 10 per cent lowest-paid workers. In developed economies, the demand for skilled workers outstripped the growth in supply and unskilled workers found it hard to move out of declining sectors into expanding ones.

The rise in the skilled–unskilled wage ratio in developed countries is consistent with the Heckscher–Ohlin international trade model and the economic theory of globalisation. Many analysts (Wood, 1998; Cornia, 2004; Dreher and Gaston, 2008) conclude that globalisation was a major cause of rising inequality within countries, especially in developed economics. On the other hand, Zhou *et al*. (2011) found no correlation between the GCs in 60 countries and two constructed measures of globalisation. However, the authors' choice and weighting of variables in their globalisation measures were arbitrary. Also, a simple bivariate analysis may not expose the real drivers of income inequality. For example, a high GC and low globalisation could both be related to low income levels.

Within Australia, we have also observed increasing income inequality and linked this with globalisation (Chapter 20). There have also been significant structural employment changes since 1990 due to global factors. Employment in the education and training sector rose from 591 000 in August 1990 to 859 000 in August 2010 due mainly to a large increase in international students. Over the same period, employment in mining rose from 98 000 to 193 000 (nearly all since 2003). On the other hand, employment in manufacturing fell from 1126 000 to 993 000 as tariff and non-tariff barriers were removed and China became the major supplier of manufactured goods. The retail sector is also now affected significantly by the growth of Internet trading. However other major non-trade sectors, such as utilities, health services, construction, transport and public administration have been little affected to-date.

Conclusions. Market integration closes the gaps between product and factor prices in developed and developing economies. It generally raises the wages of low-skilled workers in poor countries but reduces wages of these workers in rich countries. Also, workers in previously protected industries lose from policy liberalisation. In practice, markets are far from fully integrated. Differences in levels of capital intensity and knowledge, as well as restrictions on factor mobility, result in substantial differences in incomes between similarly skilled workers in different countries.

Globalisation has generally reduced international poverty although some countries have not experienced these benefits. Thus, the overall trend in international inequality is not clear. However, globalisation has most likely increased inequality within developed economies.

Globalisation and Government Revenue

Globalisation may be expected to reduce government revenues for several reasons. The most important reason is the increased mobility of factors of production. The more mobile are factors, the more easily they change jurisdictions to escape tax. This reduces the tax base. To

attract or retain mobile marginal capital and labour, governments must provide competitive tax rates. Usually this means lower tax rates. Thus, to retain revenue, governments must either broaden the tax base or increase tax rates on immobile factors.

Second, modern technology facilitates the movement of money and enables tax havens and tax evasion. People can be paid more easily in non-domestic accounts and evade income tax, purchase online and evade sales taxes, or transfer wealth around the world and evade tax on income from capital. However, technology combined with increasing cooperation between taxation agencies also increases government capacity to monitor financial transactions.

Third, tariff reductions reduce tax revenue on trade. In some developing countries, trade taxes account for a high proportion of government revenue, partly because other tax bases (such as wages and profits) are small or easily evaded. However, trade taxes typically account for only 1 to 2 per cent of government revenue in developed economies.

Taxes on factors of production

Accordingly, we focus here on the major issue, namely taxes on factors or production. A principal hypothesis of the globalisation literature (for example, Gemmell *et al*., 2008) is that tax rates will fall on mobile factors of production (capital) and rise on less mobile factors (labour and land).

Taxation of capital. When a country taxes international capital, the gross rate of return required by lenders increases. Owners of capital, especially portfolio capital, can shift their capital to another country until the local after-tax return meets the international rate of return on capital (or more if there is a local risk premium). Therefore, a tax on foreign capital is borne by local firms that pay more for capital and by immobile local factors of production, which have less capital to work with.

In theory, this tax should affect only supply of non-resident capital because governments can tax their residents on their worldwide income from capital. Thus, residents have no tax incentive to invest overseas. However, cross-border investments may evade tax. Despite double-taxation agreements, the introduction of withholding taxes on interest income in Germany, the US and other countries led to capital flight from these countries because income from some assets was evading national taxation. Schulze and Ursprung (1999, p. 313) reported that:

> Interest income had already been liable to personal income tax prior to the introduction of the withholding tax; and the withholding tax payments are fully credited against the personal income tax. Thus, the withholding tax was targeted only at domestic tax evasion, which triggered international tax evasion.

Tax evasion is possible when investment income is taxed in the country of the recipient but not taxed, or only lightly taxed, at source and when there are limited communications between national tax authorities.

Income generated by FDI is taxed through company income tax. Some research shows that FDI is sensitive to corporate tax rates. Cassou (1997) found a significant negative relationship between the US corporate tax rate and FDI in the United States and a significant positive relationship between foreign corporate tax rates and FDI from the United States. Devereux and Griffith (1998) found that effective average tax rates influence the location of US firms in Europe. A one per cent increase in the effective average tax rate in the UK reduced the probability of a US firm producing there by an estimated 1.3 percentage points.

To attract corporations, and gain the benefits of increased investment, governments may reduce their corporate tax rates. However, taxation is only one factor in a firm's investment decision. Governments may also provide favourable regulations or service inducements.

Table 34.3 Corporate income tax rates in the OECD

Results	*1986*	*1991*	*1998*	*2011*
Average unweighted tax rate	43.5	36.3	33.2	25.5
Range (maximum to minimum)	28.0	27.0	20.0	20.2
Standard deviation	7.0	6.8	4.5	6.0

Source: OECD (2011) Tax Database.

The overall effect is that globalisation increases tax competition and lowers corporate tax rates. As shown in Table 34.3, the average (unweighted) rate of corporate income tax in OECD countries fell from 47 per cent in 1982 to 25.5 per cent in 2011. Smaller countries lowered their rates most whereas the larger economies, such as the US, the UK and Japan, lowered their rates by lesser amounts.[6] In Australia, the corporate tax rate fell from 49 per cent in the mid-1980s to 30 per cent in 2001, which was still the rate through 2018. Also, corporate tax rates have converged. In the 1990s, the gap between the highest and lowest corporate tax rates in the OECD decreased significantly and the standard deviation, which measures the degree of dispersion of tax rates from the mean, also fell.

Although corporate tax rates have fallen, corporate income tax has not declined as a percentage of GDP. From 1990 to 2010, the weighted average company tax revenues in the OECD remained at about 3 per cent of GDP (Henry Tax Review, 2010). The unweighted average (which reduces the influence of the US) rose to about 4 per cent of GDP. This reflected both the rise in corporate profits as a percentage of GDP and a broadening of the corporate tax base. The fall in statutory tax rates overstates the decline in effective tax rates when depreciation rules, investment allowances and other corporate supports are tightened.

Taxation of labour. There remain substantial barriers to cross-border movements of labour. However, skilled high-income workers are usually in international demand and can afford to relocate. Some labour-short countries in the Middle East need labour at all skill levels. Governments must offer competitive tax rates to attract and retain labour, especially skilled labour. Over the last 25 years, there has been a worldwide decline in high marginal tax rates and the progressivity of income tax structures.

Rodrik (1997) examined whether economic integration affects **relative** tax rates for capital and labour. He measured integration by the ratio of the sum of exports and imports to GDP.[7] Tax rates for capital and labour were obtained by dividing respective tax revenues by the gross pre-tax income of each factor. Thus, he tested for the average tax rate on capital and labour. Analysing 18 OECD countries from 1965 to 1991, he found that average tax rates on labour increased with openness and that average tax rates on capital fell with openness. He concluded that, as economic integration increased, the tax burden shifted from capital to labour.

Summary. Tax rates on mobile factors (capital and skilled high-income workers) have generally fallen absolutely and relatively and there is increasing pressure to reduce corporation tax. On the other hand, tax rates on less mobile factors (unskilled workers, property or natural resources) have risen. Less mobile workers tend to be lower-income workers. Property owners are generally higher-income persons.

---

[6] As we saw in Chapter 31, the US has now legislated to reduce the corporate tax rate radically to 20%.

[7] Note that this is a measure of trade mobility, not capital mobility (the main factor in tax competition).

Taxes on consumption

Consumption is generally less mobile than factors of production. Most consumption takes place in the home country, and most imported goods can be taxed. Therefore. it is harder to avoid consumption taxes. However, avoidance has increased with the rapidly rising purchases over the Internet, most of which have been free of consumption taxes.

On balance, because of the relative immobility of most consumption activities notwithstanding untaxed purchases over the Internet, globalisation may be expected to lead to higher consumption taxes. Also, user charges may be expected to increase, for example for road use, education, health services and so on. Both phenomena have occurred. As we saw in Chapter 31, over the last 30 years, OECD countries including Australia have significantly increased consumption taxes, notably by use of the value-added tax (or GST).

Shifts in the tax structure towards taxation of consumption and away from taxation of capital affect the burden of taxation. It increases the tax burden of those outside the workforce including pensioners, students and the unemployed, and reduces the tax burden on savers.

Tax competition and economic efficiency

We have examined the impacts of tax competition on the incidence of tax, but how does it affect economic efficiency?

When a factor of production is not mobile, lowering the tax rate has no effect on the tax base. The efficient tax rate will depend only on local effects, that is, the marginal social benefit of taxation and the marginal social cost, including any DWL associated with impacts on labour supply or saving.

When factors of production are mobile, the tax rate set in one country can affect the tax base in another. A reduction in a local tax rate because of tax competition reduces the tax base of other jurisdictions and is therefore a negative externality. However, the taxing jurisdiction is generally concerned only with its own revenue and ignores this externality. Tax competition is likely to reduce tax rates for mobile factors of production. Ignoring the negative externality of tax competition may result in under-taxation and hence under-provision of public goods in both the taxing jurisdiction and in the other jurisdictions. However, the impact may be only minor.

This result applies especially when income is taxed at source, as is commonly the case. Differential taxation of income at source affects the allocation of mobile capital and the size of the tax base. If income is taxed according to the destination principle, a resident's worldwide capital income is taxed at the same rate, regardless of its source. Tax competition is then less likely to affect the tax base because residents are less mobile than capital.

Most countries, including Australia, tax income under both the residence and source principle. In Australia, residents are taxed on their worldwide income and imports bear the GST, whereas exports are GST free. Law-abiding residents have no tax incentive to move capital overseas. The Australian government also taxes income from Australian sources that accrues to foreigners. Thus, the Australian system has a tax-competitive source-based element, which may discourage foreign capital and marginally distort the worldwide allocation of capital and productive efficiency.

Concluding observations

Globalisation is likely to put downward pressure on tax rates, especially on capital and on high personal income tax rates. This might be expected to reduce tax revenues. However, tax revenues have not fallen. As a proportion of GDP, tax revenues in OECD countries have risen slightly from an unweighted average of 33 per cent in the mid-1970s to around 35 per cent between 2002 and 2008. Moreover, government revenues are generally higher in countries with open economies.

Evidently, conclusions about globalisation and tax revenues must be drawn cautiously. Tax revenue is the product of the tax base and tax rate(s). A broadening of the tax base, or a rise in its value, may offset a decline in the tax rate. Open efficient economies that encourage trade may have broad tax bases. Second, high tax rates on immobile factors may offset low tax rates on mobile factors. Third, and most importantly, tax revenues depend also on the demand for government services. As discussed below, global forces increase the demand for some government expenditures. Government revenue may rise in response to this increase in demand despite the impact of tax competition on tax rates.

Globalisation and Government Expenditure

In Chapter 2, we identified four main economic determinants of government expenditure: household income, demographics, welfare needs and production costs. The main political determinants are political ideology, the discretionary decisions of politicians and public officials, and interest groups.

Table 34.4 shows the possible impact of globalisation on these eight drivers of expenditure. Two factors stand out as possible drivers of increased government expenditure. One is the effect of globalisation on welfare needs. The 'compensation hypothesis' (Gemmell *et al*., 2008) suggests that economic inequality and insecurity resulting from globalisation increase public expenditure. Related to this is the rise in interest group pressure as groups seek to protect their interests from foreign competition. On the other hand, the 'efficiency effects' of globalisation may reduce the discretionary power of politicians or public servants to raise expenditure or the tax to finance it. Also, increased competition may reduce production costs. Therefore, globalisation could increase or decrease total government expenditure, depending on which expenditure drivers are most influential as well as on the revenue constraint. This will depend also on the ideology of the governing party.

Total government expenditure. Several econometric studies have assessed the impact of globalisation on government expenditure. As usual with such studies, issues arise with definition of the variables as well as with the relationships between them. In most studies the dependent variable is total government expenditure as a share of GDP. Arguably, total non-defence expenditure would be a more relevant dependent variable. The variables representing global economic integration include the share of imports plus exports in GDP, the volume of capital flows, covered interest rate differentials and estimates of the degree of capital market liberalisation. Some studies, for example Garrett (1995) and Cusack (1997), include variables representing left–right political ideology.

Table 34.4 Globalisation and factors affecting government expenditure as percentage of GDP

Determining factors	*General influence on public expenditure*	*Possible impact of globalisation*
Economic factors		
Income	Income elasticity about one	No effect on expenditure as % of GDP
Demographics	Expenditure rises with ageing	No effect of globalisation
Welfare needs	Expenditure rises with poverty and inequality	May significantly increase expenditure
Cost of production	Expenditure rises with input prices	Could reduce expenditure slightly
Political factors		
Political ideology	Expenditure a function of ideology	Impact important but ambiguous
Political discretion	Expenditure a function of discretion	Could reduce expenditure
Public officers' discretion	Expenditure a function of discretion	Minimal impact
Interest groups	Interest groups raise expenditure	Could increase expenditure

Schulze and Ursprung (1999) surveyed seven studies of government expenditure, including three that incorporate political and economic factors as independent variables (politico-economic models) and four that focus on economic factors (apolitical models). They reported that, in the former models, it is often difficult to separate the impacts of market integration and politics. Garrett (1998) found that a combination of globalisation and left-wing power was associated with increased public expenditure, especially at high levels of globalisation, but that the separate influences of globalisation and politics are not clear. Swank (1997) found that capital market integration was associated positively with public expenditure in countries characterised by high corporatism (e.g. Norway, Sweden), high central authority (e.g. Denmark, Finland) and high consensus democracy (e.g. Belgium, the Netherlands).

Early apolitical models focusing on economic factors found that public spending was usually associated positively with market integration. Quinn (1997) found that capital market integration usually increased government share of expenditure. Rodrik (1998) and Garrett (1998) found that government spending rose as a share of GDP with openness to trade. However, Gemmell *et al*. (2008) found that no clear relationship between government expenditure and globalisation. Overall, the compensatory demand for more social insurance programs and expenditure due to globalisation appears to be broadly offset by efficiency factors that tend to reduce government expenditure. Governments can accommodate the compensatory demands despite the constraints on taxation due to more mobile factors. However, the net impact is broadly neutral.

Composition of government expenditure. It follows that globalisation is likely to change the composition of government expenditure with an increase in social security expenditures, especially personal transfer payments. In most countries, transfer payments have risen as a share of both GDP and government spending since the 1970s. However, this may have reflected increased welfare needs or changes in the political views of governments independently of global economic forces. Econometric studies are needed to separate out the influence of market integration and other factors.

Based on their review of such studies, Schulze and Ursprung (1999) concluded that market integration has led to higher social security expenditures. Quinn (1997) found that financial integration increased welfare and security payments. Rodrik (1998) found that trade integration increased most forms of government spending, including spending on health, education and social security. In a more recent study, Gemmell *et al*. (2008, p. 168) found:

> strong support for the compensation hypothesis on the mix of spending. Increases in globalisation measured by a country's stock of inward FDI are associated with a shift away from productive expenditures such as education and transport and communications and towards social welfare and public services spending.

Globalisation and Government Policies

Finally, how is globalisation likely to affect public policy? Some effects result from technology. Others arise from government agreements that are necessary for the international conduct of business. While technology both constrains and empowers government, international agreements are mainly constraints. Nevertheless, a key message is that most governments retain substantial autonomy over the policy spectrum.

Technology impacts. The volume and mobility of short-term capital, along with modern communications, make it very hard for governments to control cross-border capital movements as they did routinely between 1945 and the late 1970s. With free capital movements, governments cannot set both interest rates and the exchange rate. Regarding fiscal policy, if a government adopts budget deficits that financial markets judge inflationary

and likely to produce exchange rate depreciation, capital flight will raise interest rates and trigger a fall in the exchange rate. Therefore, globalisation of financial markets has simultaneously increased the need for financial regulation and deprived governments of some policy autonomy. This was evident in the Asian financial crisis in the late 1990s (see Box 34.2) and even more strongly evident in the global financial crisis that started in late 2007 and ran for several years. Of course, markets have always punished unsustainable policies such as over-valued exchange rates or persistent large budget deficits. What is new is the strength and speed of the punishment.

Modern factor mobility and communications have also reduced the capacity of governments to regulate the economy in other ways. Policy making implies a capacity to coerce from which the subject bodies cannot escape. Today, if the return on investment is capped by price regulations, capital investment goes overseas. For example, in the early 2000s price regulations encouraged the Australian Gas Light Company to redirect capital investment from Australia to New Zealand. If regulation constrains profits, it may also discourage entry into the market. Another area of greatly reduced government power is information regulation. It is extremely difficult to regulate the information or entertainment (including pornographic 'entertainment') that individuals may receive.

On the other hand, in some policy areas, such as competition policy, there is less need for regulation. Globalisation generally increases competition and reduces market failures associated with lack of competition. In the past there were many natural local monopolies in communications, construction and transport in Australia. These are now more contested markets. Written and oral communications that used to be the natural monopoly of the postal or telephone service are now possible by numerous electronic means. Telephony is a highly competitive industry requiring less regulation than was traditionally required.

Moreover, government has gained powers from technological advances. Computers have greatly increased government's capacity to hold or discover information on individuals. Governments can readily monitor traffic speeds. Genetic screening enhances government's capacity to provide improved health services and to detect criminals. Although technology is conceptually different from globalisation, it would be perverse to count only the reductions in government powers associated with technology-driven globalisation and to ignore how technology has increased the power of government in other ways.

International issues. International treaties are the basis for international action on matters of global concern. Australia is a signatory to over 900 international treaties. Global political concerns influence the use of labour and natural resources. Rules governing labour, for example child labour, are driven by international concern for human rights.[8] Rules governing the use of resources are responses to market failures connected with global public goods and externalities.

