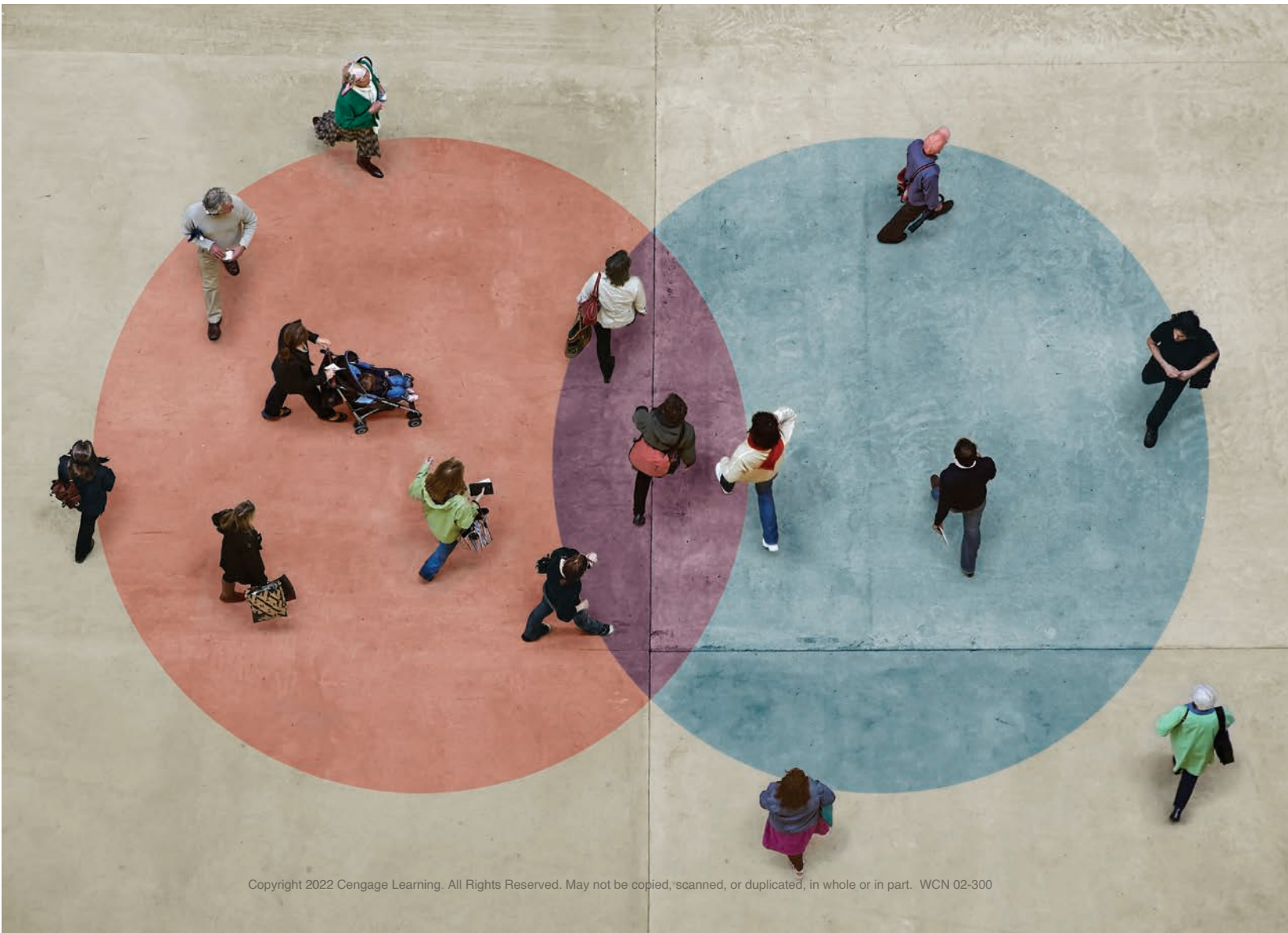
7th Asia-Pacific Edition

economics for today

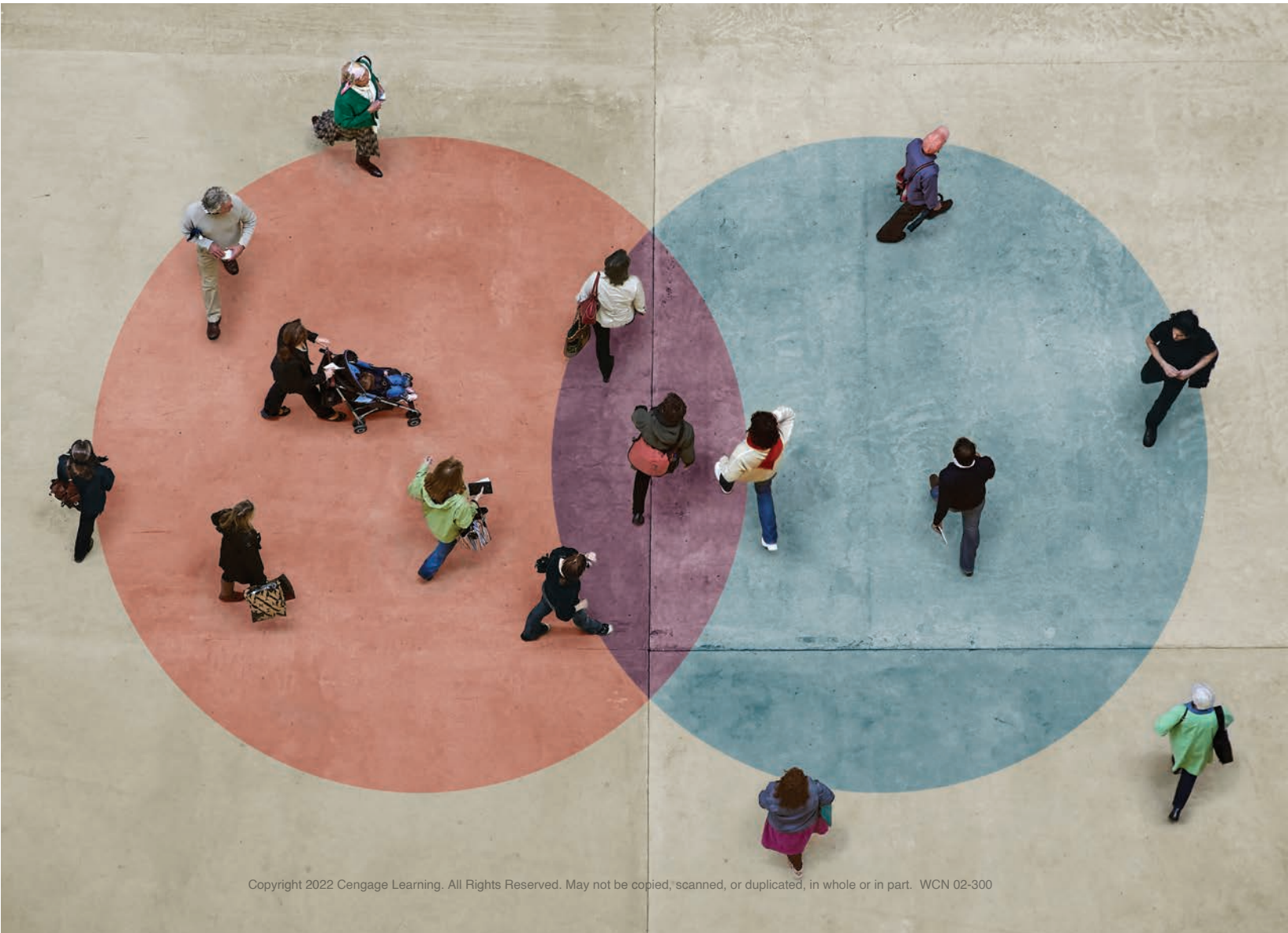
Allan **Layton** _ Tim **Robinson** _ Tommy **Tang** _ Dinusha **Dharmaratna** _ Irvin B. **Tucker**