Box 34.2 Financial markets and the Asian meltdown in 1997

The rapid capital flight from Asian countries and the crash in exchange rates in 1997 showed that globalisation of financial markets (free capital flows) is not compatible with highly regulated domestic financial markets and fixed exchange rates. All the countries mainly affected by the meltdown, Indonesia, Malaysia, South Korea and Thailand, had regulated financial markets and fixed exchange rates. On the other hand, none of those countries had adequate prudential supervision of their banks and other financial institutions and none had a legal framework that allowed creditors to realise on the assets of bankrupt borrowers. The meltdown was a consequence of these poor regulations.

[8] The International Labour Organization, based in Geneva, is the main international agency responsible for labour.

International agreements on endangered species, whaling activity, and the use of resources in Antarctica are examples of semi-successful policy responses to market failures. It remains to be seen whether international agreement will be as effective on climate change.

Any government that wants good relations with other countries must accept some limitations on its ability to act autonomously. Certainly, international bodies have weak powers to enforce international agreements. Enforcement often depends more on the sanctions applied by other signatories of the treaty. However, the Australian High Court has ruled that Australian laws are invalid if they conflict with United Nations treaties entered into by the Commonwealth government. These treaties empowered the Australian government to stop construction of the proposed Franklin dam in Tasmania and the flooding of world heritage areas and to limit commercial activities around the Barrier Reef.

The importance of trade has also led many countries to cede power over trade restrictions to supra-national agencies such as the WTO. Global market integration requires global rules and adherence to these rules. Supra-national organisations are needed to deal with non-cooperative behaviour, such as tariffs which reduce the general benefits of trade. Access to the international benefits depends on agreeing to the international rules of trade. Inevitably, regulations over tariffs extend to equivalent non-tariff barriers to trade. Again, the WTO has little direct power and must rely on the will of independent nations for enforcement. In 1999 the WTO ruled that Australia's quarantine restrictions on salmon imports (a non-tariff barrier) contravened the rules of trade to which nearly all nations have agreed. Australia did not have to remove the restrictions, but if it failed to do so it would face sanctions from countries whose exports have been affected. International economic unions, such as the European Union, or free trade agreements, such as the North American Free Trade Association, are other ways in which countries agree to forgo national autonomy to gain the benefits of increased trade, specialisation and economies of scale. Australia and New Zealand have a free trade agreement which aims to integrate the two markets.

In summary. Integrated economic markets put pressure on national governments to bring their economic policies into line with those of other countries and restrict a government's macroeconomic and microeconomic policy options. The strength of the tendency towards policy convergence depends on a country's share of global trade and capital flows and its size. Small open countries that rely heavily on international trade and capital are under more pressure to adjust policies to meet market expectations. But in the global financial crisis even large countries, such as the United States, and major economic groupings of countries, such as the European Union, found their autonomy to set policies restricted by their capacity to attract international capital.

However, globalisation has not prevented governments from conducting policies consistent with their ideological stance. Governments can still determine the level and pattern of public spending. Governments often adopt compensating mechanisms to offset the distributional effects of globalisation. Also, technology has increased government's power to monitor and regulate its citizens in various ways. Democratic institutions and party ideologies remain important in shaping public policies. As I observed in previous editions of this book, in 2004, 2008 and 2012, "globalisation is not about to eliminate the nation state". As Brexit and America First have shown, this is even more true in 2018.

Summary

- Globalisation means the worldwide integration of markets. It is characterised by knowledge transmission, a global capital market, international production systems and international trade.
- It is driven by technological change that reduces distance costs. It may be promoted or hindered by public policy.
- Globalisation allows producers to specialise in their comparative advantage and achieve economies of scale, allows capital to be employed in the most productive opportunities, increases competition and reduces prices.
- Globalisation has contributed significantly to reduce the number of people in poverty around the world, especially in more open economies. Open economies have had significantly faster rates of economic growth than closed economies.
- The main beneficiaries from globalisation are exporters and consumers. The main losers are uncompetitive businesses. In terms of labour, the main winners are usually skilled workers in developed countries and low wage workers in open developing countries. The main losers are less-skilled workers in developed countries.
- On a worldwide population basis, globalisation has most likely reduced inequality. However, it has significantly increased income inequality in many developed countries.
- Globalisation constrains government revenue by creating tax competition, which puts downward pressure on tax rates on mobile factors, such as capital and high-skilled workers. However, taxes on consumption have risen.
- In part to offset the unequal impacts of globalisation, governments have increased compensatory public welfare spending.
- Reflecting the revenue constraints on the one hand and compensatory expenditure on the other hand, total government expenditure has stayed an approximately constant percentage of GDP.
- Integrated economic markets put pressure on governments to align their economic policies with those of other countries and restrict their macroeconomic and microeconomic policy options.
- Nevertheless, nationalism is very much alive. Most governments retain a high degree of policy autonomy.

Questions

1. In the words of Adam Smith, 'it is the maxim of every prudent master of a family, never to attempt to make at home what it will cost him more to make than to buy'. Explain how the theorem of comparative advantage is an application of this maxim.
2. Mercantilists argued that exports increase a country's wealth whereas imports reduce it. Why is this view of the economy misleading?
3. According to Benjamin Franklin, 'No nation was ever hurt by trade'. Is this true?
4. What are the main features of globalisation? Can these features be measured? Is it possible to generate a robust single measure of globalisation?
5. Australia and the United States have signed a bilateral free trade agreement. Are all such moves towards free trade always beneficial? What are likely to be the main efficiency advantages and disadvantages of this proposed agreement? Can these advantages and disadvantages be quantified?
6. What problems arise in trying to determine whether globalisation affects economic growth?
7. Why do open economies tend to grow faster than closed economies? Given your answer, should governments ever put restraints on openness?
8. How is globalisation likely to affect the main tax rates for mobile and immobile factors of production?
9. What are the advantages and disadvantages of international corporate tax competition? Overall, is this competition desirable?
10. Globalisation may both reduce global inequality and increase inequality in rich countries. True or false? If true, is this desirable?
11. How does globalisation influence government expenditure?
12. Economists generally argue that government should work with market forces rather than work against them. Is this also true of global economic forces?

Further Reading

Dreher, A. and Gaston, N. (2008) 'Has globalisation increased inequality?', *Review of International Economics*, 16, 515–536.

Gemmell, N., Kneller, R. and Sanz, I. (2008) 'Foreign investment, international trade and the size and structure of public expenditures', *European Journal of Political Economy*, 24, 151–171.

Harrison, A. (2006) *Globalisation and Poverty*, Working Paper 12347, National Bureau of Economic Research, Cambridge, MA.

Organisation for European Cooperation and Development, 2005, *Measuring Globalisation: OECD Handbook on Economic Globalisation Indicators*, OECD, Paris.

Organisation for European Cooperation and Development, 2011, *An Overview of Growing Income Inequalities in OECD Countries: Main Findings*, OECD, Paris.

Snower, D.J., Brown, A.J.G. and Merkl, C. (2009) 'Globalisation and the welfare state: a review of Han-Werner Sinn's *Can Germany be Saved*?', *Journal of Economic Literature*, 47, 136–158.

Stiglitz, J. (2002) *Globalization and its Discontents*, Chapters 1 and 9, Penguin Books, New York.

Taylor, J.B., Moosa, I. and Cowling, B. (2000) *Microeconomics*, Chapters 17–19, John Wiley, Brisbane.

World Bank (2002) *Globalisation, Growth and Poverty*, World Bank, Washington DC.

Epilogue

Gouverner, c'est choisir.

Gaston, Duc de Lévis, 1812, *Maximes politiques*

Everybody has won, and all must have prizes.

The Dodo, *Alice's Adventures in Wonderland*, Lewis Carroll

The fundamental problem of government is the fundamental problem of economics: wants exceed resources. Therefore, choices must be made. In the words of the Duc de Lévis two hundred years ago: "to govern is to choose".

Some people, often philosophers or ascetics and occasionally economists, propose to solve the fundamental problem of economics by eliminating wants. There has long been a view that the route to happiness lies in the elimination of desire. Indeed, one of the messages of the economics of happiness literature (Layard, 2005) is that, despite all the goods we enjoy today, people in developed economies are no happier than they were 50 years ago. However, without an extraordinary change in human nature, it is a very safe prediction that our wants will continue to exceed our means of supplying them.

The market response to the fundamental problem of economics is to use prices to balance demand with supply. This does not mean that wants equal supply. Rather, prices ration and allocate scarce resources.

For many goods, competitive markets provide an efficient response to the fundamental economic problem. By trading freely in markets, both for factors of production and for goods, individuals can choose how to work and spend their income to maximise their own welfare. As Adam Smith showed over two hundred years ago, remarkably, in competitive markets individuals generally act not only to their own benefit but also to the benefit of the community as well. And there is no doubt that markets have encouraged the innovation and energy that have driven world economic growth over the centuries.

However, there are many circumstances (market failures) where markets fail to supply the goods that we want. More critically, given an unequal endowment of resources across individuals, in many cases from birth or the first few years of life, or simply due to misfortune, income and goods are often distributed most unevenly. Therefore, some form of collective action is essential to achieving efficient and socially inclusive outcomes.

As the famous parable of the Good Samaritan suggests, some collective action may be achieved by individual effort without government intervention. However, for most purposes, government at some level has a central and indispensable role in the rules by which we live and in the provision of social welfare.

But, as have seen, government is not a simple process. Collective decision-making processes are complex. And, regrettably, with good reason politicians are often not trusted to improve the situation.

Some readers may recall Lewis Carroll's satire of the political process in the Caucus race in *Alice's Adventures in Wonderland* (Lewis Carroll, 1865). To get dry after a swim in the river, the Dodo proposed that Alice and everyone else in the party should run a Caucus race, in which they would run in patterns of any shape, starting and leaving off wherever they wished.[1] After they had been running around for half an hour or so and were quite dry again, the Dodo called out 'the race is over', whereupon everyone crowded round and demanded to know who had won the race. After a period of silence and thought, the Dodo announced at last, 'Everybody has won, and all must have prizes'.

Giving out prizes seems to be a universal political practice. In the words of Nikita Kruschev, Chairman of the Soviet Union from 1958 to 1964: 'Politicians are the same the whole world over. They promise to build a bridge even when there is no river.' And Kruschev did not even have to get elected other than by a few party officials!

On the other hand, economics is a science of choice. We try to show how individuals make choices in markets and how governments do and could make choices over the use and distribution of resources for the betterment of individuals in our society.

This is an important task though generally a less popular task than handing out prizes! But. hopefully, the study of public economics will encourage a rational and generous view about how governments can secure and enhance our welfare both as individuals and as a community.

[1] The Dodo was the nickname that Lewis Carroll, whose real name was Lewis Dodgson, gave to himself in his stories. As a professor of mathematics at Oxford University, Lewis Dodgson wrote among other things about the voter paradox, the impossibility of rational collective choice.

Glossary

Access pricing A charges for access to a service.

Accrual accounting Expenses and income are recorded when they occur, not when they are paid.

***Ad valorem* tax** A tax on the value of sales.

Additive utilitarian social welfare function Social welfare expressed as the sum of individuals' utilities.

Adverse selection A problem in insurance markets that occurs because those taking up insurance contracts are likely to be worse risks than those who do not take up insurance.

Allocative efficiency Allocating resources so as to produce the optimal mix of goods.

Arrow's Impossibility Theorem No collective decision-making rule compiles fully with six generally accepted ethical criteria for voting.

Asymmetric information A situation in which economic agents do not have the same information.

Benchmarking Comparing an agency's costs with the costs of similar providers.

Benefit-cost ratio The ratio of the present value of net recurrent benefits to the present value of capital expenditure.

Black economy Unreported commercial transactions that do not enter into gross domestic product calculations.

Budget constraint The limit to expenditure.

Budget sector Government services that are financed out of the annual government budget.

Capital gain An increase in the value of an asset.

Capitalisation The process whereby future income or payments are incorporated into the price of an asset.

Categorical income transfer A payment that depends on the recipient belonging to a specified category of person.

Certainty equivalent The amount of income that an individual would give up in exchange for an uncertain prospect.

Coase theorem Private actions will correct for externalities.

Common property resources Non-excludable resources available freely to anyone who wants to use them.

Comparative advantage Exists when the opportunity cost of production of a good (in terms of other goods forgone) is lower in one country than in another. The concept applies generally to all economic agents.

Compensated demand The demand for goods as a function of price, while holding utility constant.

Compensated labour supply The supply of labour as a function of wages when the income effect of wage changes is removed.

Compensating variation The amount of money that ensures an individual is as well off with a change as he or she is without it.

Computable general equilibrium model A model with equations for demand and supply for all sectors of the economy.

Concentration ratio A measure of the proportion of a market taken up by the largest firms in it.

Consolidated revenue The government account into which all taxes are paid unless they are ear-marked for a special purpose.

Constitution Defines the basic institutions and rules of the state and the methods of appointment and powers of the head of state and government.

Constant prices A common set of prices used to value economic variables at different points in time.

Consumer surplus The difference between the maximum amount that a consumer is willing to pay for a good and the amount that he or she actually pays.

Contestable market A market in which entry and exit are free.

Contingent valuation A survey method used to elicit individual values for a non-market good, contingent on it being available.

Corrective tax A tax that corrects for market failure by adjusting the marginal private cost of a good to internalise a negative externality.

Cost-benefit analysis An evaluation method that compares the total social costs and benefits of a proposal as far as possible in money terms.

Cost-effective analysis A technique for comparing the costs of alternative proposals to find the minimum cost solution which achieves the given objective.

Cost function A function relating total cost to the level of output.

Cross-sectional data Data for individuals, firms, regions or countries at a given point in time.

Current prices Using the prices prevailing or forecast for each point in time to measure economic magnitudes at that time.

Deadweight loss A loss for which there is no offsetting benefit.

Depreciation The loss of value of a capital asset in any given period.

Destination taxes Income or consumption is taxed where received or consumed at point of destination.

Direct democracy A public choice in which all adult citizens are entitled to participate by voting.

Direct tax A tax that is borne by the entity paying the tax.

Discount rate The interest rate used to discount future sums to present sums.

Earned income tax credit A tax credit per dollar of earned income.

Economic incidence of taxation Indicates who actually bears the tax.

Economic (pure) profit The excess of a firm's revenue over the opportunity cost of all inputs. cf: accounting profit which is the return on equity invested.

Economic rent The return to any factor of production in excess of its opportunity cost. For a firm, economic rent is the same as economic profit.

Effective tax rate Actual taxes paid divided by the taxable base.

Effective marginal tax rate The percentage of an extra dollar that an individual loses due to tax and loss of grants.

Elasticity The percentage change in one variable brought about by a one per cent change in another variable.

Elasticity of (technical) substitution A measure of the responsiveness of input combinations, such as labour and capital, to changes in their relative prices, holding output constant.

Equal marginal sacrifice rule Tax should be arranged so that all persons paying tax have an equal marginal sacrifice.

Equilibrium Occurs when no economic agent has an incentive to change their behaviour.

Equity When used in welfare economics, a term that is synonymous with fairness.

Equivalent taxes Two or more sets of taxes with the same incidence.

Equivalent variation The amount of money that ensures an individual is at least as well off without a change as he or she is with it.

Excess burden The loss of welfare that occurs due to the imposition of a tax, over and beyond the tax collected.

Excise tax A tax on consumption of a selected commodity.

Expenditure substitute A tax concession that substitutes for direct government expenditure.

Experimental studies Studies based on a sample of individuals randomly assigned to a treatment and a control group.

Externality Any effect that market exchanges have on firms or individuals who do not participate in those exchanges.

Factor tax A tax on a factor used in the production process, such as capital or labour.

Factor price equalisation theorem Given free trade and no transport costs, prices of factors of production will be equalised across trading partners.

Financial analysis An account of the cash flows of a project for a specified agency.

Fiscal illusion Occurs when an agent does not understand the real impact of a fiscal change.

Fixed cost Costs independent of the level of production.

Flat-rate tax A tax where the marginal tax rate is constant across all levels of income.

Free riding Occurs when someone enjoys the benefits of a good that is provided by others.

General equilibrium A model of an economy that portrays the operation of many markets simultaneously.

General purpose grant A grant with no conditions attached.

Generational accounting Measuring the effects of a policy on the members of each generation.

Gini coefficient Measures the degree of inequality as the ratio of the area between the Lorenz curve for a given income distribution and the 45 degree line of equality and the total area under the line of equality.

Globalisation The worldwide integration of markets.

Goods and services tax An Australian tax that is levied on the value added at each stage of production (see also value-added tax).

Head count ratio The number of households below the poverty line as a proportion of total households.

Horizontal equity Individuals in similar positions should be treated equally.

Horizontal fiscal imbalance Occurs when sub-national governments have unequal fiscal capacity to provide public services.

Horizontal tax competition Tax competition between two jurisdictions to attract or retain resources.

Human capital model An individual invests in education to increase the present value of their lifetime income.

Imperfect competition Occurs when an individual buyer or seller can influence the market price.

Imputed rent The monetary value of services that an owner of an asset receives from use of the asset, for example from owner-occupation of a dwelling.

Income (Haig-Simons definition) The value of goods that can be consumed in any period without any change in net wealth.

Income effect The change in consumption due to a change in real income.

Income in-kind Income in the form of goods and services rather than cash payments.

Indexation The use of an index, such as the consumer price index, to adjust nominal prices and maintain constant real values.

Indifference curve The set of consumption bundles that yields the same total utility to an individual.

Indirect tax A tax levies on one party that may be borne by another party.

Inflation tax The effect of inflation on the value of an individual's monetary assets.

Internal rate of return The rate of return that equates discounted net benefits to discounted capital costs.

Internalisation Where the marginal private benefit, or cost, of goods is adjusted so that the users consider the actual marginal social benefit or cost of their decisions.

Intransitivity Inconsistent rankings of preferences.

Isoquant A curve which shows the different combinations of inputs that produce a given quantity of output.

Labour intensive A firm or industry in which the ratio of labour to capital inputs is relatively high.

Linear tax structure A tax structure that has a constant marginal tax rate at all levels of income.

Local public good A public good that provides benefits to residents in a local geographical area.

Logrolling The trading of votes to obtain passage of a package of legislative proposals.

Long run marginal cost The cost of supplying an additional unit of output in the long run.

Lorenz curve A curve that plots the proportion of total income received against the proportion of the adult population or income recipients from the poorest to richest.

Low income tax offset An increase in the threshold at which low earners are taxed.

Lump sum tax A tax that is independent of an individual's behaviour.

Marginal Additional, extra, incremental.

Marginal cost The incremental cost of producing one more unit of output.

Marginal physical product The addition to output in physical units per unit increase in any factor input.

Marginal rate of substitution The rate at which an individual is willing to trade one good for another while remaining equally well off.

Marginal rate of transformation The amount by which output of one good can be increased if the output of another good is reduced by a small amount.

Marginal revenue The additional revenue obtained with the sale of one more unit of output.

Marginal revenue product The addition to revenue from a unit increase in factor input.

Marginal social benefit The benefit accruing to all parties from production of an extra unit of a good.

Marginal social cost The cost accruing to all parties from production of an extra unit of a good. .

Marginal tax rate The extra tax paid expressed as a percentage of an additional dollar value of the tax base.

Marginal wage cost The cost of employing an extra worker.

Mark–up pricing Prices are marked up until revenues are sufficient to cover costs.

Market disequilibrium A situation in which the plans of buyers and sellers do not match and some economic agents have an incentive to alter their behaviour.

Market failure Occurs when a market fails to generate an efficient outcome.

Market power When a firm can raise price above long-run average cost and obtain a return in excess of competitive rates of return.

Matching grants Grants that that are conditional on the recipient providing a specified contribution.

Median voter The voter whose preferences lie in the middle of the set of all voters' preferences.

Median voter theorem In some specified conditions, the preference of the median voter is decisive

Merit goods Goods that government considers individuals should consume even if they do not demand them.

Minimum wage regulation Regulation that sets the minimum wage at which firms can employ labor.

Mixed public goods Goods with some private good and some public good characteristics.

Moral hazard Occurs when insurance against the consequences of risk changes an individual's behaviour.

Multi-criteria analysis An evaluation method in which weights are attached to various objectives, and to how well these objectives are met, in order to assess the relative merits of alternative projects.

Natural monopoly A single firm that can supply a market's total output at lower unit cost than two or more firms can.

Negative gearing Occurs when the interest payment on a property loan creates a negative net income from the property.

Negative income tax system A tax system that includes income grants as well as taxes.

Net present value The estimated value of a stream of benefits net of costs discounted to present value terms.

Net public debt Financial liabilities of government less financial assets.

Net operating budget Current operating revenue minus current expenses.

Net public worth The value of all publicly-owned assets less liabilities at any point in time.

Nominal interest rates Market interest rates inclusive of allowance for inflation.

Non–categorical income transfer A payment that depends only on the income of the recipient and not on whether the recipient belongs to a particular category.

Non-excludable good Individuals cannot be excluded from consuming the good.

Non-linear pricing Where consumers are charged different prices per unit of consumption depending on their level of consumption.

Non-linear tax structure A tax structure which has varying marginal tax rates.

Non-rival good A good is non-rival in consumption when consumption by one individual does not reduce the availability of the good to others.

Normative economics Economic analysis that requires a value judgement.

Numeraire A standard unit of measure used to make comparisons.

Observational studies Studies based on data from observed economic behaviour

Opportunity cost The value of output or leisure foregone by using a resource, or by behaving, in one way rather than another.

Opportunity set The set of outcomes that can be achieved.

Ordinal preferences Rankings or orderings of alternatives.

Ordinary (or observed) demand curve A demand curve that shows how demand varies with price.

Pareto efficient An allocation of resources such that a reallocation cannot make anyone better off without making someone else worse off.

Pareto improvement A reallocation of resources that makes at least one person better off and no one worse off.

Partial equilibrium A model of a single market that ignores possible spillover effects into other markets.

Partial factor tax A tax that is levied on a factor in only some of its uses.

Participation ratio The ratio of the number of persons in the workforce to the number of persons aged 15-64.

Payroll tax A tax levied on the value of wages and salaries paid by employers to employees.

Pecuniary externalities Actions that impact on third parties which are transmitted through the price mechanism.

Perfect competition Where there are a large number of buyers and sellers in the market for any good and each agent is a price taker.

Pigouvian tax A tax levied on each unit of a polluter's output in an amount equal to the marginal damage that it inflicts at the efficient level of pollution.

Point utility possibilities frontier The utilities obtained from alternative distributions of one basket of goods.

Positive economics Economic analysis of how an economy actually functions

Poverty line The level of income required for a minimally adequate standard of living.

Predatory pricing Occurs when a firm underprices a product with the object of limiting competition.

Present value The capital value now of outcomes occurring in the future.

Price discrimination Occurs when a seller can separate markets and charge different prices for the same product in the separate markets.

Price taker An agent unable to affect the price of a good.

Principle-agent problem Occurs when an agent pursues his own interests at the expense of the principal.

Private good A good that is excludable and rival in consumption.

Private time preference rate The rate at which an individual is willing to trade present for future consumption.

Privatisation Occurs when ownership of assets changes from the public to the private sector.

Producer surplus The difference between the price that a producer receives and production cost.

Product (output) mix efficiency Occurs when an economy produces the mix of goods that individuals most want. A necessary condition is that the marginal rate of transformation equals the marginal rates of substitution of consumers.

Production function An equation that shows the relationship between a firm's output and its inputs.

Production possibilities frontier The maximum quantity of goods that can be produced with given resources and technology.

Productive (technical) efficiency Producing a given output of goods with minimum use of resource (or producing maximum output from given resources).

Progressive tax system A tax system where the average tax rate rises as the tax base increases.

Property right An entitlement defining the owner's rights and limitations for use of a resource, including the power to exclude others from using the resource.

Proportional tax system A tax system in which the average tax rate is the same for all values of the tax base.

Public choice theory Uses methods of economic analysis to analyse public choices.

Public good A good that is non-excludable and non-rival in consumption.

Public sector Includes the budget sector and public trading enterprises.

Purchaser–provider split Occurs when delivery of a program is separated from its funding.

Quasi experimental studies Studies that draw on observational data where external or natural events randomly assign individuals to different groups.

Rate of return regulation Regulation of prices based on a rate of return that a regulator allows firms to make.

Real income Income measured in constant prices.

Real interest rates Nominal interest rate corrected for changes in the level of prices.

Regressive tax system A tax system where the average tax rate falls as the tax base rises.

Regulatory capture Occurs when government regulates an industry for the benefit of the producers.

Rent Payment for use of land or buildings.

Rent seeking Attempting to gain economic rents.

Rent-seeking costs Resources expended by firms in order to gain economic rents.

Replacement rates Income from social assistance as a percentage of income from employment.

Representative democracy A system in which voters elect representatives to make collective decisions on their behalf.

Revealed preference valuations Inferring preferences or willingness to pay values from behaviour.

Revenue sharing Occurs when various levels of government share tax revenue on a predetermined basis, with no restrictions on the use of the funds.

Risk aversion A preference for safe, low returns over prospects with higher mean returns but more variance.

Risk pooling Occurs when investments are pooled over a large number of projects and total risk is minimised.

Second-best theorem When there is more than one market failure, the correction of one market failure may not increase economic efficiency.

Selective tax A tax that is levied on only part of the tax base, or that allows exemptions and deductions from the general tax base.

Sensitivity analysis A method of assessing risk by indicating how the estimated net present value of a project changes with variations in the values of particular inputs.

Shadow prices The underlying real economic cost of an economic variable.

Short run marginal cost The cost of supplying an additional unit of output in the short run.

Social indifference curve The combination of outcomes between which society is indifferent.

Social opportunity cost of capital The full rate of return on capital, including the return to government.

Social time preference rate The discount rate that maximises inter-temporal social welfare.

Social welfare The collective welfare of society.

Social welfare function The level of social welfare expressed as a function of the utility of each member of society or of other factors that contribute to welfare.

Source taxes Income or production is taxed at source.

Specific purpose grants Grants with conditions attached to the use of the grant.

Stated preference Eliciting individuals' willingness to pay for goods from their stated values or choices.

Statutory incidence Indicates who is legally responsible for paying a tax.

Subsidiarity principle Public services should be supplied at the level of government closest to the users of the service unless this is not cost-effective.

Substitution effect The change in consumption of a good due to a change in relative prices.

Substitute goods Two goods are substitutes if a change in the price of one good leads to a change in demand for another good.

Taxable income The amount of income that is subject to tax.

Tax avoidance The legal rearrangement of business activities so as to minimise tax liabilities.

Tax base The economic activity or item on which a tax is levied.

Tax bracket Range of income subject to a given marginal tax rate.

Tax capitalisation The process by which a stream of present and expected future taxes is incorporated into the present capital value of an asset.

Tax deduction Expense or allowance that may be deducted from income and so reduces taxable income.

Tax effort The ratio of tax collections to tax capacity.

Tax evasion Noncompliance with the tax laws.

Tax expenditure A tax concession.

Tax exporting Shifting some of the burden of taxation onto firms or households in other jurisdictions.

Tax hypothecation Links the revenue from a tax to a specified kind of expenditure.

Tax imputation Recipients of dividends receive tax credits for profit tax paid by the company.

Tax indexation Automatic adjustment of the tax schedule to compensate for inflation.

Tax rate The rate at which a tax is levied on the tax base.

Tax rate structure The relationship between the taxes collected over a given period and the tax base, typically categorised as proportional, progressive or regressive.

Tax rebate Deductions from income tax payments.

Tax schedule Tax liability associated with each level of taxable income.

Tax unit The unit on which the tax is levied.

Tax wedge The difference between what consumers pay for a good and what producers receive for it (after tax).

Theorem of comparative advantage Two parties gain from trade whenever the parties have a comparative advantage in the production of some good.

Time series data Data for a variable or set of variables at different times.

Transaction costs The costs associated with negotiating, monitoring and enforcing contracts.

Transfer payment A payment that is independent of the performance of any service.

Transfers in-kind Governments payments to individuals in the form of goods or services rather than as cash payments.

Transfer pricing The price that one part of an organisation charges to another part.

Transitional equity Fairness in changing policies or taxes.

Transitivity Consistent ranking of alternatives.

Turnover taxes Taxes levied at a fixed rate on the gross value of sales.

Two–part tariff A pricing system where users are charged both for access to a system and for its use.

Unit tax A given amount of tax per unit of commodity purchased.

Unitary state A state where the central government has supreme authority to make laws for the whole country.

User charges Prices paid for goods or services.

Utility A term for a persons' welfare or satisfaction..

Utility function An equation that relates an individual's utility to the amounts of goods that he or she consumes (or to their income and leisure).

Utility possibilities frontier Shows the maximum utility that one person can achieve given the utility of the other person.

Value-added tax A percentage tax on value added at each stage of production.

Variable costs Costs that change in response to changes in the level of output.

Vertical equity Individuals with greater economic capacity should pay more tax than individuals with lower capacity.

Vertical fiscal imbalance The imbalance between the expenditure responsibilities and taxing powers of different levels of government.

Wage tax credit Payment to a firm to subsidise wage payments usually for low wage rate workers.

Welfare economics Economic theory concerned with the social desirability of alternative economic states.

Willingness to accept The minimum amount that an individual is willing to accept as compensation for the loss of a good and be no worse off than before.

Willingness to pay The maximum amount an individual is willing to pay for a good and be as well off as he is without it.

References

These references do not include the further readings at the end of each chapter unless they are referred to in the text.

Aaron, H.A. and Gale W.G. (eds) (1996) *Economic Effects of Fundamental Tax Reform*, Brookings Institute, Washington, DC.

Abelson, P. (1981a) 'Some benefits of small local government areas', *Publius*, Winter, 129–39.

Abelson, P. (1981b) *The Role and Size of Local Government Authorities in New South Wales*, Occasional Paper No.19, Centre for Research on Federal Financial Relations, Australian National University.

Abelson, P. (2003) 'The value of life and health for public policy', *Economic Record*, 79, Special Edition, S2–S16.

Abelson. P. (2005) *Outsourcing of Public Services in Australia: Seven Case Studies*, Macquarie Economics Research Paper, No. 3, Macquarie University, Sydney.

Abelson, P. (2008), *Establishing a Monetary Value for Lives Saved: Issues and Controversies*, Working paper 2008–02, Office of Best Practice Regulation, Department of Finance and Deregulation, Canberra.

Abelson, P. (2010) 'The high cost of taxi regulation with special reference to Sydney', *Agenda: A Journal of Policy Analysis and Reform*, 17, 73–94.

Abelson, P. (2011) 'Estimating the revenue raising capacity of the states and territories and the implications for the equitable distribution of GST revenue', *Economic Papers*, *30*, *1–12*.

Abelson, P. and Joyeux, R. (2007) 'Price and efficiency impacts of taxes and subsidies for Australian housing', *Economic Papers*, 26, 147–69.

Abelson, P., Joyeux, R. and Mahuteau, S. (2012) 'Modelling house prices across Sydney with estimates for access, property size, public transport, urban density and crime', *Australian Economic Review*.

Abelson, P. and R. Joyeux (2015) "Smoke and Mirrors: Fallacies of the NSW Government Views on Local Government Financial Capacity", *Public Money and Management*, July, 315-20.

Abelson, P. (2016) "The Optimal Size of Local Government with Special Reference to New South Wales", *Agenda*, A Journal of Policy Analysis and Reform, 23, 1, 31-46.

Acemoglu and Robinson, J.A. (2013), *Why Nations Fail, The Origins of Power, Prosperity and Poverty*, Crown Publishers, Random House, USA.

Acemoglu, D., Johnson, S. and Robinson, J. (2005) 'Institutions as a fundamental cause of long-run economic growth', in P. Aghion and S.N. Durlauf (eds), *Handbook of Economic Growth,* Elsevier, Amsterdam.

Access Economics (2008) *The Growing Cost of Obesity in 2008: Three Years On*, prepared for Diabetes Austria.

Aghion, P. and Howitt, P. (2009) *The Economics of Growth*, MIT Press, Cambridge, MA.

Aghion, P., Alesina, A. and Trebbi, F. (2008) 'Democracy, technology and growth', in E. Helpman (ed.), *Institutions and Economic Performance*, Harvard University Press, Cambridge, MA.

Aghion, P., Blundell, R., Griffith, R., Howitt, P. and Prantl, S. (2006) 'The Effects of Entry on Incumbent Innovation and Technology', CEPR Discussion Paper No. 5323.

Aghion, P., Burgess R., Redding S. and Zilibotti, F. (2006) 'The Unequal Effects of Liberalisation: Evidence from Dismantling the License Raj in India', CEPR Discussion Paper No. 5492.

Akerloff, G. (1970) 'The market for lemons: quality uncertainty and the market mechanism', *Quarterly Journal of Economics*, 80, 488–500.

Albouy, D. (2010) *Evaluating the Efficiency and Equity of Federal Fiscal Equalisation*, Working Paper 16144, National Bureau of Economic Research, New York.

Alesina, A. and Wacziary, R. (1998) 'Openness, country size and government', *Journal of Public Economics*, 69, 305–21.

Anand, S. and Segal, P. (2008) 'What do we know about global income inequality?', *Journal of Economic Literature*, 46, 57–94.

Andersson, H. (2005) 'The value of safety as revealed in the Swedish car market: an application of the hedonic pricing approach', *Journal of Risk and Uncertainty*, 30: 3, 211–39.

Angrist, J.D. and Krueger, A.B. (1991) 'Does compulsory school attendance affect schooling and earnings?', *Quarterly Journal of Economics*, 106, 979–1014.

Angrist, J., Bettinger, E., Bloom E., King, E. and Kremer, M. (2001) *Vouchers for Private Schooling in Columbia: Evidence from a Randomised Natural Experiment*, Working Paper 8343, National Bureau of Economic Research, Cambridge, MA.

Applied Economics (2003a) *Cost-Benefit Analysis of Proposed New Health Warnings on Tobacco Products*, report for the Commonwealth Department of Health and Ageing.

Applied Economics (2003b) *Returns on Investment in Public Health: An Epidemiological and Economics Evaluation*, report ISBN 0 6428219 1 7 to Department of Health and Ageing, Canberra (see www.health.gov.au).

Apps, P. (2006) *Family Taxation: An Unfair and Inequitable System*, Working Paper, Law School, Sydney University, Sydney.

Apps, P. (2010) 'Why the Henry Review fails on family tax reform', pp. 103–128 in C. Evans, R. Krever and P. Mellow eds *Australia's Future Tax System: The Prospects after Henry*, Thomson Reuters, Sydney.

Apps, P. and Rees, R. (2010) 'Australian tax reform and the targeting fallacy', *Australian Economic Review*, 43, 153–175.

Araujo, S. and Sutherland, D. (2010) *Public Private Partnerships and Investment in Infrastructure*, OECD Economics Department Working Papers No. 803, OECD, Paris.

Aristei, D., Bacchettini, V. and Calzoni, G. (2009) 'Income Distribution, Growth and Well-being: Evidence for Selected OECD Countries', *Argumenta Oeconomica*, 22, 5–54.

Aristotle, c.340 BC, *The Politics*, reprinted in 1962, Penguin Classics.

Arnott, R. and Grieson, R.E. (1981) 'Optimal fiscal policy for state and local government', *Journal of Urban Economics*, 9, 23–48.

Arrow, K.J. (1951) *Social Choice and Individual Values*, John Wiley, New York.

Arrow, K. (1973) 'Higher education as a filter', *Journal of Public Economics*, 2, 193–216.

Arrow, K.J. and Lind R.C (1970) 'Uncertainty and the evaluation of public investment decisions', *American Economic Review*, 60, 364–78.

Aschauer, D.A. (1989) 'Is public expenditure productive?' *Journal of Monetary Economics*, 23, 177–200.

Ashenfelter, O. and Krueger, A. (1994) 'Estimates of the economic return to schooling from a New York sample of twins', *American Economic Review*, 84, 1157–1173.

Ashenfelter, O., Harmon, C. and Oosterbeek, H. (1999) 'A review of estimates of the schooling/earnings relationship, with tests for publication bias', *Labour Economics*, 6, 453–478.

Atkinson, A.B. (1970) 'On the measurement of income inequality', *Journal of Economic Theory*, 2, 244–263.

Atkinson, A.B. (1995) *Public Economics in Action, The Basic Income/Flat Tax Proposal*, Clarendon Press, Oxford.