7th Asia-Pacific Edition

economics for today

Allan **Layton** _ Tim **Robinson** _ Tommy **Tang** _ Dinusha **Dharmaratna** _ Irvin B. **Tucker**



Economics for Today
7th Edition
Alan Layton
Tim Robinson
Tommy Tang
Dinusha Dharmaratna
Irvin B. Tucker

Portfolio lead: Fiona Hammond
Product manager: Rachael Pictor
Content developer: Talia Lewis
Project editor: Alex Chambers
Cover designer: Leigh Ashforth
Text designer: Rina Gargano
Editor: Jade Jakovcic
Indexer: Julie King
Proofreader: Carolyn Leslie
Permissions/Photo researcher: Catherine Kerstjens
KnowledgeWorks Global Ltd.

Any URLs contained in this publication were checked for currency during the production process. Note, however, that the publisher cannot vouch for the ongoing currency of URLs.

Sixth edition published in 2019

Authorised adaptation of *Economics for Today* 10th edition by Tucker, published by Cengage Learning 2019 [ISBN 9781337613040].

Acknowledgements
Cover image courtesy Getty Images/David Malan.

Part-opening images courtesy [1] iStock.com/aldomurillo; [2] iStock.com/pixelfit; [3] iStock.com/Inok; [4] iStock.com/Henrik5000; [5] iStock.com/FatCamera; [6] iStock.com/Orbon Alija.

National Library of Australia Cataloguing-in-Publication Data
ISBN: 9780170449021
A catalogue record for this book is available from the National Library of Australia.

Cengage Learning Australia
Level 7, 80 Dorcas Street
South Melbourne, Victoria Australia 3205

Cengage Learning New Zealand
Unit 4B Rosedale Office Park
331 Rosedale Road, Albany, North Shore 0632, NZ

For learning solutions, visit cengage.com.au

Printed in Singapore by 1010 Printing International Limited.
1 2 3 4 5 6 7 25 24 23 22 21

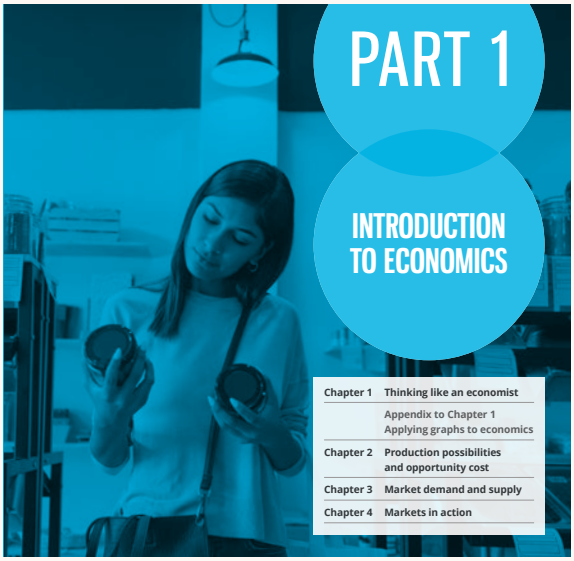
Brief contents

Contents

Guide to the text

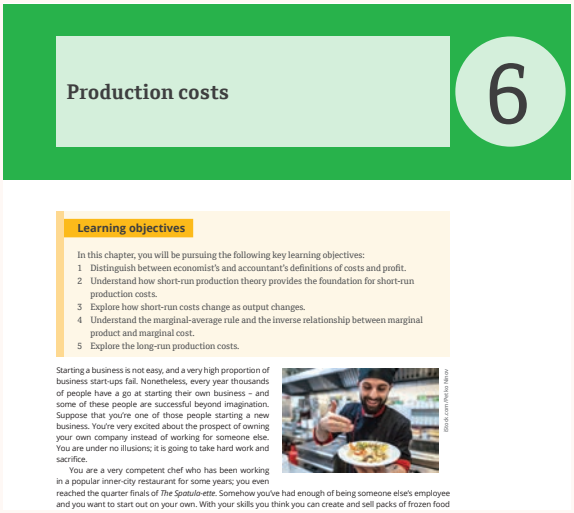
As you read this text you will find a number of features in every chapter to enhance your study of Economics and help you understand how the theory is applied in the real world.