Atkinson, A.B. (1997) 'Bringing income distribution in from the cold', *Economic Journal*, 107, 297–321.

Atkinson, A.B. and Leigh, A. (2007) 'The distribution of top incomes in Australia', *Economic Record*, 83, 247–61.

Atkinson, A.B., Piketty, T. and Saez, E. (2011) 'Top incomes in the long run of history', *Journal of Economic Literature*, 49, 3–71.

Atkinson, A.B. and Stiglitz, A.B. (1976) 'The design of tax structure: direct versus indirect taxation', *Journal of Public Economics*, 6, 55-75.

Atkinson, A.B. and Stiglitz, J.E. (1980) *Lectures on Public Economics*, McGraw-Hill, Maidenhead, England.

Atkinson, A.B. (2015) *Inequality: What can be Done?*, Harvard University Press, Cambridge, M.A.

Attanasio, O.P., Banks, J., Meghir, C. and Weber, G. (1995) *Humps and Bumps in Lifetime Consumption*, Working Paper 5350, National Bureau of Economic Research, Cambridge, MA.

Auerbach, A.J. (2008) *Federal Budget Rules: The US Experience*, Working Paper No. 14288, National Bureau of Economic Research, Cambridge, MA.

Auerbach, A.J. and Slemrod, J. (1997) 'The economic effects of the Tax Reform Act of 1986', *Journal of Economic Literature*, 35, 589–632.

Australian Bureau of Statistics (2005) *Household Expenditure Survey, Australia: Summary of Results, 2003–04*, Cat. No. 6530.0, AGPS, Canberra.

Australian Bureau of Statistics (2007) *Government Benefits, Taxes and Household Income*, Cat. No. 6530.0, AGPS, Canberra.

Australian Bureau of Statistics (2009) *Not-for-profit Organisations, Australia, 2006–07* (Re-issue), Cat. No. 8106.0, AGPS, Canberra.

Australian Bureau of Statistics (2009 and 2011) *Household Income and Income Distribution in Australia,* Cat. No. 6523.0, AGPS, Canberra.

Australian Bureau of Statistics (2017) *Household Income and Income Distribution, 2015-16*, Cat. No. 6523.0, AGPS, Canberra.

Australian Competition and Consumer Commission (2011) *Merger Guidelines*, November 2008 updated, ACCC, Melbourne.

Australian Energy Regulator (2016) *United Energy – Determination, 2016-20 – Final Decision*.

Australian Institute of Health and Welfare (2010) *Australia's Health*, AIHW, Canberra.

Australian Taxation Office (2007) *Tax Basics for Non-Profit Organisations*, NAT 7966–06, 2005, Canberra.

Australian Taxation Office (2011) *Taxation Statistics*, ATO, Canberra.

Austroads (2003) *Valuing Environmental and Other Externalities*, AP–R229, Austroads, Melbourne.

Autor, V.J. and Duggan, M. (2006) *The growth in the social security disability rolls: a fiscal crisis unfolding*, Working Paper 12436, National Bureau of Economic Research, Cambridge, MA.

Bailey, S.J. (2002) *Public Sector Economics: Theory Policy and Practice*, Macmillan, London.

Bajada, C. (2001) *The Cash Economy and Tax Reform*, Australian Tax Research Foundation, Sydney.

Bardsley, P. and Olekhans, N. (1999) 'Cigarette and tobacco consumption: have anti-smoking policies made a difference?' *Economic Record*, 75, 225–40.

Barr, N. (1992) 'Economic Theory and the Welfare State: A Survey and Interpretation', *Journal of Economic Literature*, 30, 741–803.

Barr, N. (2004) *The Economics of the Welfare State*, 4th ed., Oxford University Press, Oxford.

Barr, N. and Whynes, D. (1993) 'Introductory Issues', in N. Barr and D. Whynes (eds), *Current Issues in the Economics of Welfare*, Macmillan, London, pp. 1–19.

Barrett, G. (2012), 'The Return to Cognitive Skills in the Australian Labour Market', *Economic Record*, 88, 1–17.

Barro, R.J. (1974) 'Are government bonds net worth?', *Journal of Political Economy*, 82, 1095–1117.

Barro, R.J. (1978) *The Impact of Social Security on Private Saving*, American Enterprise Institute of Public Policy Research, Washington, DC.

Barro, R.J. (1991) 'Economic growth in a cross section of countries', *Quarterly Journal of Economics*, 106, 407–443.

Barro, R.J. (1997) *Determinants of Economic Growth: A Cross Country Empirical Study*, MIT Press, Cambridge, MA.

Bateman, H., Kingston, G. and Piggott, J. (2001) *Forced Saving, Mandating Private Retirement Incomes*, Cambridge University Press, Cambridge.

Bateman, I., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D., Sugden, R. and Swanson, J. (2002) *Economic Valuation with Stated Preference Techniques*: *A Manual*, Edward Elgar, Cheltenham, UK.

Baumol, W.J. (1967) 'Macroeconomics of unbalanced growth: the anatomy of urban crisis', *American Economic Review,* 62, 307–322.

Baumol. W.J. (2002) *The Free-Market Innovation Machine*, Princeton University Press, Princeton, NJ.

Baumol, W.J., Panzar, J. and Willig, R. (1982) *Contestable Markets and the Theory of Industry Structure*, Harcourt Brace Jovanovich, New York.

Becker, G. and Murphy, G.M. (1988) 'A theory of rational addiction', *Journal of Political Economy*, 96, 675–700.

Becker, G. and Chiswick, B.R. (1966) 'Education and the distribution of earnings' *American Economic Review*, 56, 358–369.

Bennett, J. (1996) 'Estimating the recreational use values of national parks', *Tourism Economics*, 2, 303–320.

Bernheim, B.D. (2002) 'Taxation and Saving', in A.J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics*, Vol. 3, Elsevier Science, Amsterdam, pp. 1173–1249.

Bernheim, B.D. and Rangel, A., 2007, 'Behavioural Public Economics: Welfare and Policy Analysis with Non-standard Decision Makers', pp. 7–84 in P. Diamond and H. Vartiainen (eds) *Behavioural Economics and its Applications,* Princeton University Press, Princeton, N.J.

Bernheim, B.D. and Scholze, J. (1993) 'Private saving and public policy', *Tax Policy and the Economy*, 7, 73–110.

Besley, T.J. and Rosen, H.S. (1998) 'Vertical externalities in tax setting: evidence from gasoline and cigarettes', *Journal of Public Economics* 70, 383–398.

Bettinger, E., Kremer, M. and Saavedra, J.E. (2010), 'Are educational vouchers only redistributive?', *Economic Journal*, 120, F204–F228.

Beveridge, W. (1942) *Report to the [UK] Parliament on Social Insurance and Allied Services*, London.

Bhattacharyya, A., Parker, E. and Raffiee, K. (1994) 'An examination of the effect of ownership on the relative efficiency of public and private utilities', *Land Economics*, 70, 197–209.

Bhattacharyya, D.K., and Wassmer, R.W. (1995) 'Fiscal dynamics and elected officials', *Public Choice*, 83, 221–249.

Bickel, P., Friedrich, R., Link, H., Stewart, L. and Nash, C. (2006) 'Introducing environmental externalities into transport pricing: measurement and implications', *Transport Reviews*, 26, 389–415.

Bils, M. and Klenow, P. (2000) 'Does schooling cause growth or the other way around?' *American Economic Review*, 90, 1160–1183.

Black, J. (2002) *A Dictionary of Economics*, 2nd ed, Oxford University Press, Oxford.

Blankart, C.B. (2000) 'The process of government centralization: a constitutional view', *Constitutional Political Economy*, 11, 29–37.

Blau, F.D. and Kahn, L.M. (2007) "Changes in the labor supply behavior of married women: 1980-2000", *Journal of Labor Economics,* 393–438.

Blaug, M. (1989) 'Review of economics of education: research and studies', *Journal of Human Resources,* 24, 331–335.

Blomquist, G. (1979) 'Value of life savings: implications of consumption activity', *Journal of Political Economy*, 87, 540–558.

Blomquist, G. Miller, T. and Levy, D. (1996) 'Values of risk reduction implied by motorist use of protection equipment: new evidence from different populations', *Journal of Transport Economics and Policy*, 29, 55–66.

Blomquist, S. and Selin, H. (2010) 'Hourly wage rates and taxable labour income responsiveness to changes in marginal tax rates', *Journal of Public Economics*, 94, 878–889.

Bloom, H.S., Ladd, H.F. and Yinger, J. (1983) *Property Prices and House Values*, Academic Press, New York.

Blundell, R. (1996) 'Labour supply and taxation', in M.P. Devereux (ed,) *The Economics of Tax Policy*, Oxford University Press, Oxford, pp. 107–136.

Boardman, A., Greenberg, D., Vining A. and Weimer, D. (2011), *Cost-Benefit Analysis, Concepts and Practice*, 4th ed., Prentice Hall, NJ.

Boardman, A. and Vining, A. (1989) 'Ownership and performance in competitive environments: a comparison of the performance of private, mixed and state-owned enterprises, *Journal of Law and Economics*, 32, 1–33.

Bohm, P. (1972) 'Estimating demand for public goods: an experiment', *European Economic Review,* 3, 773–776.

Borcherding, T.E. (1985) 'The causes of government expenditure growth: a survey of the US evidence', *Journal of Public Economics*, 28, 359–382.

Borcherding, T.E., Burnaby, B.C., Pommerehne, W.W. and Schneider, F. (1982) 'Comparing the efficiency of private and public production: the evidence from five countries', *Zeitschrift fur Nationalokonomie*, 89, 127–56.

Borland, J. (1999) 'Earnings inequality in Australia: changes, causes and consequences', *Economic Record*, 75, 177–202.

Boskin, M.J. (1975) 'Efficiency aspects of the differential tax treatment of market and household economic activities', *Journal of Public Economics*, 4, 1–25.

Boskin, M.J. (1978) 'Taxation, saving and the rate of interest', *Journal of Political Economy*, 86, S3–S27.

Boston, J., Martin J., Pallot, J. and Walsh, P. (1996) *Public Management: the New Zealand Model*, Oxford University Press, Oxford.

Bowles, S. and Gintis, H. (2002) 'Social capital and community governance', *Economic Journal*, 112, F419–F436.

Braden, J.B. and Kolstad, C.D. (eds) (1991) *Measuring the Demand for Environmental Quality*, Elsevier, Amsterdam.

Brennan, G. and Pincus, J. (1998) 'The myth of vertical fiscal imbalance', 27th Annual Conference of Economists, Sydney.

Brittle, S. (2010) 'Ricardian equivalence and the efficacy of fiscal policy in Australia', *Australian Economic Review*, 43, 254–269.

Brunner, E.J and Ross, S.L. (2010) 'Is the median voter decisive? Evidence from referenda voting patterns', *Journal of Public Economics*, 94, 898–910.

Buchanan, J.M. (1965) 'An economic theory of clubs', *Economica*, 32, 1–14.

Buchanan, J.M. (1975) 'Public finance and public choice', *National Tax Journal*, 28, 383–394.

Buchanan, J.M. and Tullock, G. (1962) *The Calculus of Consent*, University of Michigan Press, Ann Arbor, MI.

Buchanan and Musgrave (1999) *Public Finance and Public Choice: Two Contrasting Visions of the State*, MIT Press, Cambridge, USA.

Buhmann, B., Rainwater, L., Schmaus G. and Smeeding, T.M. (1989) 'Equivalence scales, well-being, inequality and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database', *Review of Income and Wealth*, 34, 115–142.

Bureau of Transport and Communications Economics (1996) *Benefits of Private Sector Involvement in Road Provision: A Look at the Evidence*, Working Paper 33, BTCE, Canberra.

Burgess, R. and Stern, N. (1993) 'Taxation and development, *Journal of Economic Literature*, 31, 762–830.

Burtless, G. (1995) 'International trade and the rise in earnings inequality', *Journal of Economic Literature*, 33, 800–816.

Butler, J.R.G. (2010) 'New opportunities in health finance and governance', *Australian Economic Review*, 43, 1–76.

Campbell, A. and Bond, K. (1997) 'The cost of public funds in Australia', *Economic Record*, 74, 22–34.

Carcillo, S. and Grubb, D. (2006) *From Industry to Work: The Role of Active Labour Market Policies*, Directorate for Employment, Labour and Social Affairs, OECD, Paris.

Card, D. (1999) 'The causal effect of education on earnings', in O. Ashenfelter and D. Card (eds) *Handbook of Labour Economics*, Vol. 3A, Elsevier, Amsterdam, pp. 1801–1863.

Card, D. and Krueger, A.D. (1994) 'The economic return to school quality: a partial survey', Working Paper No. 379, Industrial Relations Section, Princeton University, Princeton, NJ.

Cardak, B.A. and Ryan, C. (2006) *Why are high ability individuals from poor backgrounds under-represented at universities?*, Discussion Paper No. A06.04, School of Business, La Trobe University, Melbourne.

Carling, K. and Richardson, K. (2001) *The relative efficiency of labour market programs: Swedish experience from the 1990s*, IFAU—Office of Labour Market Policy Evaluation, Working Paper 2001:2.

Carroll, L. (1865) *Alice's Adventures in Wonderland.*

Carson, R. and Louviere, J.J. (2010) 'Experimental design and the estimation of willingness to pay in choice experiments for health policy evaluation', in E. McIntosh, P.M. Clarke, E.J. Frew and J.J. Louviere (eds) *Applied Methods of Cost-Benefit Analysis in Health Care*, Oxford University Press, New York, pp. 185–210.

Cashin, P. (1995) 'Government Spending, Taxes, and Economic Growth', *IMF Staff Papers*, 42, 237–269.

Cassou, S.P. (1997) 'The link between tax rates and foreign direct investment,' *Applied Economics,* 29, 1295–1301.

Caves, R.E. and Krause, L.B. (eds) (1984) *The Australian Economy: A View from the North*, Allen & Unwin, Sydney.

Chapman, B. (2005) *Income Contingent Loans for Higher Education: International Reform*, Discussion Paper No 491, Centre for Economic Policy Research, Australian National University, Canberra.

Chevalier, A. and Lanot, G. (2002) 'The relative effect of family characteristics and financial situation on educational achievement', *Education Economics*, 10, 161–185.

Chirinko, R.S. and Wilson, D.J. (2008) 'State investment tax incentives: a zero sum game?', *Journal of Public Economics*, 92, 362–384.

Coase, R.H. (1960) 'The problem of social cost', *Journal of Law and Economics*, 3, 1–44.

Coelli, T., Rao, P. and Battese, G. (1998) *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Dordrecht, Holland.

Colegrave, A.D and Giles, M.J. (2008) 'School cost functions: a meta-regression analysis', *Economics of Education Review*, 27, 688–696.

Commission of Inquiry into Third London Airport (1971) *Report*, HMSO, London.

Commonwealth Competitive Neutrality Complaints Office (1998) *Rate of Return Issues*, CCNCO Research Paper, Canberra.

Conneely, K. and Uusitalo, R. (1999) *Estimating heterogeneous treatment effects in the Becker schooling model*, Working Paper, Department of Economics, Helsinki.

Connolly, S. and Munro, A. (1999) *Economics of the Public Sector*, Pearson, London.

Cordes, J.J. and Watson, H.S. (1998) 'The theory and practice of tax incidence', in F. Thompson and M.T. Green (eds), *Handbook of Public Finance*, Marcel Dekker, New York, pp. 93–134.

Corlett, W.J. and Hague, D.C. (1953) 'Complementarity and the excess burden of taxation', *Review of Economic Studies*, 21, 21–390.

Cornia, G.A. (ed) (2004) *Inequality, Growth and Poverty in an Era of Liberalisation and Globalisation*, Oxford University Press, New York.

Covick, O. (2004) 'Put not your trust(s) in tax reform: rather do the opposite', *Economic Papers*, 23, 257-270.

Cowling, K. and Mueller, D.C. (1978) 'The social costs of monopoly power', *Economic Journal*, 88, 727–748.

Creedy, J. (2004) 'The excess burden of taxation', *Australian Economic Review*, 37, 454–464.

Creedy, J. (2010) 'Tax-and-transfer tensions: Designing direct tax structures', *Australian Economic Review*, 43, 103–113.

Cromwell, J. and Mitchell, J. (1986) 'Physician induced demand for surgery', *Journal of Health Economics*, 5, 293–313.

Crosby, M., Brown, D. and Malady, C. (1997) 'Political and partisan cycles in the Australian economy', *Australian Economic Papers*, 36, 42–54.

Crowle, J. and Turner, E. (2010) *Childhood Obesity: An Economic Perspective*, Productivity Commission Staff Working Paper, Melbourne.

Cullis, J. and Jones, P. (1998) *Public Finance and Public Choice*, Oxford University Press, Oxford.

Currie, J. and Gahvari, F. (2008) 'Transfers in cash and in-kind: theory meets the data', *Journal of Economic Literature*, 46, 333–383.

Cusack, T.R. (1997) 'Partisan politics and public finance: changes in public spending in the industrialised democracies, 1995-1989', *Public Choice*, 91, 375–395.

Cutler, D.M. (2002a) 'Health care and the public sector', in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Volume 4, Elsevier Science, Amsterdam.

Cutler, D.M. (2002b) 'Equality, efficiency, and market fundamentals: the dynamics of international medical-care reform', *Journal of Economic Literature*, 40, 881–906.

Dahlby, B. and Wilson, L. (1996) 'Tax assignment and fiscal externalities in a federal state', in P. Boothe (ed.) *Reforming Fiscal Federalism for Global Competition: A Canada-Australia Comparison*, University of Alberta Press, Canada.

Dandie, S. and Mercante, J. (2007) *Australian Labour Supply Elasticities: Comparison and Critical Review,* Working Paper 2007–04, Treasury, Canberra.

Dardis, R., 1980, 'The value of a life: new evidence from the marketplace', *American Economic Review*, 70, 1077–82.

Davidoff, I. and Leigh, A. (2008) 'How much do public schools really cost? Estimating the relationship between house prices and school quality', *Economic Record*, 265, 193–206.

Dawkins, P., Beer, G. Harding, A., Johnson, D. and Scutella, R. (1998) 'Towards a negative income tax system for Australia', *Australian Economic Review*, 31, 237–57.

Dawkins, P., Duncan A.S. and Freebairn, J. (2003) 'Modifying income support in the Australian tax and transfer system: some options and an evaluation', Centre for Policy Evaluation Working Paper 1/04, University of Nottingham.

Day, K. (1992) 'Interprovincial migration and local public goods', *Canadian Journal of Economics*, 92, 123–44.

Day, B. and Prades, J-L., P. (2010) 'Ordering anomalies in choice experiments', *Journal of Environmental Economics and Management*, 59, 271–285.

Deacon, R.T. (1978) 'A demand model for the local public sector', *Review of Economics and Statistics*, 60, 184–92.

Deaton, A. (1979) 'Optimally uniform commodity taxes', *Economics Letters*, 2, 357–361.

Denny, K. and Harmon, C. (2000) *The impact of education and training on labour market experiences of young adults*, Working Paper 00/08, Institute of Fiscal Studies, London.

Department of Finance and Administration (2000) *The Outcomes and Outputs Framework Guidance Document*, DoFA, Canberra.

Department of Finance and Administration (DoFA) (2006) *Handbook of Cost-benefit Analysis*, 2nd ed., DoFA, Canberra.

Devereux, M.P. and Griffith, R. (1998) 'Taxes and the location of production: evidence from a panel of US multinationals,' *Journal of Public Economics,* 68, 335–367.

Dewey, J., Husted, T.D. and Kenny, L. (2000) 'The ineffectiveness of school inputs: a product of misspecification?', *Economics of Education Review*, 19, 27–45.

Diamond, P. (1977) 'A framework for social security analysis, *Journal of Public Economics*, 8, 211–224.

Diamond, P. (1998) 'Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates', *American Economic Review*, 88, 83–95.

Diamond, P. and Hausman, J. (1994) 'Contingent valuation: is some number better than no number?', *Journal of Economic Perspectives*, 8, 45–64.

Diamond, P. and Mirrlees, J.A. (1971), 'Optimal taxation and public production I: production efficiency and II: tax rules', *American Economic Review*, 61, 8–27 and 261–278.

Diamond, P. and Vartiainen, H. (2007) 'Introduction', in Diamond, P. and Vartiainen, H. (eds) *Behavioural Economics and its Applications,* Princeton University Press, Princeton, NJ, pp. 1-6.

Dickert, S., Hauser, S. and Scholz, J.K. (1995) 'The earned income tax credit and transfer programs: a study of labour market and program participation', in J. M. Poterba (ed.) *Tax Policy and the Economy*, National Bureau of Economic Research, MIT Press, pp. 1–50.

Diewart, W. and Lawrence, D.A. (1994) *The Marginal Cost of Taxation in New Zealand*, Swan Consultants, Report for the NZ Roundtable, Canberra.

Diewert, W.E., Lawrence, D.A. and Thompson, F. (1998) 'The marginal costs of taxation and regulation', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York, pp. 135–171.

Dixon, P., Picton, M.R. and Rimmer, M.T. (2002) 'Efficiency effects of inter-governmental grants in Australia', *Australian Economic Review*, 35, 304–15.

Dixon, R., Gunther, A. and Mahmood, M. (2001) 'The cost of monopoly in Australian manufacturing', *Australian Economic Review*, 34, 390–402.

Dobes, L. and J. Leung (2015) 'Wider Economic Benefits in Transport Infrastructure Cost-Benefit Analysis, - A Bridge too Far?' *Agenda*, 22, 1, 75-95.

Doiron, D. and Kalb, D. (2005) 'Demands of childcare and household labour supply in Australia', *Economic Record,* 81, 215–236.

Dollar, D. (2001) *Globalisation, Inequality, and Poverty since 1980*, Development Research Group, World Bank, Washington, DC.

Dollar, D. and Kraay, A. (2001) *Growth is Good for the Poor*, Policy Research Working Paper No.2587, World Bank, Washington, DC.

Domberger, S. (1994) *Competitive Tendering and Contracting in the NSW Public Sector: The 1994 Survey Findings*, Graduate School of Business, University of Sydney.

Domberger, S. and Piggott, J. (1986) 'Privatisation policies and public enterprise: A survey', *Economic Record*, 173, 145–162.

Domberger, S., Hall, C. and Li, E.A.L. (1995) 'The determinants of price and quality in competitively tendered contracts', *Economic Journal*, 105, 1454–1470.

Donato, R. and Scotton, R. (1998) 'The Australian health care system', in G. Mooney and R. Scotton (eds) *Economics and Australian Health Care Policy*, Allen & Unwin, Sydney, pp. 20–39.

Dougherty, C. and Jorgensen, D.W. (1996) 'International comparisons of the sources of economic growth', *American Economic Association, Papers and Proceedings*, 86, 25–29.

Douglas, N. and B. O'Keeffe (2016) 'Wider Economic Benefits – When and if they should be used in evaluation of transport projects', *Australian Transport Research Forum Proceedings*.

Downs, A. (1957) *An Economic Theory of Democracy*, Harper & Row, New York.

Dowrick, S. (1992) 'Technological catch-up and diverging incomes: patterns of economic growth', *Economic Journal*, 102, 600–610.

Dowrick, S. (1996) 'Estimating the impact of government consumption on growth: growth accounting and endogenous growth models', *Empirical Economics*, 21, 163–186.

Dowrick, S. (2003) *Ideas and Education: Levels and Growth Effects*, Working Paper No. 9709, National Bureau of Economic Research, Cambridge, MA.

Dreher, A. and Gaston, N. (2008) 'Has globalisation increased inequality?', *Review of International Economics*, 16, 515–536.

Dupuit, J. (1844) 'De la mesure de l'utilité des travaux publics', *Annals des Ponts et Chauses*, 2nd series, Vol. 8.

Durlauf, S.N. (2002) 'On the empirics of social capital', *Economic Journal*, 112, F459–F479.

Durlauf, S.N., Johnson, P.A. and Temple, J.R.W. (2005) 'Growth econometrics', in P. Aghion and S.N. Durlauf (eds) *Handbook of Economic Growth*, Elsevier, Amsterdam, pp. 555–677.

Eaton, B.C., Eaton, D.F. and Allen, D.W. (2002) *Microeconomics,* 5th ed., Prentice Hall, Toronto.

Economic Planning and Advisory Commission (1995) *Private Infrastructure Task Force, Interim Report,* EPAC, AGPS, Canberra.

Economic Planning and Advisory Commission (1996) *Future Child Care Provision in Australia*, Task Force Final, AGPS, Canberra.

Edmonds. E. (2001) 'Will child labour decline with improvement in living standards?', Dartmouth College Working Paper No. 01–09.

Edwards, S. (1998) 'Openness, productivity and growth: what do we really know?', *Economic Journal*, 108, 383–398.

Ehrlich, I. and Kim, J. (2007) *Has social security influenced family formation and fertility in OECD countries? An economic and econometric analysis*, Working Paper 12896, National Bureau of Economic Research, Cambridge, MA.

Eissa, N. and Liebman, J.B. (1996) 'Labour supply response to the earned income tax credit', *Quarterly Journal of Economics*, 111, 605–637.

Emy, H.V. and Hughes, O.E. (1991) *Australian Politics: Realities in Conflict*, Macmillan, Sydney.

Engen, E. and Skinner, J. (1992) 'Fiscal policy and economic growth', *Working Paper No. 4223*, National Bureau of Economic Research, Cambridge, MA.

Ferraro, P.J. and Vossler, C.A. (2010) 'The source and significance of confusion in public goods experiments', *B.E. Journal of Economic Analysis and Policy, Contributions*, 10, 1–40.

Feldstein, M. (1996) 'The missing piece in policy analysis: social security reform', *American Economic Review*, 68, 1–14.

Feldstein, M. (1996) 'Social security and savings: new time series evidence', *National Tax Journal*, 49, 151–164.

Feldstein, M.S. (2008) *Effects of Taxes on Economic Behaviour*, Working Paper 13745, National Bureau of Economic Research, New York.

Feldstein, M. and Liebman, J.B. (2002) '*Social security*' in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Vol. 4, Elsevier Science, Amsterdam, pp. 2246–2323.

Financial Crisis Inquiry Commission (2011) *The Financial Inquiry Crisis Report: Final Report of the National Commission on the Causes of the Financial and Economic Crisis in the United States*, US Government Printing Office.

Foster, C.D. (2005) *British Government in Crisis*, Hart Publishing, London.

Freebairn, J. (1995) 'Reconsidering the marginal welfare costs of taxation', *Economic Record*, 71, 121–131.

Freebairn, J. (1998) 'Efficiency Issues' in P. Abelson (ed), *The Tax Reform Debate*, Allen & Unwin, Sydney.

Freebairn, J. (2005) *Equity in Survey Revenue: A Comparison of Alternative Tax Bases*, Working Paper 11/2005, Melbourne Institute of Applied Economic and Social Research, Melbourne.

Freeman, A.M. (1986) 'On assessing the state of the arts of the contingent valuation method of valuing environmental changes', in R.G. Cummings, D.S. Brookshire and W.D. Schulze (eds) *Valuing Environmental Goods,* Rowman & Allanheld, Totowa, NJ.

Freeman, R. (1991) 'How much has de-unionization contributed to the rise in male earnings inequality?', NBER Working Paper No. 3826, August.

Freyens, B. and Oslington, P. (2007) 'Dismissal costs and their impact on employment: evidence from Australian small and medium enterprises', *Economic Record*, 83, 1–15.

Friedman, L.S. (2003) *The Microeconomics of Public Policy*, Princeton University Press, Princeton, NJ.

Friedman, M. and Friedman, R. (1980) *Free to Choose*, Penguin Books, Melbourne.

Fu, S. (2007) 'Smart café cities; testing human capital externalities in the Boston Metropolitan Area', *Journal of Urban Economics*, 61, 86–111.

Funk, P. and Gathmann, C. (2011) 'Does direct democracy reduce the size of government? New evidence from historical data, 1890–2000', *Economic Journal*, 121, 1252–1280.

Fukuyama, F. (1999) *The Great Disruption*, Simon & Schuster, New York.

Galbraith, J.K. (1955) *The Great Crash, 1929*, Hamish Hamilton, London.

Gans, J.S. and Leigh, A. (2006) *Born on the First of July: An (Un)natural Experiment in Birth Timing*, Discussion Paper No. 529, Centre for Economic Policy Research, Australian National University, Canberra.

Garrod, G. and Willis, K.G. (1999), *Economic Evaluation of the Environment: Methods and Case Studies*, Edward Elgar, Cheltenham, UK.

Garnaut, R. and Fitzgerald, V. (2002) *Review of Commonwealth–State Funding,* Treasuries of NSW, Victoria and Western Australia.

Garrett, G. (1995) 'Capital mobility, trade, and the domestic politics of economic policy', *International Organisation*, 49, 657–687.

Garrett, G. (1998) *Governing the Global Economy: Economic Policy and Market Integration around the World*, mimeo, Yale University.

Garrett, G. and Mitchell, D. (1997) *International Risk and Social Insurance: Reassessing the Globalisation-Welfare State Nexus*, mimeo, Yale University.

Gassner, K., Popov, A. and Pushak, N. (2008) *An Assessment of Private Sector Participation in Electricity and Water and Sanitation Services in Developing and Transition Countries*, Public Private Infrastructure Advisory Facility, World Bank, Washington, D.C.

Gaston, N. and Rajaguru, G. (2009) 'The long-run determinants of Australian income inequality', *Economic Record*, 85, 260–275.

Gautier, A.H. and Hatzius, J. (1997) 'Family benefits and fertility: an econometric analysis,' *Population Studies*, 51, 295–306.

Gemmell, N., Kneller, R. and Sanz, I. (2008) 'Foreign investment, international trade and the size and structure of public expenditures', *European Journal of Political Economy*, 24, 151-171.

George, H. (1879) *Progress and Poverty*, Garden City, New York.

Glaeser, E., Laibson, D. and Sacerdote, B. (2001) *The Economic Approach to Social Capital*, Discussion Paper No. 1926, Harvard Institute of Economic Research, Cambridge, MA.

Glaeser, E. and Ujhelyi, G. (2010) 'Regulating misinformation', *Journal of Public Economics*, 94, 247–257.

Glaister, S. and Lewis, D.L. (1978) 'An integrated fares policy for transport in Greater London', *Journal of Public Economics*, 9, 341–353.

Glittleman, M. and Wolff, E.N. (1993) 'International comparisons of inter-industry wage differentials', *Review of Income and Wealth*, 39, 295–312.

Godspeed, T.J. (2000) 'Tax structure in a federation', *Journal of Public Economics*, 75, 493–506.

Goldberg, P.K., and Maggi, G. (1999) 'Protection for sale: an empirical investigation', *American Economic Review*, 89, 1135–1155.

Gomez-Ibanez, J.A. (1999) 'Pricing', in J.A. Gomez-Ibanez, W.B. Tye and C. Winston (eds) *Essays in Transportation Economics and Policy, A Handbook in Honour of John R. Meyer*, Brookings Institution Press, Washington, DC.

Gong, X., Breunig, R. and King, A. (2011) *Partnered Women's Labour Supply and Child Care Costs in Australia: Measurement Error and the Child Care Price*, Discussion Paper No. 652, Centre for Economic Policy Research, Australian National University, Canberra.

Goodwin, P.B. (1992) 'A review of new demand elasticities with special reference to short and long run effects of price changes', *Journal of Transport Economics and Policy*, 26, 155–170.

Goos, M. and Manning, A. (2007) 'Lousy and lovely jobs: the rising polarisation of work in Britain', *Review of Economics and Statistics*, 89, 118–33.

Goos, M., Manning, A. and Salomons, A. (2010) *Explaining Job Polarisation in Europe: The Roles of Technology, Globalisation and Institutions*, Centre for Economic Performance Discussion Paper No. 1026, LSE, London.

Gottschalk, P. and Smeeding, T. (1997) 'Cross national comparisons of earnings and income inequalities', *Journal of Economic Literature*, 35, 633–687.

Gramlich, E.M. (1987) 'Subnational fiscal policy', in M.J. Quigley (ed.), *Perspectives on Local Public Finance and Public Policy*, JIA Press, Greenwich, CT, pp. 3–27.

Grilliches, Z. (1977) 'Estimating the returns to schooling: some econometric problems', *Econometrica*, 45, 1–22.

Groves, T. and Loeb, M. (1975) 'Incentives and public inputs', *Journal of Public Economics*, 3, 211–226.

Gruber, J. (2011) *Public Finance and Public Policy*, 3rd ed., Worth Publishers, New York.

Gruber, J. and Wise, D.A. (1999) 'Introduction and Summary', in J. Gruber and D.A. Wise (eds) *Social Security and Retirement Programs around the World*, University of Chicago Press, Chicago.

Gruen, F.M. (1982) 'The welfare expenditure debate: economic myths of the left and right', *Economic Record*, 58, 207–223.

Guasch, J.L. (2000) *Concessions bust of boom? An empirical analysis of ten years of experience in concessions in Latin America and Caribbean*, The World Bank Institute, Washington, DC.

Guest, R. and McDonald, I. (2002) 'Would a decrease in fertility be a threat to living standards in Australia?', *Australian Economic Review*, 35, 29–44.

Guiso, L., Sapienza, P. and Zingales, L. (2005), *Trusting the Stock Market*, NBER Working Paper 11648, National Bureau of Economic Research, New York.

Gujarati, D.N. (2003) *Basic Econometrics*, McGraw-Hill, New York.

Hackl, F., Schneider, F. and Withers, G.A. (1993) 'The public sector in Australia, a quantitative analysis', in N. Gemmell (ed.) *The Growth of the Public Sector, Theories and International Evidence*, Edward Elgar, Aldershot, UK.

Hall, R.E. and Jones, C.I. (1997) 'Levels of economic activity across countries. What have we learned from recent empirical growth research?', *American Economic Association, Papers and Proceedings*, 87, 173–177.

Hancock, W.K. (1930) *Australia*, reprinted in 1961, Jacaranda, Brisbane.

Hanemann, W.M. (1991) 'Willingness to pay and willingness to accept: how much can they differ?', *American Economic Review*, 81, 635–647.

Hanson, G.H. (2008) *The Economic Consequences of the International Migration of Labour*, Working Paper 14490, National Bureau of Economic Research, Cambridge, MA.

Hanushek, E.A (1986) 'The economics of schooling: production and efficiency in public schools', *Journal of Economic Literature*, 24, 1141–1177.

Hanushek, E.A. (2002) 'Publicly provided education', in A.J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics*, Vol. 4, Elsevier Science, Amsterdam, pp. 2046–2141.

Hanushek, E.A., (2010) *The Economic Value of Higher Teacher Quality*, Working Paper No. 16606, National Bureau of Economic Research, New York.

Hanushek, E.A. and Woessmann, L. (2010) *How Much Do Educational Outcomes Matter in OECD Countries?* Working Paper No. 16515, National Bureau of Economic Research, New York.

Harberger, A.C. (1954) 'Monopoly and resource allocation', *American Economic Review*, 45, 77–87.

Harberger, A.C. (1962) 'The incidence of the corporation tax', *Journal of Political Economy*, 70, 125–140.

Harberger, A.C. (1974) 'The incidence of the corporation income tax', in A.C. Harberger (ed.) *Taxation and Welfare*, Little Brown, Boston.

Harberger, A.C. (1990) 'Reflections on uniform taxation', in R.W. Jones and A.O. Krueger (eds) *The Political Economy of International Trade, Essays in Honour of Robert E. Baldw*in, Basil Blackwell, Cambridge, MA, pp.75-89.

Hardin, G. (1968) 'The tragedy of the commons', *Science*, 162, 1243–1248.

Harding, A., Vu, Q.N., Payne, A. and Percival, R. (2009) 'Trends in effective marginal tax rates in Australian from 1996–97 to 2006–07', *Economic Record*, 85, 449–461.

Harmon, C. and Walker, I. (1995) 'Estimates of the return to schooling for the United Kingdom', *American Economic Review*, 85, 1278–1286.

Harrington, W., Krupnick, A. and Spofford, W. (1989) 'The economic losses of a waterborne disease outbreak', *Journal of Urban Economics*, 25, 116–127.

Harrison, A. (2006) *Globalisation and Poverty*, Working Paper 12347, National Bureau of Economic Research, Cambridge, MA.