PART OPENING FEATURES



The Chapter list outlines the chapters contained in each part for easy reference.

CHAPTER OPENING FEATURES


Production costs

6

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Distinguish between economist's and accountant's definitions of costs and profit.
2 Understand how short-run production theory provides the foundation for short-run production costs.
3 Explore how short-run costs change as output changes.
4 Understand the marginal-average rule and the inverse relationship between marginal product and marginal cost.
5 Explore the long-run production costs.

Starting a business is not easy, and a very high proportion of business start-ups fail. Nonetheless, every year thousands of people have a go at starting their own business – and some of these people are successful beyond imagination. Suppose that you're one of those people starting a new business. You're very excited about the prospect of owning your own company instead of working for someone else. You are under no illusions; it is going to take hard work and sacrifice.

You are a very competent chef who has been working in a popular inner-city restaurant for some years; you even reached the quarter finals of *The Spatula-ette*. Somehow you've had enough of being someone else's employee and you want to start out on your own. With your skills you think you can create and sell packs of frozen food

Identify the key concepts that the chapter will cover with the **Learning objectives** at the start of each chapter.

The **Chapter overview** introduces you to the key topics throughout each chapter and relates them to the real world.

FEATURES WITHIN CHAPTERS

Examine how real-world issues can be solved using the theoretical concepts you have learnt throughout each chapter in the **Analyse the issue** boxes.

Analyse the issue Applicable concepts: change in resources and sustainability

Immigration and the PPF


Getty Images/Michael Masters

Over the past decade the world has seen larger migrant and refugee flows than have been witnessed since the great post-war migrations of the late 1940s and 1950s. Whether it is North African and Middle Eastern migrants seeking to live in Europe, Mexicans migrating to the US, Eastern Europeans migrating to the UK, or refugees and migrants wanting to live in Australia, it seems that all around the world people are on the move.

One of the most important consequences of these events has been the worldwide increase in the popularity of political leaders who have promised to halt or at least significantly restrict migrant inflows (think of the likes of Donald Trump or Pauline Hanson). These events are also reflected in Brexit – the decision of a majority of British voters to support Britain's exit from the European Union (EU). When it is no longer part of the EU the UK will be in a better position to control immigration from European countries.

The headlines dealing with the consequences of large-scale people movements typically emphasise their social and political aspects. For example, in New Zealand, net migration rates continued to rise since 2012, peaking at 1.32 per cent of its population of just under 4.7 million at end of 2016. Would the new migrants flood the country with Asians and take away jobs from Kiwis? These were the views expressed by politicians during the 2017 general election campaign.

You're the economist provides opportunities to consider hypothetical situations, putting you in the shoes of an economist and applying what you have learnt to real-life scenarios. Check your work with the answers provided at the end of each chapter.

You're the economist How can better pricing policies help solve our infrastructure problems?

With strong growth in the Australian economy during the first two decades of this century, the demands on infrastructure provided by governments have increased enormously. Power outages, water shortages and heavily congested roads are just some of the manifestations of this tendency. Many solutions have been suggested for these problems, including calls for state governments, which are responsible for most of these services, to fund massive increases in public infrastructure. There might be political reasons for governments funding more infrastructure. However, a policy of facilitating large increases in infrastructure is fraught with dangers – not the least of which are the adverse environmental effects (known as negative externalities, which we will discuss in the next section) that accompany construction of capital assets, such as power plants and highways.

To reduce gridlocks in major capital cities, in 2019 Infrastructure Australia called on the government to invest more on roads and rails.[1] Rather than proposing increases in infrastructure as a solution to the problem, many economists suggest that an increase in the prices of these services is what is required. They say that in many cases the problems would be solved if prices were increased, but only during certain seasons, at certain times of the day or on certain days of the week. Thus, electricity prices could be higher when peak load occurs in the afternoon or evening, and charges for toll roads and tunnels could be raised during weekday peak periods. Such a policy would echo the private sector where price increases during high demand times are common. You probably know that Uber has surge pricing for busy periods, and that holiday rents are nearly always higher during peak periods, such as school holidays.

As far as road pricing goes, economists also point out that many of Australia's new urban roads, bridges and tunnels are taking less traffic and having a lesser effect on diversion of traffic from existing congested roads than is possible because these new urban infrastructures are subject to user-pay tolls that discourage demand.

In relation to tolls discouraging use of new road infrastructure designed to, among other things, reduce traffic

Answers to 'You're the economist'

How can better pricing policies help solve our infrastructure problems?

As the population grows rapidly in Sydney and Melbourne, new infrastructures, such as roads, bridges and tunnels, are needed to cope with the increasingly congested cities. Recently, for efficiency reasons, the private sector has been more involved in these infrastructure projects. And to recoup their investment, these private construction firms are granted the right to charge their road, bridge and tunnel users a toll. The problem is that tolls discourage travel on these new roads that need to be highly patronised so that congestion on existing roads can be reduced. Some alternative ways for these constructors to raise revenue to recoup their investment would need to be devised if tolls were not imposed on new roads. An obvious solution would be to make travel free on new, congestion-busting roads while charging a toll on congested sections of existing roads. The tolls will increase patronage of those new roads, reduce that of existing congested roads, and at the same time the tolls can be collected to fund the construction of these congesting-busting roads. Of course, there would need to be many, sometimes complex, changes in government policy, legislation and regulation to allow this to happen. If you said that new private-sector-funded road infrastructure could involve contracts with government that gave construction firms the right to toll existing untolled roads but not the new infrastructure, YOU ARE THINKING LIKE AN ECONOMIST.

Explore global issues influences on economics with the **Global perspective** boxes. These present ways in which international phenomena affect the chapter topic.

Global perspective Applicable concepts: barriers to entry, patents, and economies of scale

Tesla: swapping one barrier for another?

In June 2014, Elon Musk, the CEO and co-founder of Tesla – a small but highly innovative car manufacturer in the US – announced that his company would be making all of its patents freely available. As well as producing a series of sexy battery-electric saloon cars with dazzling performance, Tesla also owns an increasing number of its own charging stations, which are the electric equivalent of the familiar petrol-retailing service stations.

An interesting and valuable characteristic of Tesla's technology is that it enables its electric cars to travel much further than its rivals before the batteries needs charging. This characteristic does much to reduce the degree of so-called 'range anxiety' – anxiousness about the possibility that the car's batteries will go flat before the driver's destination is reached. (Tesla cars are powered by batteries alone.) Tesla has also been working on a battery swap program that allows a new bank of batteries to be installed at one of its charge stations in less time than it takes to fill the tank of a conventional car.

Tesla's announcement that it would be making its patents freely available caused a flurry of wide-ranging explanations for this move. These included suggestions that Musk had seen the light and was simply spreading the love by embracing the idea that a world in which ideas are freely exchanged would be a better place.

Contrary to this view, and consistent with their assumption that individuals rationally pursue their own self-interest, economists looked for a more plausible explanation of Tesla's move. Of the many interpretations proffered by economists there are two that we shall look at here.

In order to understand these explanations, we need to be aware of one more important feature of Tesla's business. This is their investment in a huge battery plant in the US. This plant has become known as the 'Gigafactory'. With assistance from leading Japanese lithium-ion battery producer Panasonic, Tesla has built a huge plant employing 6000 people that will reap enormous economies of scale. It is estimated that these economies will reduce battery prices by at least 30 per cent, leading to a significant decrease in the price of new electric cars as well as a decrease in their overall running costs. In addition, this new plant will produce lower-cost batteries for a wide range of other

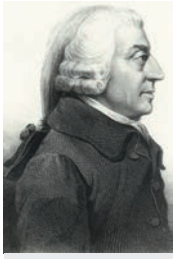
FEATURES WITHIN CHAPTERS

Does economics promote greed?

Economics and ethics

As you study this book you will discover that the economic model is based on the assumption that economic agents pursue their own self-interest. This self-interest is said to be enlightened (i.e., **enlightened self-interest**) in that individuals pursue their own ends subject to the set of laws and social mores that prevails in society. Critics of economics sometimes argue, however, that the economic model promotes unenlightened self-interest – sometimes described as greed. Indeed, some individuals have used the notion of economic self-interest to justify their own greed – as did Gordon Gekko, played by actor Michael Douglas, when he famously declared that 'greed is good' in the film *Wall Street*. With the onset of the global financial and economic crisis (GFEC), greed again reared its ugly head when a number of individuals and organisations were accused of unethical behaviour and some were even convicted for criminal activities.

In spite of some opportunistic interpretations of the relevance of the assumption of self-interested behaviour, the mainstream economic model does not set about promoting greed. It is concerned to promote the efficient working of free markets subject to conformity to an underlying set of ethical or moral principles – a set of principles that is not to be traded for increased material welfare. The founder of modern economics, Adam Smith, pointed out over two centuries ago that the gaining of economic advantage by nonconformity to these principles does not promote the economic welfare of the community. Management guru Charles Handy, in *The Empty Raincoat*, sums up the case by referring to Smith's 18th-century writings:



alamy stock Photo/Pictorial Press Ltd

Adam Smith

Explore real-world ethical issues and how they apply to economic theory in the **Economics and ethics** boxes.

In summary

- Under the **law of supply**, any change in price will cause a direct change in the quantity supplied by a firm, ceteris paribus.
- The **market supply curve** is the horizontal summation of individual supply curves.
- Under the **law of supply**, any change in price will cause a **change in quantity supplied**, represented by a movement along the supply curve.
- Changes in **non-price determinants** produce a shift in the supply curve and not a movement along the supply curve.
- An **increase in supply** causes a rightward shift of the supply curve. Vice versa.

Review your understanding of key ideas with the **In summary** boxes at the end of each main section throughout the chapter.

Labour

Labour is the mental and physical capacity of workers to produce goods and services. The services of farmers, factory workers, lawyers, professional footballers and economists are all *labour*. Both the number of people available for work and the skills or quality of workers measure the labour resource. One reason why nations differ in their ability to produce is that human characteristics, such as the education, experience, health and motivation of workers, differ among nations. And it is this fundamental difference in labour productivity that explains the wide disparity of income and living standard between nations.

Not only do the characteristics of labour determine its productivity but this productivity is also affected by the way in which labour is organised in the workplace. During the Industrial Revolution of the 18th and 19th centuries, the productivity of labour increased enormously as workers took on specialised tasks in factories.

Labour
The mental and physical capacity of workers to produce goods and services.

Important **Key terms** are marked in bold in the text and **defined in the margin** when they are used in the text for the first time. A full list of key terms is also available in the glossary, which can be found online on CourseMate Express.

END-OF-CHAPTER FEATURES

At the end of each chapter you will find several tools to help you to review, practise and extend your knowledge of the key learning objectives.

Key concepts

Business cycle	Economic growth	Solow model of economic growth
Recession	Leading indicators	Exogenous technological progress
Expansion	Coincident indicators	Endogenous growth model
Peak	Lagging indicators	Goals of macroeconomic policy
Trough	GDP gap	

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand the concept of the business cycle 'rollercoaster' and its two phases

- The **business cycle** refers to alternating periods of economic growth and contraction, which can be dated by changes in output, income, sales and employment measures. Each business cycle can be divided into two distinct phases: **recession** and **expansion**. A *recession* is a downturn in the business cycle in which output, sales and employment decline. In practice, a recession is usually defined as occurring when real GDP contracts in each of at least two consecutive quarters. An *expansion* is an upturn in the business cycle during which real GDP, employment and other measures of aggregate economic activity rise. In practice, an expansion is regarded as having commenced when real GDP expands for at least two consecutive quarters. The endpoint of an expansion is referred to as the **peak**, and the endpoint of a recession is referred to as the **trough**.

2. Develop an understanding of business cycle indicators, including how they are used to determine the phase of the business cycle and to forecast where it may be headed in the near future

- Economic indicator variables are used to help better understand and predict the business cycle. They are usually classified into three categories:
  - **Leading indicators**: variables that change direction before the economy shifts from one phase into another (say, from an expansion into a recession).