Harsanyi, J.C. (1955) 'Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility', *Journal of Political Economy*, 73, 309–321.

Hart, O., Shleifer, A. and Vishny, R.W. (1996) 'The proper scope of government: theory and an application to prisons', Working Paper 5744, National Bureau of Economic Research, Cambridge, MA.

Haveman, R. (1996) *Reducing Poverty while Increasing Employment: A Primer on Alternative Strategies, and a Blueprint*, OECD Economic Studies, No. 26, 7–41.

Haveman, R. and Wolfe, B. (2000) 'The economics of disability and disability policy', in A.J. Culyer and J.P. Newhouse (eds) *Handbook of Health Economics*, Elsevier Science, Amsterdam, pp. 995–1051.

Hawke, L. (2007) 'Performance budgeting in Australia', *OECD Journal on Budgeting*, 7, 2–15.

Head, J.G. (1974) *Public Goods and Public Welfare*, Duke University Press.

Heady, C. (1996) 'Optimal taxation as a guide to policy', in M.P. Devereux (ed.) *The Economics of Tax Policy*, Oxford University Press, Oxford, pp. 23–54.

Heckman, J.J. (1993) 'What has been learned about the labour supply in the United States in the past twenty years?', *American Economic Review*, 83, 116–121.

Heckman, J.J., Lalonde, R.J. and Smith, J.A. (1999) 'The economics and econometrics of active labour market programs', in O. Ashenfelter and D. Card (eds) *Handbook of Labour Economics*, Vol. 3A, Elsevier Science, Amsterdam.

Heckman, J.J, Moon, S.S., Pinto, R., Savelyev, P.A. and Yavitz, A. (2010) 'The rate of the return to the Highscope Perry Preschool Program', *Journal of Public Economics*, 94, 114–128.

Hedges, L.V., Laine, R.D. and Greenwald, R. (1994) 'Does money matter? A meta-analysis of studies of differential school inputs on student outcomes', *Educational Researcher*, 23, 5–14.

Henderson, R.F. (Chair), Commission of Inquiry into Poverty (1975) *First Main Report, Poverty in Australia*, AGPS, Canberra.

Henry, K. (Chair) (2010) *Australia's Future Tax System, Report to the Treasurer*, Commonwealth of Australia, Canberra.

Hensher, D.A. (1987) 'Productive efficiency, ownership and contracting out of transport services', in P. Abelson (ed.), *Privatisation: An Australian Perspective*', Australian Professional Publications, Sydney.

Hensher, D.A. (2007) 'Productivity in the urban bus sector', Chapter 13 in *Bus Transport*, Elsevier, Amsterdam.

Hensher, D.A. (2011) 'Valuation of travel time savings', in A. de Palma, R. Lindsey, E. Quinet, R. Vickerman (eds), *Handbook in Transport Economics,* Edward Elgar, Aldershot, UK, pp. 135–158.

Hensher D.A., Rose J.M., de Dios Ortuzar, J. and Rizzi Li (2009) 'Estimating the willingness to pay and value of risk reduction for car occupants in the road environment', *Transportation Research Part A: Policy and Practice*, 43, 692–707.

Hibbs, D.A. (1987) *The Political Economy of Industrial Democracies*, Harvard University Press, Cambridge, Mass.

Hicks, J.R. (1935) 'The theory of monopoly', *Econometrica*, 3, 1–20.

Hicks, J.R. (1939) 'Foundations of Welfare Economics', *Economic Journal*, 49, 696–712.

Hilmer, F. Rayner M., and Taprell, Q. (1993) *National Competition Policy*: *Towards a New Order*, Report by the Independent Committee of Inquiry, AGPS, Sydney.

Hindricks, J. and Myles, G.D. (2006) *Intermediate Public Economics*, MIT Press, Cambridge, MA.

Hirshleifer, J. (1983) 'From weakest link to best-shot: the voluntary provision of public goods', *Public Choice*, 41, 371–386.

Ho, J.J., Liu, J-T. and Wang, J-D. (2005) 'Stated preferences for the removal of physical pain resulting from permanently disabling occupational injuries: A contingent valuation study of Taiwan', *Accident Analysis and Prevention*, 37, 537–548.

Hoang, M., Bolto, B., Haskard, C., Barron, O., Gray, S. and Leslie, G. (2009) *Desalination in Australia*, National Research Flagships, CSIRO, Sydney.

Hobbes, T. (1651) *Leviathan*, reprinted in R. Tuck (ed.), 1991, Cambridge University Press, Cambridge.

Hodge, G.A. (2000) *Privatization, An International Review of Performance*, Westview Press, Boilder, CO.

Hoehn, J. P. and Randall, A. (1987) 'A satisfactory benefit cost indicator from contingent valuation', *Journal of Environmental Economics and Management*, 17, 213–229.

Holcombe, R.G. (1983) 'Applied fairness theory: comment', *American Economic Review*, 73, 1153–1156.

Holcombe, R.G. (1998) 'The Foundations of Normative Public Finance', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York.

Horowitz, J.K. and McConnell, K.E. (2002) 'A review of WTA/WTP studies', *Journal of Environmental Economics and Management*, 44, 426–447.

Hotze, V.J., Mullin, C.H. and Scholz, J.K. (2002) 'Welfare employment and income: evidence on the effects of benefit reductions in California', *American Economic Review Papers and Proceedings*, 380–384.

Hotze, V.J., Mullin, C.H. and Scholz, J.K. (2006) *Examining the effect of the earned income tax credit on the labour market participation of families on welfare*, Working Paper 11968, National Bureau of Economic Research, Cambridge, MA.

Hoxby, C.M. (1994) *Does competition among public schools benefit students and taxpayers*?, Working Paper 4979, National Bureau of Economic Research, Cambridge, MA.

Hoxby, C. and Rockoff, J.E. (2004) *The Impact of Charter Schools on Student Achievement*, Working Paper, Harvard Institute of Economic Research.

Hurrell, A. and Woods, N. (1995) 'Globalisation and Inequality', *Journal of International Studies*, 24, 447–470.

Hyman, D.N. (1999) *Public Finance*, 6th ed., Harcourt Brace, Orlando, FL.

Imber, D., Stevenson, G. and Wilks, L. (1991) *A Contingent Valuation Survey of the Kakadu Conservation Zone*, Research Paper, No. 3, Resource Assessment Commission, Canberra.

Industry Commission (1994) *Urban Transport*, Report No. 37, AGPS, Canberra.

Industry Commission (1995) *The Growth and Revenue Implications of Hilmer and Related Reforms*, Report to the Council of Australian Governments, AGPS, Canberra.

Infrastructure Partnerships Australia (2007) *Performance or PPPs and Traditional Procurement in Australia*, Infrastructure Partnerships Australia, Sydney.

Ingles, D. (1997) 'Low income traps for working families', Discussion Paper No. 363, Centre for Economic Policy Research, Australian National University.

Ingles, D. (1998) 'Overcoming anomalies in the interaction of tax and social security', *Australian Economic Review*, 31, 271–280.

Inman, R.P. (1989) 'The local decision to tax: evidence from large U.S. cities', *Regional Science and Urban Economics*, 19, 455–491.

Inman, R.P. and Rubinfield, D.L. (1996) 'Designing tax policy in federalist economies: an overview', *Journal of Public Economics*, 60, 307–334.

International Monetary Fund (2017) 'A Greater Role for Fiscal Policy', Chapter 1 in *Fiscal Monitor,* IMF, Washington.

Intriligator, M., Bodkin, R. and Itsiao, C. (1996) *Econometric Models, Techniques and Applications*, 2nd ed., Prentice Hall International, New Jersey.

Irwin, T. and Yanamoto, C. (2004) *Some Options for Improving the Governance of State-Owned Electricity Utilities*, Discussion Paper No, 11, Energy and Mining Sector Board, World Bank, Washington, DC.

James S. and Nobes, C. (2002) *The Economics of Taxation*, 7th ed., Pearson, London.

Jay, P. (2000) *Road to Riches or the Wealth of Man*, Weidenfeld & Nicolson, London.

Jha, R. (1998) *Modern Public Economics*, Routledge, London.

Johansen, L. (1977) 'The theory of public goods: misplaced emphasis', *Journal of Public Economics*, 7, 147–152.

Johnes, G. (1993) *The Economics of Education*, Macmillan, London.

Johnson, D. (1995) 'Economics and health services', *Australian Economic Review*, 28, 123–139.

Johnson, F.R., Banzhaf, M.R. and Desvouges, W.H. (1999) 'Willingness to pay for improved respiratory and cardiovascular health: a multiple format, stated preference approach', *Health Economics*, 9, 295–317.

Johnson, R. and Libecap, G. (1989) 'Agency growth, salaries, and the protected bureaucrat' *Economic Inquiry*, 27, 5–50.

Jones, C.I. (2002) *Introduction to Economic Growth*, W.W. Norton, New York.

Jones, R, and Whalley, J. (1988) 'Regional effects of taxes in Canada: an applied general equilibrium approach', *Journal of Public Economics*, 37, 1–28.

Jones–Lee, M.W. (1994) 'Safety and the saving of life: the economics of safety and physical risk', in R. Layard and S. Glaister (eds) *Cost-Benefit Analysis*, Cambridge University Press, Cambridge.

Jones–Lee, M. and Loomes, G. (2004) 'Eliciting measures of value for health and safety', 9th Australian Workshop on Safety-Related Programmable Systems, Brisbane, *Conferences on Research and Practice in Information Technology*, Vol. 47.

Jorgenson, D.W. and Wilcoxen, P.J. (1990) 'The economic impact of the Clean Air Act amendments of 1990', *Energy Journal*, 14, 159–82.

Jud, G.D. and Watts, J.M. (1981) 'Schools and housing values', *Land Economics*, 57, 459–470.

Kahnemann, D. and Knetsch, J.C. (1992) 'Valuing public goods: the purchase of moral satisfaction', *Journal of Environmental Economics and Management*, 22, 57–70.

Kahnemann, D. and Tversky, A. (1979) 'Prospect theory: an analysis of decisions under risk', *Econometrica*, 47, 263–291.

Kalb, G. (2007) 'Interaction of the tax and social security systems in Australia: the effect on effective marginal tax rates', *Australian Economic Review*, 31, 186–193.

Kaldor, N. (1939) 'Welfare propositions of economics and interpersonal comparisons of utility', *Economic Journal,* 49, 549–552.

Kamerschen, D.R. (1966) 'An estimation of the welfare losses from monopoly in the American economy', *Western Economic Journal*, 4, 221–236.

Kaplow, L. (2008) *The Theory of Taxation and Public Economics*, Princeton University Press, Princeton, NJ.

Kay, J.A. and Thompson, D.J. (1986) 'Privatisation: A policy in search of a rationale', *Economic Journal*, 96, 18-32.

Keating, M. and Lambert, S. (1998) 'Improving incentives: changing the interface of tax and social security', *Australian Economic Review*, 31, 281–9.

Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, 1964, Macmillan, London.

Khursheed, A. and Borcherding, T. (1998) 'Organising government supply: the role of bureaucracy', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York, pp. 43–92.

King, S. and Pitchford, R. (1998) 'Privatisation in Australia: understanding the incentives in public and private firms', *Australian Economic Review*, 31, 313–328.

King, S. (1997) 'National competition policy', *Economic Record*, 73, 270–284.

KPMG Econtech (2010) *CGE Analysis of the Current Australian Tax System*, Final Report to the Henry Tax Review (2010).

Krupnick, A., Alberini, A., Cropper, M, Simon, N., O'Brien, B., Goeree, R. and Heintzelman, M. (2000) *Age, Health and Willingness to Pay for Mortality Risk Reduction: A Contingent Valuation of Ontario Residents*, Discussion Paper 0–0–37, Resources for the Future, Washington, DC.

Krueger, A.B. (1998) *Experimental Estimates of Education Production Functions*, Working Paper No.379, Industrial Relations Section, Princeton University, Princeton, NJ.

Krueger, A. N. (2003) 'Economic considerations and class size', *Economic Journal*, 113, F34–F63.

Krueger, A.N. and Lindahl, M. (2001) 'Education for growth: why and for whom?', *Journal of Economic Literature*, 39, 1101–1136.

Krueger, A.N. and Meyer, B.D. (2002) 'Labour Supply Effects of Social Insurance', in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Vol. 4, Elsevier Science, Amsterdam, pp. 2327-2391.

Laffer, A.B. (1979) 'Statement Prepared for the Joint Economic Committee, May 20', reprinted in 1979 in A.B. Laffer and J.P. Seymour (eds) *The Economics of the Tax Revolt: A Reader*, Harcourt, Brace, Jovanovich, New York.

Lane, J-E. (2000) *New Public Management*, Routledge, London.

Lange, O. (1942) 'The foundations of welfare economics', *Econometrica*, 10, 215–228.

Lansdell, N. and Gangadharan, L. (2003) 'Comparing travel cost models and the precision of their consumer surplus estimates: Albert Park and Maroondah Reservoir', *Australian Economic Papers*, 42, 399–417.

Lau, S-H.P. and Sin, C-Y. (1997) 'Public infrastructure and economic growth: time series properties and evidence', *Economic Record*, 73, 125–135.

Laubach, T. (2005) *Fiscal Relations Across levels of Government in the United States*, Working Paper 462, Economics Department OECD, Paris.

Lavy, V. (2004) *Performance Pay and Teachers' Effort, Productivity and Grading Ethics*, Working Paper 10622, National Bureau of Economic Research, New York.

Layard, R. (2005) *Happiness: Lessons from a New Science*, Penguin, London.

Layard, R. (2008) 'Happiness and public policy: a challenge to the profession', *Economic Journal*, 116, C24–C33.

Layard, R., Nickell, S. and Mayraz, G. (2008) 'The marginal utility of income', *Journal of Public Economics*, *Special Issue, Happiness and Public Economics*, (eds) T. Besley and E. Saez, 92, 1846–1857.

Le, A.T. and Miller, P.W. (2000) 'Australia's unemployment problem', *Economic Record*, 76, 74–104.

Lee, Y.L. (2000) 'Optimal schooling investments and earnings: an analysis using Australian twins data', *Economic Record*, 76, 225–235.

Leigh, A. (2008) 'Returns to education in Australia', *Economic Papers*, 27, 233–49.

Leigh, A. (2010a) 'Estimating teacher effectiveness from two-year changes in students' test scores', *Economics of Education Review*, 29, 480–488.

Leigh, A. (2010b) 'Who benefits from the earned income tax credit? Incidence among recipients, coworkers and firms, *B.E. Journal of Economic Analysis*, Berkeley Electronic Press, 10, Article 45.

Leigh, A. and Ryan, C. (2008a) *Estimating Returns to Education: Three Natural Techniques Compared*, *Economics of Education Review*, 27, 149–160.

Leigh, A. and Ryan, C. (2008b) 'How and why has teacher quality changed in Australia?' *Australian Economic Review*, 41, 141–159.

Lerner, A.P. (1948) 'The burden of the national debt' in L.A. Metzler *et al.* (eds) *Income Employment and Public Policy: Essays in Honour of Alvin H. Hansen*, W.W. Norton, New York.

Levitt, S.D, and List, J.A. (2008*) Field Experiments in Economics: the Past, the Present, and the Future*, Working Paper 14356, National Bureau of Economic Research, Cambridge, MA.

Lewis, P. (2006) *Minimum Wages and Employment: A Report for the Australian Fair Pay Commission*, Centre for Labour Market Research, University of Canberra, Canberra.

Li, M.M. and Brown, H.J. (1980) 'Micro-neighbourhood externalities and hedonic housing prices', *Land Economics*, 56, 125–141.

Liebenstein, H. (1966) 'Allocation efficiency and inefficiency', *American Economic Association*, 56, 392–415.

Lindahl, E. (1919) 'Just taxation—a positive solution', reprinted in 1958 in R.A. Musgrave and A.T. Peacock, (eds), *Classics in the Theory of Public Finance*', St Martin's Press, New York.

Lindert, P. and Williamson, J. (2001) *Does Globalisation Make the World More Unequal?,* Working Paper No. 8228, National Bureau of Economic Research, Cambridge, MA.

Lipsey, R. and Lancaster, K. (1957) 'The general theorem of second best', *Review of Economic Studies*, 24, 11–32.

Little, I.M.D. (1957) *A Critique of Welfare Economics'*, 2nd ed., Oxford University Press, Oxford.

Lochner, L. (2011) *Non-Production Benefits of Education: Crime, Health and Good Citizenship*, Working Paper 16722, National Bureau of Economic Research, Cambridge, MA.

Locke, J. (1690) *The Second Treatise of Government*, reproduced in 1946 by J.W. Gough, Basil Blackwell, Oxford.

Lopez, R.A. and Pagoulatos, E. (1994) 'Rent seeking and the welfare cost of trade barriers', *Public Choice*, 79, 149–160.

Louviere, J.J. and Fiebig, D. (2010) 'Benefit assessment for cost-benefit analysis studies in health care using discrete choice experiments: estimating welfare in a health care setting', in E. McIntosh, P.M. Clarke, E.J. Frew and J.J. Louviere (eds) *Applied Methods of Cost-Benefit Analysis in Health Care*, Oxford University Press, New York, pp. 211–229.

Lucas, R.E. and Sargent, T.J. (1981) *Rational Expectations and Econometric Practice*, University of Minnesota Press, Minneapolis, MN.

Lundsgaard, J. (2002) *Competition and Efficiency in Publicly Funded Services*, Economics Department Working Papers, No, 331, OECD, Paris.

Lybeck, J.A. (1986) *The Growth of Government in Developed Economies*, Gower, Hants, UK.

Mailath, G.J. and Samuelson, L. (2006) *Repeated Games and Reputations: Long-run Relationships*, Oxford University Press, Oxford.

MacFarlan, M. and Oxley, H. (1996) 'Social transfers: spending patterns, institutional arrangements and policy responses', *OECD Economic Studies*, No. 27, 147–193.

Mankiw, N.G. (1995) 'The growth of nations', *Brookings Papers on Economic Activity*, 1, 275–1326.

Mankiw, N.G., Weinzierl, M. and Yagan, D. (2009) 'Optimal taxation in theory and practice', *Journal of Economic Perspectives*, 23, 147–1174.