Key concepts provide a list of the key terms throughout each chapter.

The **summary** links back to each learning objective and summarises the important concepts covered throughout the chapter.

END-OF-CHAPTER FEATURES

Study questions and problems and **multiple-choice questions** your learning.

Study questions and problems

1 What is the business cycle, and what is its fundamental cause?

2 What are the two phases of the business cycle? Explain. What do you call the point at which a business cycle expansion finishes? What is the point at which a business cycle recession finishes?

3 What are the three main types of business cycle indicators? Explain.

4 Shown here are real GDP figures for each of 10 quarters.

Quarterly real GDP (billions of dollars)

Quarter	Real GDP (billions of dollars)	Quarter	Real GDP (billions of dollars)
1	250	6	400
2	300	7	500
3	350	8	600
4	325	9	575
5	300	10	620

Plot these data points and identify the two phases of the business cycle. What are some of the consequences of a prolonged decline in real GDP? Is the decline in real GDP from $600 billion to $575 billion a recession? Why or why not?

Multiple-choice questions

(Note: For some of the following more than one answer may be correct)

1 The phases of a business cycle are:
- **a** upswing and downswing.
- **b** full employment, depression, expansion and plateau.
- **c** full employment and unemployment.
- **d** recession and expansion.

2 Which of the following is *not* a coincident indicator?
- **a** Personal income.
- **b** Industrial production.
- **c** Money supply, M3.
- **d** Manufacturing and trade sales.

3 The point at which a recession ends is called the:
- **a** expansionary phase.
- **b** trough.
- **c** contractionary phase.
- **d** peak.

Guide to the online resources

FOR THE INSTRUCTOR

Cengage is pleased to provide you with a selection of resources that will help you to prepare your lectures and assessments, when you choose this textbook for your course.
Log in or request an account to access instructor resources at au.cengage.com/instructor/account for Australia or nz.cengage.com/instructor/account for New Zealand.

MINDTAP

Premium online teaching and learning tools are available on the *MindTap* platform – the personalised eLearning solution. *MindTap* is a flexible and easy-to-use platform that helps build student confidence and gives you a clear picture of their progress. We partner with you to ease the transition to digital – we're with you every step of the way.

The *Cengage Mobile App* puts your course directly into students' hands with course materials available on their smartphone or tablet. Students can read on the go, complete practice quizzes or participate in interactive real-time activities.

MindTap for Layton's Economics for Today is full of innovative resources to support critical thinking, and help your students move from memorisation to mastery! Includes:

- Layton's Economics for Today eBook
- Quizzes
- Graph tutorials
- GraphBuilder workships
- Aplia Problem Sets
- Concept clips
- Case Projects
- And more!

MindTap is a premium purchasable eLearning tool. Contact your Cengage learning consultant to find out how *MindTap* can transform your course.

CENGAGE | MINDTAP

INSTRUCTOR'S MANUAL

The Instructor's manual includes:

- Learning objectives
- Chapter outlines
- Chapter summaries
- Key concepts
- Hints for effective teaching
- Tutorial discussion questions
- Solutions for all activities and questions from the text

COGNERO® TEST BANK

A bank of questions has been developed in conjunction with the text for creating quizzes, tests and exams for your students. Create multiple test versions in an instant and deliver tests from your LMS, your classroom, or wherever you want using Cognero. Cognero test generator is a flexible online system that allows you to import, edit, and manipulate content from the text's test bank or elsewhere, including your own favourite test questions.

POWERPOINT™ PRESENTATIONS

Use the chapter-by-chapter PowerPoint slides to enhance your lecture presentations and handouts by reinforcing the key principles of your subject.

ARTWORK FROM THE TEXT

Add the digital files of graphs, tables, pictures and flow charts into your learning management system, use them in student handouts, or copy them into your lecture presentations.

CASE PROJECTS AND SOLUTIONS

Assign case projects for students to further explore the key chapter concepts with real-world economics problems. Detailed solutions included.

FOR THE STUDENT

MINDTAP

MindTap is the next-level online learning tool that helps you get better grades!

MindTap gives you the resources you need to study – all in one place and available when you need them. In the *MindTap Reader*, you can make notes, highlight text and even find a definition directly from the page.

If your instructor has chosen *MindTap* for your subject this semester, log in to *MindTap* to:

- Get better grades
- Save time and get organised
- Connect with your instructor and peers
- Study when and where you want, online and mobile
- Complete assessment tasks as set by your instructor

When your instructor creates a course using *MindTap*, they will let you know your course link so you can access the content. Please purchase *MindTap* only when directed by your instructor. Course length is set by your instructor.

MindTap for Layton's Economics for Today includes *Aplia* – the online homework and assessment solution that helps you to better prepare for class and exams.

Preface

Text with a mission

The original reason for writing the first edition of *Economics for Today* was to present, in an engaging style, the basic principles of microeconomics and macroeconomics to students who will take a one-semester course in economics. With the growth of business studies in areas outside economics, including a range of increasingly popular conversion masters programs, there has been a huge increase in the number of students studying introductory economics as a terminal course. While this book is aimed at these students, it also provides a firm foundation for students who will progress to further studies in economics.

Rather than taking an encyclopaedic approach to economic concepts, *Economics for Today* focuses on some of the most important tools in economics, such as supply and demand and marginal analysis, and applies them to clearly explain real-world economic issues. Every effort has been made to make *Economics for Today* the most 'reader-friendly' text on the market.

This book was written to simplify the often-confusing array of technical economic analyses that forces some students simply to memorise in order to pass. Instead, it presents an intuitive and balanced approach that effectively explains the application of fundamental economic principles. After reading this book, the reader will be able to understand and competently engage with many of today's contemporary economic issues and policy debates.

What's new in the seventh edition

The key to the success of the first six editions of *Economics for Today* was the great effort we put into making sure each edition was contemporary and fresh, with reviewer feedback from the previous edition appropriately considered. We have again done our best to ensure this is the case in this seventh edition. For example, based on reviewer feedback, we have introduced a new section in Chapter 19, 'Macroeconomic policy issues: reflecting on economic and debt crisis', where we reflect on the European sovereign debt crisis and the COVID-19 economic crisis. This chapter consolidates – and updates through to 2020 – the relevant material and also discusses the economic impact throughout Chapters 11–18. Chapter 19 therefore performs a similar pedagogic function as Chapter 10, which focuses on two important topical issues – climate change and behavioural economics. The analysis of carbon pollution and climate change also includes consideration of recent policy developments in carbon pollution abatement. In this new edition, much effort has been put into a more comprehensive examination of development of behavioural economics and its practical applications. Combining economics and cognitive psychology, behavioural economists conduct research to investigate human behaviour that contradicts or appears to contradict the rational choice theory. We will look at important insights that help government to formulate and implement more effective public policies, with special reference to the design of strategies to combat the COVID-19 pandemic. We have enjoyed undertaking this updating task and believe this latest edition represents a substantial development upon previous editions.

While we have completely updated, where necessary, all data and exhibits and included relevant recent macroeconomic and microeconomic real-world events, we have also sought to further streamline the book to ensure we remain consistent with our original goal of writing an easy-to-understand introductory text that can be used as a one-semester introduction to the fundamentals of the fascinating field of economics. In particular, in each of the microeconomics chapters, we use the same hypothetical case set within the real-world context to introduce the key economic concepts and theories so that the readers will experience more continuity in their reading. The macroeconomic chapters (11–19) have been significantly streamlined without sacrificing any

important content material. We have also updated the feature sections within each chapters using more recent cases and examples for more relevant applications of economic concepts to analyse economic and public issues.

The result is, we believe, an even more accessible presentation of the core microeconomic and macroeconomic ideas needed by modern business students. Only essential material is included and key concepts are explained in clear and simple terms. We have also sought to rework the end-of-chapter study materials, with much of this material having been refreshed.

As with the previous editions, in making the many revisions incorporated into this seventh edition, the authors owe a significant and sincere debt of gratitude to the many adopters and potential adopters who have provided feedback.

How it fits together

The text presents the core principles of microeconomics and macroeconomics in an international context. The first 10 chapters introduce the logic of economic analysis and develop the core of microeconomics. Here readers learn the role of demand and supply in determining prices in markets characterised by varying degrees of competition. This part of the book explores issues such as minimum wage laws, market failure, economies and diseconomies of scale, and competition policy. Two topical issues – climate change and behavioural economics – are discussed in depth in Chapter 10.

The next nine chapters develop the macroeconomics part of the text. Using the modern aggregate demand and aggregate supply model, the text develops a clear and workable understanding of the determinants of changes in the price level, national output and employment in the economy. The study of macroeconomics also includes a discussion of a nation's monetary and financial system, explaining how changes in the supply of and demand for liquid financial assets in financial markets influence the wider economy. It also introduces the reader to important issues relating to the conduct of modern monetary and fiscal policy. The 2008–09 GFEC, its continued aftermath through to 2017, and the associated European sovereign debt crisis are then used as major applied real-world case studies in Chapter 19 to draw together the macroeconomic chapters in a very contemporary way. In addition, the economic impacts of COVID-19 are discussed throughout the macroeconomic chapters with a detailed explanation in Chapter 19 (but of course, at the time of writing the complete impact is not yet clear).

Throughout the book, the significance of international influences on national economies is recognised, and to further underline the great importance of international considerations in understanding modern macroeconomics, Chapter 18 is devoted to international matters. For example, readers will learn how the supply of and demand for currencies determine exchange rates, the importance of international trade and how it raises the production and consumption possibilities of nations engaging in trade, whether and when external current account deficits are something to be concerned about, why protection of domestic industries from foreign competition may actually increase unemployment and reduce a country's standard of living, and what the implications are of a high or low value for the dollar.

Text flexibility

Economics for Today is easily adapted to lecturers' preferences in the sequencing of microeconomics and macroeconomics topics. The text can be used in a macroeconomic–microeconomic sequence by teaching the first four chapters and then Chapters 11–19. The microeconomics content can then be covered with Chapters 5–10. This approach allows readers to identify with macro issues – which tend often to be in the news – before moving to microeconomics.

The book has 19 chapters and will allow students to cover approximately one-and-a-half chapters each week over 12 teaching weeks. Alternatively, some chapters and some parts of chapters can be omitted at the

discretion of the lecturer. For example, each of the two topical issues discussed in depth in Chapter 10 can be used in a stand-alone manner, with either or both being included in the curriculum. Similarly, since it is very much an application of macroeconomic policy in action, Chapter 19 – or sections of it – could be omitted. Some further proposals along these lines are included in the Instructors Manual.

Finally, an alternative placement for Chapter 18, 'International trade and finance', is also possible. As well as incorporating international issues throughout, *Economics for Today* explicitly addresses international influences on national economies in Chapter 18. Some instructors may prefer to cover this chapter earlier – immediately after Part 4, for example.

Special features

Each chapter contains a number of current real-world exercises and topics for discussion:

- 'You're the economist' sections in each chapter ask readers to answer a simple question related to the topic being discussed. Answers are provided at the end of the chapter.
- Global perspective sections in each chapter highlight chapter topics in a global context.
- Ethical considerations are referenced throughout the text. These sections highlight the fact that many economic issues – and how they are dealt with by policymakers – often involve ethical dimensions.
- 'Analyse the issue' sections in each chapter provide a brief case study for readers to analyse.
- The explicit listing of learning objectives for each chapter, along with a dot-point summary at the end of each section dealing with a learning objective, followed by a detailed summary of all learning objectives' outcomes at the end of the chapter allows readers to efficiently acquire and retain the important lessons of each chapter.

Dedication

This edition of *Economics for Today* is dedicated to all our readers – past, present and future – particularly those who go on to become the economists of tomorrow.

Acknowledgements

Many people have assisted us in very many ways to bring this seventh edition to fruition. They include Dr Ted Boss (University of Alabama at Birmingham), Professors Stan Hurn and Clevo Wilson from QUT, Emeritus Professor Ted Kolsen and Dr Alan Williams. We also wish to thank the team at Cengage Learning, whose support and understanding have made the task so much easier for us. Special thanks must also go to our spouses, Monica Layton, Veronica Horgan, and Carmen Tang, for their help and ongoing support.