Markle, K.S. and Shakleford, D.A. (2011) *Cross-Country Comparisons of Corporate Income Tax Rates*, Working Paper 16839, National Bureau of Economic Research, Cambridge, MA.

Marlow, M. (1997) 'Public education supply and student performance', *Applied Economics*, 29, 617–26.

Martin, M.C. (2011) 'Welfare reform in the U.S.: a policy overview analysis', *Poverty and Public Policy*, 3, 1–38.

Martin, R. (1999) 'Fiscal Federalism in Canada: Lessons for Australia', mimeo, NSW Treasury, Sydney.

Martin, J.P. and Grubb, D. (2001) *What works and for whom: a review of OECD countries' experiences with active labour market policies*, Working Paper 2001:14, OECD, Paris.

Marx, K. (1867) *Das Kapital*, Otto Meissner, Hamburg.

Marx, K. (1875) *Critique of the Gotha Program (based on a letter to the German Social Democratic Movement)*.

Mathers, C., Vos, T. and Stevenson, C. (1999) *The Burden of Disease and Injury in Australia*, Australian Institute of Health and Welfare, AIHW, Cat. No. PHE 17, Canberra.

Mayer, C.P. and Meadowcroft, S.A. (1985) 'Selling public assets: techniques and financial implications', *Fiscal Studies*, 42–56.

McClure, C.E., Jr. (1971) 'The theory of tax incidence with imperfect factor mobility', *Finanzarchiv*, 30.

McConnell, M. (1987) 'Federalism: evaluating the Founders' design', *University of Chicago Law Review*, 54, 1484–1512.

McConnell, C.R., Bruce, S.L. and Macpherson, D.A. (1999) *Contemporary Labour Economics*, McGraw-Hill, New York.

McNair, B.J. and Abelson, P. (2010), 'Estimating the value of under grounding electricity and communications networks', *Australian Economic Review*, 43, 376–388.

McNair, B.J., Bennett, J., Hensher, D.A. and Rose, J.M. (2011) 'Households' willingness to pay for overhead-to-underground conversion of electricity distribution networks', *Energy Policy*, 39, 2560-2567.

Meade, J.E. (Chair) (1978) *The Structure and Reform of Direct Taxation*, Report of a Committee to the Institute of Fiscal Studies, Allen & Unwin, London.

Megginson, W. and Netter, J. (2001) 'From state to market: a survey of empirical studies of privatisation', *Journal of Economic Literature*, 39, 321–89.

Meyer (1995) 'Lessons from the U.S. unemployment insurance experiments', *Journal of Economic Literature*, 33, 91–131.

Meyer, B.D. and Mok, W. (2007) *Quasi-Experimental Evidence on the Effects of Unemployment Insurance from New York State*, Working Paper No. w12865, National Bureau of Economic Research, New York.

Mikesell, J.L. (1998) 'The link between tax law and tax collections', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York, pp. 173–198.

Mill, J.S. (1848) *Principles of Political Economy*, edited by W.J. Ashley in 1909, Longman Green, London.

Mill, J.S. (1859) *On Liberty,* Parker and Son, London (available on Google).

Mincer, J. (1974) *Schooling, Experience and Earnings*, National Bureau of Economic Research, New York.

Mirrlees, J.A. (1971) 'An exploration in the theory of optimum income taxation', *Review of Economic Studies*, 38, 175–208.

Mirrlees, J. (Chair) (2011) *Tax by Design, The Mirrlees Review*, Institute of Fiscal Studies, London.

Mishan, E. (1988) *Cost–Benefit Analysis*, Allen and Unwin, London.

Mitchell, D., Harding, A. and Gruen, F. (1994) 'Targeting welfare', *Economic Record*, 70, 315–340.

Mitchell, R.C. and Carson, R.T. (1989) *Using Surveys to Value Public Goods*, John Hopkins University Press, Baltimore, MD.

Mixon, F. (1995) 'Public choice and the EPA: empirical evidence on carbon emissions violations', *Public Choice*, 83, 127–137.

Moffit, R. (1992) 'Incentive effects of the U.S. welfare system: a review', *Journal of Economic Literature*, 30, 1–61.

Montesquieu, C. (1748) *The Spirit of the Laws*, edited by A.M. Cohler, B.C. Miller and H.S. Stone in 1989, Cambridge University Press, Cambridge.

Morgan, W., Mutti, J. and Partridge, M. (1989) 'A regional general equilibrium model of the United States: tax effects on factor movements and regional production', *Review of Economics and Statistics*, 71, 626–635.

Morrison, M. and Bennett, J. (2004) 'Valuing New South Wales rivers for use in benefit transfer', *Australian Journal of Agricultural and Resource Economics*, 48, 591–612.

Mueller, D.C. (2003) *Public Choice III*, Cambridge University Press, Cambridge.

Mueller, D.C. and Murrell, P. (1985) 'Interest groups and the political economy of government size', in F. Forte and A. Peacock (eds) *Public Expenditure and Government Growth*, Basil Blackwell, Oxford, pp. 13–36.

Mueller, D.C. and Murrell, P. (1986) 'Interest groups and the size of government', *Public Choice*, 13, 55–79.

Musgrave, R.A. (1939) 'The voluntary exchange theory of public economy', *Quarterly Journal of Economics*, 53, 213–237.

Musgrave, R.A. (1959) *The Theory of Public Finance*, McGraw–Hill, New York.

Musgrave, R.A and Musgrave, P.B. (1989) *Public Finance in Theory and Practice*, 5th ed., McGraw-Hill, New York.

Musso, J.A. (1998) 'Fiscal Federalism as a Framework for Government Reform', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York, pp. 347–395.

National Centre for Vocational Education Research (2001) *Australian Apprenticeships: Facts, Fiction and Future*, NCVER, Canberra.

National Health and Hospitals Reform Commission (2009) *A Healthier Future for All Australians: Final Report,* Australian Department of Health, Canberra.

National Institute of Labour Studies (2006) *Labour Market and Institutional Developments: Australia*, Global Policy Network, Flinders University, Adelaide.

Nelson, M. (1997) 'Municipal government approaches to service delivery: an analysis from a transactions cost perspective', *Economic Inquiry*, 35, 82–96.

Neumark, D. and Wascher, W. (2006) *Minimum Wages and Employment: A Review of Evidence from the New Minimum Wage Research*, Working Paper 12663, National Bureau of Economic Research, Cambridge, MA.

Neumark, D. and Wascher, W. (2007) *Minimum Wages, The Earned Income Tax Credit and Employment: Evidence from the Post-Welfare Reform Era,* Working Paper 12915, National Bureau of Economic Research, Cambridge, MA.

Nickell, S.J. (1996) 'Competition and corporate performance', *Journal of Political Economy*, 104, 724–746.

Nickell, S.J. (1998) 'Unemployment: questions and some answers', *Economic Journal*, 108, 802–816.

Niskanen, W.A. (1971) *Bureaucracy and Representative Government*, Aldine, Chicago, IL.

Niskanen, W.A. (1994) *Bureaucracy and Public Economics*, Edward Elgar, Brookfield, VT.

Nordhaus, W.J. (1991) 'To slow or not to slow: the economics of the greenhouse effect', *Economic Journal*, 101, 920–937.

Nordhaus, W.J. (2007) 'The Stern Review on the economics of climate change', *Journal of Economic Literature*, 45, 686–702.

Northwood, K., Rawnsley, T. and Chen, T. (2002) 'Experimental estimates of the distribution of household wealth in Australia, 1994–2000', 31st Annual Conference of Economists, Adelaide.

Nozick, R. (1974) *Anarchy, State and Utopia*, Basil Blackwell, Oxford.

NSW Department of Information Technology and Management (DITM) (2000) *The Surveyors Act 1929 and Related Regulations, Issues Paper*, National Competition Principles Review, DITM, Sydney.

NSW Independent Pricing and Regulatory Tribunal (2015) *Socially Optimal Consumption Prices*, IPART, Sydney.

NSW Independent Pricing and Regulatory Tribunal (2018) *Review of out WACC Method*, IPART, Sydney.

NSW Roads and Traffic Authority (2006) *Economic Analysis Manual*, RTA, Sydney.

NSW Treasury (2001) *Guide to Economic Performance Measurement for General Government Sector Agencies*, TPP 01-03, NSW Treasury, Sydney.

Oates, W.E. (1969) 'The effects of property taxes and local spending on property values: an empirical study of tax capitalization and the Tiebout hypothesis', *Journal of Political Economy*, 77, 957–71.

Oates, W.E. (1989) 'Searching for Leviathan: a reply and some further reflections', *American Economic Review*, 79, 579–83.

Oates, W.E. (1999) 'An essay on fiscal federalism', *Journal of Economic Literature*, 37, 1120–1149.

Okun, A.M. (1975) *Equality and Efficiency*, Brookings Institution, Washington, DC.

Olsen, M. (1982) *The Rise and Decline of Nations: Economic Growth, Stagflation and Social Rigidities*, Yale University Press, New Haven. NJ.

Organisation for Economic Cooperation and Development (OECD) (1982) *The OECD List of Social Indicators*, OECD, Paris.

OECD (1993) 'Some measures to improve the quality of government spending', *Economic Outlook*, No. 54, OECD, Paris.

OECD (1997) 'Trade earnings and employment, assessing the impact of trade with emerging countries on OECD labour markets', *Employment Outlook*, OECD, Paris.

OECD (1998) *The OECD Report on Regulatory Reform*, *Synthesis*, OECD, Paris.

OECD (2000) *From Initial Education to Working Life: Making Transitions Work*, OECD, Paris.

OECD (2001a) *The Well-being of Nations: The Role of Human and Social Capital*, OECD, Paris.

OECD (2001b) *Innovations in Labour Market Policies: The Australian Way*, OECD, Paris.

OECD (2001c) *Taxing Wages*, OECD, Paris.

OECD (2002) *OECD Employment Outlook*, OECD, Paris.

OECD (2003) *The Impact of Ageing Populations on Public Finances: Overview of Analysis Carried out at EU level and Proposals for a Future Work Programme*, OECD, Paris.

OECD (2007) *Health Data 2007*: *Statistics and Indicators for 30 Countries*, OECD, Paris.

OECD (2008) *Growing Unequal?*: *Income Distribution and Poverty in OECD Countries*, OECD, Paris.

OECD (2010) *Education at a Glance*, OECD, Paris.

OECD (2011) *Divided We Stand*: *Why Inequality Keeps Rising*, OECD, Paris.

OECD (2016) 'Using the Fiscal Levers to Escape the Low-Growth Trap', Chapter 2 in *Economic Outlook 2016,* OECD, Paris.

OECD (2017) *Education at a Glance*, OECD, Paris.

Osborne, D. and Gaebler, T. (1993) *Reinventing Government*, Penguin, New York.

Oxley, H. and Martin, J.P. (1991) 'Controlling government spending and deficits: trends in the 1980s and prospects for the 1990s', *OECD Economic Studies*, No.17, Autumn, OECD, Paris.

Pareto, W. (1909) *Manuel d'Economie Politique*, Girard & Briere, Paris.

Parker, D. (1990) 'The 1988 Local Government Act and compulsory competitive tendering', *Urban Studies*, 27, 653–668.

Pattugalan, K. and Ellis, D. (2010) 'Australia's marginal tax rates, tax offsets and the Medicare levy', *Economic Round-Up*, 4, 1–15.

Peacock, A.T. and Wiseman, J. (1961) *The Growth of Public Expenditure in the United Kingdom*, Princeton University Press, Princeton, N.J.

Pearce, D.W. (1983) *Cost-Benefit Analysis*, Macmillan, London.

Pearce, D. and Markandya, A. (1989) *Environmental Policy Benefits, Monetary Valuation*, OECD, Paris.

Peltzman, S. (1980) 'The Growth of Government', *Journal of Law and Economics*, 23, 209–288.

Persson, T. and Tabellini, G. (2002) 'Political economy and public finance', in A.J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics*, Vol. 3, Elsevier Science, Amsterdam, pp. 1551–1658.

Piggott, J. and Chapman, B. (1995) 'Costing the job compact', *Economic Record*, 71, 313–328.

Pigou, A.C. (1920) *Economics of Welfare*, 4th ed., (1938), Macmillan, London.

Pincus, J.J. (1983) 'Regulation and public enterprise' in *State Enterprise and Deregulation*, Special Study No. 5, Centre of Policy Studies, Monash University, pp. 33–42.

Pindyck, R.S. and Rubinfield, D.L. (1989) *Microeconomics*, Macmillan, New York.

Plotnick, R. (1989) 'An interest group model of direct income redistribution', *Review of Economics and Statistics*, 68, 581–584.

Pommerehne, W.W. and Schneider, F. (1982) 'Unbalanced growth between public and private sectors: an empirical examination', in R.H. Haveman (ed) *Public Finance and Public Employment*, Wayne State University Press, Detroit, MI, pp. 309–326.

Pope, J. (1997) 'Policy implications of research on compliance costs of taxation', in J.G. Head and R. Krever (eds) *Taxation Towards 2000*, Australian Tax Research Foundation, Sydney.

Pope, J. (2005) 'Reform of the personal income tax system in Australia', *Economic Papers*, 24, 316–331.

Porter, M.E. (1990) *The Competitive Advantage of Nations*, Macmillan, London.

Portnoy, F. (1997) *F.I.A.S.C.O. Blood in the Water in Wall Street*, W.W. Norton, New York.

Posner, R.A. (1975) 'The social costs of monopoly and regulation', *Journal of Political Economy*, 83, 807–27.

Prebisch, R. (1970) *Change and Development*: *Latin America's Great Task*, Praeger Publishers, Westpoint, CT.

Price Waterhouse (1989) *Privatisation: Learning the Lessons from the U.K. Experience*, Price Waterhouse, London.

Productivity Commission (1999) *Micro Reforms and Australian Productivity*, AGPS, Canberra.

Productivity Commission (2001a) *Review of the Prices Surveillance Act 1983*, Report No. 14, Ausinfo, Canberra.

Productivity Commission (2001b) *Cost Recovery by Government Agencies*, Report No. 15, Ausinfo, Canberra.

Productivity Commission (2002) *Independent Review of Job Network,* Report No. 21, Ausinfo, Canberra.

Productivity Commission (2005) *Review of National Competition Policy Reforms,* Report No. 33, Canberra.

Productivity Commission (2007) *Public Support for Science and Innovation*, Research Report 03/07, Canberra.

Productivity Commission (2010) *Gambling*, Report No. 50, Canberra.

Productivity Commission (2011a) *Trade and Assistance Review 2009–10,* Annual Report Series, Canberra.

Productivity Commission (2011b) *Report on Government Services*, Canberra.

Productivity Commission (2016), *Increasing Australia's Future Prosperity*, Discussion Paper, Canberra.

Productivity Commission (2017), *Shifting the Dial, 5 Year Productivity Review*, Report No. 84, Canberra.

Psacharopoulos, G. and Patrinos, H.A (2004) 'Returns to investment in education: a further update', *Education Economics*, 12, 111–134.

Putnam, R. (2000) *Bowling Alone*, Simon & Schuster, New York.

Quiggin, J. (1996) *Great Expectations*, Allen & Unwin, Sydney.

Quiggin, J. (1999) 'Human capital theory and education policy in Australia', *Australian Economic Review*, 32, 130–144.

Quiggin, J. (2010) 'Privatisation' pp. 174–205 in *Zombie Economics*, Princeton University Press, Princeton, N.J.

Quinet, E. and Vickerman, R.W. (2004) *Principles of Transport Economics*, Edward Elgar, Cheltenham, UK.

Quinn, D. (1997) 'The correlates of change in international financial regulation,' *American Political Science Review,* 91, 531–551.

Rainey, H.G. and Chun, Y.H. (2005) 'Public and private management compared', in E. Ferlie, L. Lynn Jr. and C. Pollitt (eds), *Oxford Handbook of Public Management*, Oxford University Press, Oxford, pp. 72–102.

Rammohan, A. and Whelan, S. (2007) 'The impact of childcare costs on the full-time/part-time employment decisions of Australian mothers', *Australian Economic Papers*, 46, 152–168

Ramsey, F. (1927) 'A contribution to the theory of taxation', *Economic Journal*, 37, 47–61.

Rawls, J. (1971) *A Theory of Justice*, Harvard University Press, Cambridge, MA.

Rees, R. (1984) *Public Enterprise Economics,* 2nd ed., Weidenfield and Nicolson, London.

Reid, B.G. (1998) 'Endogenous elections electoral budget cycles and Canadian provincial governments', *Public Choice*, 97, 35–48.

Reinhart, C.M. and Rogoff, K.S. (2010a) *From Financial Crash to Debt Crisis*, Working Paper 15795, National Bureau of Economic Research, New York.

Reinhart, C.M. and Rogoff, K.S. (2010b) 'Growth in a time of debt', *American Economic Review*, 100, 573–578.

Resource Assessment Commission (1991) *Commentaries on the Resource Assessment Commission's Contingent Valuation Survey of the Kakadu Conservation Zone*, Resource Assessment Commission, Canberra.

Richardson, S. and Harding, A. (1999) 'Poor workers?', in Richardson (ed.) *Reshaping the Labour Market*, Cambridge University Press, Melbourne.

Robbins, L. (1938) 'Interpersonal comparisons of utility: a comment', *Economic Journal*, 48, 635–41.

Robinson, M. (2002) *Best Practice in Performance Budgeting,* Discussion Paper No. 124, Queensland University of Technology, Brisbane.

Rodgers, J.R., Siminski, P. and Bishop, J. (2009) 'Changes in poverty rates during the Howard era', *Australian Economic Review*, 42, 300–320.

Rodrik, D. (1997) *Trade, Social Insurance, and the Limits to Globalisation*, Working Paper No. 5905, National Bureau of Economic Research, Cambridge, MA.

Rodrik, D. (1998) 'Why do more open economies have bigger governments?', *Journal of Political Economy*, 106, 997–1032.

Rodrik, D. (2015) *Economics Rules: The Rights and Wrongs of the Dismal Science*, W.W. Norton, New York.

Roed, K. and T. Zhang, 2003, 'Does unemployment compensation affect unemployment duration?', *Economic Journal*, 113, 190–206.

Rolfe, J. and Dyack, A. (2011) 'Valuing recreation in the Coorong, Australia, with travel costs and contingent valuation methods', *Economic Record*, 87, 282–293.