Thanks are also due to the many reviewers who have provided numerous insightful suggestions for the current and previous editions, which we have attempted to take up wherever possible:

- Andrea Henderson – RMIT
- Jayanath Ananda – Central Queensland University
- Galina Williams – Central Queensland University
- Martin O'Brien – University of Wollongong
- Bradley Steffenssen – PIBT
- Andrew Grinko – Western Sydney University
- Philip Gibbs – UNSW Foundation Studies

Allan Layton

Tim Robinson

Irvin B. Tucker

Dinusha Dharmaratna

Tommy Tang

Cengage Learning would like to thank the following people for their work on the online resources for this seventh edition:

- Bencigar Arachi (Central Queensland University) for case projects and practice quizzes
- Boon Lee (Queensland University of Technology) for the instructor manual, testbank and PowerPoints.

About the authors

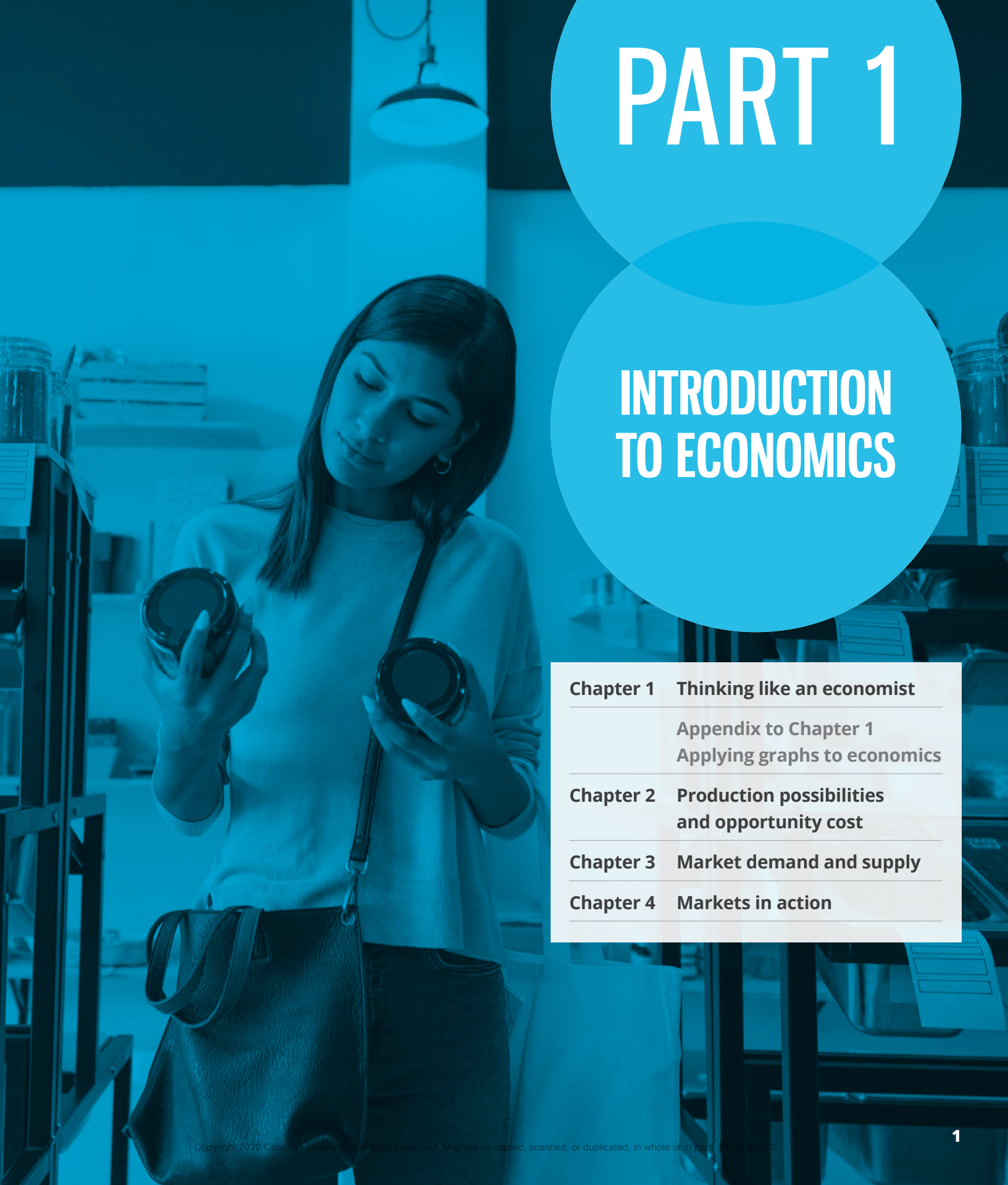
Professor Allan Layton, BEcon. (Hons), MEcon., PhD, received his doctorate from the University of Queensland. He has held academic posts at La Trobe University, Griffith University, Macquarie University, Columbia University in the USA, the University of Queensland, the Queensland University of Technology (QUT) and the University of Southern Queensland (USQ). At QUT he was Head of School of Economics and Public Policy (1992–95), and then Head of School of Economics and Finance (1995–2006). In 2006 he moved to USQ where he was Pro Vice-Chancellor (Institutional Partnerships) (2006–10), Pro Vice-Chancellor (Research Development) (2011–12) and, concurrent with these appointments, Dean of the Faculty of Business and Law through until 2013, after which he continued at USQ as Professor of Economics until 2018 at which time he was awarded the title of Professor Emeritus by the University. His research and teaching interests span international business-cycle analysis, financial and monetary economics, macroeconomic policy and applied econometrics. He is the author of *Modern Australian Macroeconomics* and has published widely in international academic journals. He has also served on a number of boards and acted as consultant to both private- and public-sector organisations.