Romer, P.M., 1996 'Why, indeed, in America? Theory, history and the origins of modern economic growth', *American Economic Association, Papers and Proceedings*, 86, 202–206.

Rose-Ackerman, S. (1999) *Corruption and Government; Causes, Consequences and Reform*, Cambridge University Press, Cambridge.

Rosen, H.S. and Gayer, T. (2010) *Public Finance*, 9th ed., McGraw-Hill, New York.

Rouse, C. (1998) 'Private school vouchers and student achievement: an evaluation of the Milwaukee parental choice program', *Quarterly Journal of Economics*, 113, 553–602.

Rousseau, J-J. (1762) *The Social Contract*, reprinted in 1966 by Hafner, New York.

Rowntree, B.S. (1901) *Poverty: A Study of Town Life*. Longmans, London.

Rowley, C.K., Shughart II, W.F. and Tollinson, R.D. (2002) 'Introduction', in C.K. Rowley, W.F. Shughart II. and R.D. Tollinson (eds) *The Economics of Budget Deficits*, Volume 1, Edward Elgar, Cheltenham, UK, pp. xiii–xlviii.

Rubinfield, D.L. (1987) 'The economics of the local public sector', in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Volume II, Elsevier, Amsterdam, pp. 571–639.

Sachs, J.D. (1999) 'Managing global capitalism', *Australian Economic Review*, 32, 3–16.

Sachs, J.D. and Warner, A. (1995) 'Economic reform and the process of global integration', *Brookings Papers on Economic Activity*, 1, 1–118.

Sachs, J.D. and Warner, A. (1997) 'Fundamental sources of long-run growth', *American Economic Association, Papers and Proceedings*, 97, 184–188.

Saez, E. (2001) 'Using elasticities to derive optimal income tax rates', *Review of Economic Studies*, 68, 205–229.

Sala-I-Martin, X.X. (1997) 'I just ran two million regressions', *American Economic Association, Papers and Proceedings*, 97, 178–183.

Samuelson, P.A. (1948) 'International trade and the equalisation of factor prices', *Economic Journal*, 36, 357–359.

Samuelson, R.G. (1997) *The Good Life and its Discontents*: *the American Dream in the Age of Entitlement 1945–95*, Random House, New York

Sandford, C. (1995) *Tax Compliance Costs, Measurement and Policy*, Fiscal Publications, Bath. UK.

Sandler, T. and Tschirhart, J. (1980) 'The economic theory of clubs: an evaluative survey', *Journal of Economic Literature*, 18, 1481–1521.

Saunders, C. (1997) *The Australian Constitution*, Constitutional Centenary Foundation, Melbourne.

Saunders, P. (1999) 'Budget standards and the poverty line', *Australian Economic Review*, 32, 43–61.

Saunders, P. and Bradbury, B. (2006) 'Monitoring trends in poverty and income distribution: data, methodology and measurement', *Economic Record*, 82, 341–364.

Saunders, P. and Klau, F. (1985) 'Public expenditure and economic performance in OECD countries', *Journal of Public Policy*, 5, 10–21.

Saunders, P. and Naidoo, Y. (2009) 'Poverty, deprivation and consistent poverty', *Economic Record*, 85, 417–432.

Schanz, Georg von (1896) 'Der Einkommensbegriff und die Einkommensteuergesetze', *Finanzarchiv,* 13, 1–87.

Scherer, F.M. and Ross, D. (1990) *Industrial Market Structure and Economic Performance*, 3rd ed., Houghton Miffin, Boston, MA.

Schick, A. (1998) *A Contemporary Approach to Public Expenditure Management*, World Bank Institute, Washington, DC.

Schulze, G.G. and Ursprung, H.W. (1999) 'Globalisation of the economy and the nation state', in D. Greenway, and J. Whalley (eds) *The World Economy,* Blackwell Publishers, Oxford.

Schumpeter, J.A. (1943) *Capitalism, Socialism and Democracy*, Unwin, London.

Scitovsky, T. (1941) 'A note on welfare propositions in economics', *Review of Economic Studies*, 8, 77–88.

Scutella, R. (2004) *Moves to a basic incomes-flat system in Australia: implications for the distribution of income and supply of labour*, Melbourne Institute of Applied Economic and Social Research, Working Paper No, 5/04, University of Melbourne.

Sen, A. (1985) 'Social choice and justice', *Journal of Economic Literature*, 23, 1764–1776.

Sen, A. (1997) *On Economic Inequality*, Clarendon Press, Oxford.

Shapiro. C. and Varian, H.R. (1999) *Information Rules*, Harvard University Business Press, Cambridge, MA.

Shirley, M. and Walsh, P. (2000) *Public versus Private Ownership: the Current State of the Debate,* Working Paper, World Bank, Washington, DC.

Simons, H. (1938) *Personal Income Taxation*, University of Chicago Press, Chicago, IL.

Sinclair Knight Merz (2001) *Lane Cove Tunnel and Associated Road Improvements*, Working Paper No. 14, Economics, NSW Roads and Traffic Authority, Sydney.

Sinn, H-W. (2007) *The Welfare State and the Forces of Globalisation*, Working Paper 12946, National Bureau of Economic Research, Cambridge, MA.

Slemrod, J. and Yitzhaki, S. (2002) 'Tax avoidance, evasion and administration', in A.J. Auerbach and M. Feldstein (eds) *Handbook of Public Economics*, Vol. 3, Elsevier Science, Amsterdam, pp. 1423–1470.

Slemrod, J., Yitzhaki, S., Mayshar, J. and Lundholm, M. (1994) 'The optimal two-bracket linear income tax', *Journal of Public Economics*, 53, 269-290.

Smith, A. (1776) *The Wealth of Nations*, edited by E. Cannan in 1937, Random House, New York.

Smith, V.K. (1993) 'Non-market valuation of environmental resources: an interpretive appraisal, *Land Economics*, 69, 1–26.

Staaf, R. (1977) 'The growth of the educational bureaucracy: Do teachers make a difference?', in T. Borcherding (ed.) *Budgets and Bureaucrats: The Sources of Government Growth*, Duke University Press, Durham, NC, pp. 146–168.

Stanley, J. Hensher, D.A., Stanley, Janet, Currie, G., Greene, W.H. and Vella-Brodrick, D. (2011) 'Social exclusion and the value of mobility', *Journal of Transport Economics and Policy*, 45, 197–222.

Steinwall, R. (1999) *Trade Practices Act 1974*, Butterworths Annotated Acts, Butterworths, Sydney.

Stern, N.H. (1976) 'On the specification of models of optimum income taxation', *Journal of Public Economics*, 6, 123–162.

Stern, N.H. (1987) 'The theory of optimal commodity and income taxation: an introduction', in D. Newbury and N. Stern (eds) *The Theory of Taxation for Developing Countries*, Oxford University Press, pp. 25–59.

Stigler, G.J. (1955) 'Introduction' in National Bureau of Economic Research, *Business Concentration and Price Policy*, Princeton University Press, Princeton, NJ.

Stigler, G. (1975) *The Citizen and the State*, University of Chicago Press, Chicago, IL.

Stiglitz, J.E. (1996) 'The Role of Government in Economic Development', *Annual World Bank Conference on Development Economics*, pp. 11–23, World Bank, Washington, DC.

Stiglitz, J.E. (2000) *Economics of the Public Sector*, W.W. Norton, New York.

Stiglitz, J.E. (2002) *Globalisation and its Discontents*, Penguin Books, New York.

Stiglitz, J.E., Sen, A. and Fitoussi, J-P, (2009) *Report by the Commission on the Measurement of Economic Performance and Social Progress*, OECD, Paris.

Stromback, M. and Dockery, A. (2000) *Labour Market Programs, Unemployment and Employment Hazards: An Application Using the 1994-97 Survey of Employment and Unemployment Patterns,* ABS Occasional Paper, Catalogue No. 6293.0.00.002, Australian Bureau of Statistics, Canberra.

Sutherland, D., Price, R. and Joumard, I. (2005) *Fiscal Rules for Sub-Central Governments: Design and Impact*, Economics Department Working Paper, No. 465, OECD, Paris.

Swank, D. (1997) 'Global markets, democratic institutions, and the public economy in advanced industrial societies', mimeo, Marquette University. Milwaukee, MI.

Tanzi, V. and Schuknecht, L. (1995) *The Growth of Government and the Reform of the State in Industrial Countries*, IMF Working Paper, presented at the Conference on Inequality, the Welfare State and Social Values in Madrid, July 1995, Fiscal affairs Department, IMS, Washington, DC.

Taylor, J.B., Moosa, I. and Cowling, B. (2000) *Microeconomics*, John Wiley, Brisbane.

Temple, J. (1999) 'The New Growth Evidence', *Journal of Economic Literature*, 37, 112–56.

Thurow, L. (1996) *The Future of Capitalism*, Allen & Unwin, London.

Thurow, L. (1999) 'The third industrial revolution', excerpt from *Atlantic Monthly*, reprinted in Australian Financial Review, 2 July.

Tiebout, C. (1956) 'A pure theory of local expenditures', *Journal of Political Economy*, 64, 416–424.

Tietenberg, T. (2006) *Environmental and Natural Resource Economics*, 6th ed, Harper Collins, New York.

Tobin, J. (1970) 'On limiting the domain of inequality', *Journal of Law and Economics*, 13, 263–277.

Tolley, G., Kenkel, D. and Fabian, R. (1994) 'State of the art health values', in G. Tolley, D. Kenkel and R. Fabian, (eds) *Valuing Health for Public Policy, An Economic Approach*, University of Chicago Press, Chicago, pp. 323-344.

Treasurer (2010) *Australia to 2050*: *Future Challenges*, Treasury, Canberra.

Treasury (2002) *Intergenerational Report*, *2002–03*, Budget Paper No.5, AGPS, Canberra.

Treasury, Chairs: R.F.E Warburton and P.W. Hendy (2006) *International Comparison of Australia's Taxes*, Treasury, Canberra.

Treasury (2007) *Intergenerational Report IGR2*, Treasury, Canberra.

Treasury (Annual) *Tax Expenditures Statement*, Treasury, Canberra.

Tretz, G., Rickman, D.S., Hunt, G.L. and Greenwood, M.J. (1993) 'The dynamics of U.S. internal migration', *Review of Economics and Statistics*, 75, 209–214.

Tsuge, T., Ksishmoto, A. and Takeuchi, K. (2005) 'A choice experiment approach to the valuation of mortality', *Journal of Risk and Uncertainty*, 31, 73–95.

Tunny, G. (2000) 'An empirical evaluation of the Baumol cost disease', 29th Conference of Economic Society of Australia, Surfers Paradise, Queensland.

Tuomela, M. (1990) *Optimal Income Tax and Redistribution*, Clarendon Press, Oxford.

Twomey, A. and Withers, G. (2007) *Australia's Federal Future: Delivering Growth and Prosperity*, Report for the Council for the Australian Federation.

Tyers, R., Duncan, R. and Martin, W. (1997) 'Trade, technologly and labour markets: general equilibrium perspectives', 26th Annual Conference of Economists, Hobart.

UK Department of Health (1994) *The Importance of Price in Reducing Tobacco Consumption*, DoH, London.

UK Department for Transport (2009) *The Wider Impacts Sub-Objective*, TAG Unit 3.5.14, DfT, London.

UK Department for Transport (2010) *Values of Time and Operating Costs*, TAG Unit 3.5.6, DfT, London.

UK Department for Transport (2016a), *Wider Economic Benefits Appraisal*, TAG Unit A2.1.

UK Department for Transport (2016b) *Transport Appraisal in the Context of Dependent Development*, TAG Unit A2.3.

UK Department for Transport, 2016c, *Appraisal of Productivity Impacts*, TAG Unit A2.4.

UK Treasury (2002) *Spending Review, Public Service Agreement*, Treasury, London.

Van den Berg, G. and Van der Klaauw, B. (2001) *Counselling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment*, Free University of Amsterdam, (www.tinbergen.nl/~klaauw/CM.pdf)

Van Reenan, J. (2001) *No More Skivvy Schemes? Active labour Market Policies and the British Deal for Young Unemployed in Context*, Institute of Fiscal Studies, Working Paper 01/09, IFS, London.

Varian, H.R. (2006) *Intermediate Microeconomics: A Modern Approach*, 7th ed., W.W. Norton, New York.

Vickers, L. (2005) 'Abuse of market power', *Economic Journal*, 115, F244–F261.

Vickers, J. and Yarrow, G. (1991) 'Economic perspectives on privatisation', *Journal of Economic Perspectives*, 5, 111–132.

Victorian Department of Treasury and Finance, 2013, *Cost Recovery Guidelines*, Melbourne.

Villanyi, J., Hajos, L., Singh, M.H., Vasary, M. and Kumar, A. (2010) 'Ageing and public finances in the European Union: condition and future prophecies', *DSM Business Review*, 2, 91–120.

Viscusi, W.K., Vernon, J.M. and Harrington, J.E., Jr. (2005) *Economics of Regulation and Antitrust*, 4th ed., MIT Press, Cambridge, MA.

Wagstaff, A. and Van Doorslaer, E. (2000) 'Equity in health care finance', in A.J. Culyer and J.P. Newhouse (eds) *Handbook of Health Economics*, Elsevier, Amsterdam, 1804–1862.

Wales, T. (1968) 'Distilled spirits and interstate consumption effects', *American Economic Review*, 68, 853–863.

Walsh, R., Johnson, D. and McKean, J. (1992) 'Benefit transfer of outdoor recreation demand studies, 1968–1988', *Water Resources Research*, 28, No. 3.

Waters II, W.G. and Street, J. (1998) 'Monitoring the performance of government trading enterprises', *Australian Economic Review*, 31, 357–371.

Warren, N. (2004) *Tax Facts, Fiction and Reform*, Research Study No. 41, Australian Tax Research Foundation, Sydney.

Watson, W. and Jaksch, J. (1982) 'Air pollution: household soiling and consumer welfare losses', *Journal of Environmental Economics and Management*, 9, 248–262.

Weare, C. and Friedman, L.S. (1998) 'Public sector pricing: an institutional approach', in F. Thompson and M.T. Green (eds) *Handbook of Public Finance*, Marcel Dekker, New York, pp. 597–658.

Wei, H. (2010) *Measuring Economic Returns to Post-School Education in Australia: Evidence from the 1981-2006 Australian Censuses*, Research Paper, Economics Analysis and Reporting Branch, Australian Bureau of Statistics, Canberra.

Weimer, D.L. and Wolkoff, M.J. (2001) 'School performance and housing values: Using non-contiguous district and incorporation boundaries to identify school effects', *National Tax Journal*, 54, 231–254.

Whiteford, P. (2010) 'The Australian tax transfer system: architecture and outcomes', *Economic Record*, 86, 528-544.

Whiteford, P. and Angenent, G. (2002) *The Australian system of social protection: an overview*, Occasional Paper No. 6, Department of Family and Community Services, Canberra.

Wicksell, K. (1896) 'A new principle of just taxation', reprinted in 1958 in R.A. Musgrave and A.T. Peacock (eds) *Classics in the Theory of Public Finance*, St. Martins Press, New York.

Williams. R. (2005) 'Fiscal federalism: aims, instruments and outcomes', *Australian Economic Review*, 38, 351–369.

Willig, R.D. (1976) 'Consumer's surplus without apology', *American Economic Review*, 66, 587–597.

Willis, K.G. and Garrod, G. (1991) 'The recreational value of Forestry Commission estate in Great Britain, a Clawson-Knetsch travel cost analysis' *Scottish Journal of Political Economy*, 38, 58–75.

Winston, C. (1993) 'Economic deregulation: days of reckoning for microeconomists', *Journal of Economic Literature*, XXXVI, 1263-1289.

Winston, C. and Crandall, R.W. (1994) *Explaining Regulatory Policy*, Brookings Papers on Economic Activity, Microeconomics, Brookings Institution, Washington.

Wolf, C. (1988) *Markets or Government*, MIT Press, Cambridge, MA.

Wolfe B. and Haveman, R. (2001) 'Accounting for the social and non-market benefits of education' in J.F. Helliwell (ed.) *The Contribution of Human and Social Capital to Sustained Growth and Well Being*, International Symposium Report, Human Resources Development Canada and OECD, Paris.

Wood, A. (1998) 'Globalisation and the rise of labour market inequalities', *Economic Journal*, 108, 1463–1472.

World Bank (1994a) *Infrastructure for Development*, World Development Report, Oxford University Press, New York.

World Bank (1994b) *Averting the Old Age Crisis: Policies to Protect the Old and Promote Growth*, Oxford University Press.

World Bank (1996) *From Plan to Market,* World Development Report, Oxford University Press, Oxford.

World Bank (1997) *The State in a Changing World,* World Development Report, World Bank, Washington, DC.

World Bank (2000) *Attacking Poverty, World Development Report, 2000/2001*, World Bank, Washington, DC.

World Bank (2002) *Globalisation, Growth and Poverty*, Policy Research Report, World Bank, Washington, DC.

World Bank (2004), *Reforming Infrastructure: Privatisation, Regulation and Competition*, World Bank and Oxford University Press.

World Bank (2006) *World Development Indicator*, Oxford University Press, Oxford.

World Health Organisation (1948) *Constitution of the World Health Organisation*, adopted in New York and WHO, Geneva.

Wyckoff, P.G. (1995) 'Capitalisation, equalization and intergovernmental aid', *Public Finance Quarterly*, 23, 484–508.

Yang, Q.G. and Pickford, M. (2011) 'Safe harbours in merger guidelines: what should they be?', *Australian Economic Review*, 44, 13–35.

Yinger, J., Bloom, H.S., Borsch-Supan, A. and Laff, H.F. (1988) *Property Taxes and House Values: The Theory and Estimation of Intrajurisdictional Tax Capitalisation*, Press Inc., San Diego, CA.

Zabel, J., Schwartz, S. and Stephen, D. (2010) 'The impact of the Self-Sufficiency Project on the employment behaviour of former welfare recipients', *Canadian Journal of Economics*, 43, 882–918.

Zax, J.S. (1989) 'Initiatives and government expenditures', *Public Choice*, 63, December, 267–277.

Zhou, L., Bowles, T., Biswas, B. and Saunders, P. (2011) 'Impact of globalisation on income distribution inequality in 60 countries', *Global Economy Journal*, 11, 1–16.