Professor Tim Robinson, BEcon. (Hons), PhD (Qld), worked in the private sector for seven years prior to embarking upon an academic career. Part of his time in the private sector was spent running his own small business. He is currently Emeritus Professor of Economics at Queensland University of Technology (QUT) where he continues to lecture in first year economics. Prior to his becoming Emeritus Professor he held the position of Professor and Head of School of Economics and Finance at QUT. As well as having taught introductory and intermediate economics at a number of universities over many years, he has also published in the areas of environmental economics, applied welfare economics, history of economic thought, and the teaching of economics. Books he has authored or co-authored include *Economic Theories of Exhaustible Resources*; *Work, Leisure and the Environment*; *Macroeconomics: A Contemporary Introduction* and *Microeconomics: A Contemporary Introduction*. He has also undertaken economic consulting work in the areas of curriculum development, financial investment, taxation and the environment.

Dr Irvin B. Tucker, BS, SC, PhD, has taught introductory economics at the University of North Carolina at Charlotte and the University of South Carolina. He is former Director of the Center for Economic Education at the University of North Carolina at Charlotte and a long-time member of the National Council on Economic Education. He has published widely in professional journals and is the author of *Survey of Economics*.

Dr Dinusha Dharmaratna BSc (Hons), MEd, PhD worked at the Monash University for 6 years as the first year Economics coordinator prior to joining Monash College. She is currently working as a unit leader for Macroeconomics and teaches economics and finance. Not only she has taught and developed curriculum for first- and second-year level micro and macroeconomics at Monash University and Monash College but also authored a book and a book chapter in the area of environmental and resources economics. In addition to her interest in economics, she also actively engages in many teaching and learning research projects in blended learning strategies such as gamification, peer-review assessments, principles of cognitive psychology to promote collaboration and engagement.

Associate Professor Tommy Tang, B.Soc.Sc. (Hons), MEd, PhD, SFHEA has taught economics and statistics at the undergraduate and postgraduate levels for more than two decades, with a main research interest in economics education. His contribution to teaching in economics and statistics was recognised at the university and national levels, winning several teaching awards, including the prestigious Australian Learning and Teaching Council Teaching Excellence Award in 2011, and his appointment to the Australian Office of Learning and Teaching Economics Learning Standards Working Party to develop the learning standards for undergraduate and postgraduate economics programs in 2012–13. He has taken several administrative roles at university, has been selected as one of the 24 Inaugural reviewers in the QUT ExPRET (Expert Peer Review of Educational Practice) Program and is a senior fellow of the UK Higher Education Academy.



PART 1

INTRODUCTION TO ECONOMICS

1 Thinking like an economist

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand why the problem of scarcity is central to economics.
2 Explore the differences between macroeconomics and microeconomics.
3 Discover the methodology of economics.
4 Understand the hazards of the economic way of thinking.
5 Understand why economists disagree.
6 Discover the practical applications of economics.

Economics provides a framework for explaining and predicting human behaviour closely connected with their material wellbeing. Other disciplines (e.g., psychology and sociology) also have a keen interest in explaining human behaviour. However, economists have a particular way of looking at the world that is different from other disciplines. As you become familiar with the economic approach, you will begin to see the world around you through different eyes. You will discover a whole range of social and economic issues requiring more powerful tools than just common sense. The economic way of thinking is important because it provides a logical framework that can be used to help us make better decisions at the individual, business and government levels. As you master the methods explained in this book, you will appreciate that economics provides a valuable approach to solving many real-world puzzles and issues.

Getty Images/FAYEZ NURELDINE

There is no better way to discover the relevance of economics in the world today than to look at the deliberations of world leaders at the G20 annual summit. The G20 is an international organisation consisting of the world's leading 19 economies plus the European Union. Together these economies are responsible for about 85 per cent of world output and two-thirds of world population. The G20 addresses major economic challenges faced by the world. At the end of the 2020 Saudi Arabia

G20 summit meeting (held virtually due to the COVID-19 pandemic) the world leaders issued a joint declaration summarising the global challenges arising from the coronavirus pandemic and the concerned efforts and strategies necessary to overcome them. The key initiatives of the declaration are:

- An urgent need to bring the COVID-19 pandemic under control. (In **Chapter 10** we will look at how behavioural economics can help design and implement public health polices to influence behaviour for more effective containment of the virus.)
- Taking exceptional fiscal and monetary measures to combat the sharp pandemic-induced economic contraction and to provide the impetus for smooth and sustainable post-pandemic recovery. (**Part 5** of this book discusses the working of the macroeconomy, and the use of monetary and fiscal policies to deal with economic fluctuations and their effectiveness.)
- Promoting sustainable, balanced and inclusive economic growth. (The drivers of economic growth, and the associated issues of unemployment and sustainability will be discussed in **Chapters 2, 12** and **13**.)
- Appropriate treatment of government debts following the unprecedented fiscal stimulus spending to ensure smooth global recovery, especially for small, poor countries. (The issue of sovereign debts will be discussed in **Chapter 19**.)
- Support of international trade and investment conducted in a free and fair environment. (Trade and international finance will be the focus of **Chapter 18**.)
- Tackling pressing environmental challenges, such as climate change and biodiversity loss in the course of promoting economic growth and energy security. (**Chapter 10** discusses climate change and the energy problem.)

In addition to the topics mentioned above, in this text you will find out why competition is good for society (**Chapter 7**) and how governments endeavour to control the excesses of large companies that wield great economic power (**Chapter 8**). You will also apply the economist's toolbox to shed light on strategic pricing and output decisions by businesses (**Chapters 5** and **9**). You will learn why protection of domestic industries and jobs in the face of competition from emerging economies like China and India is becoming an increasingly contentious issue (**Chapter 18**). And there are many more of these important issues discussed in the pages that follow. In short you will see just how economic theories and policies affect our daily lives.

In this chapter we look first at the concepts of scarcity. Next, we examine the steps involved in building an economic model to explain aspects of the economy we observe around us. Then we look at some new developments in economics and explain why economists may disagree with one another.

1 The problem of scarcity

To begin with, we look at the many important social and public issues relevant to economics. Economics can provide us with a framework to address these issues. However, you may ask: What is economics? At a general level, economics is the study of how human beings provide for their material wellbeing. What distinguishes economics from other social sciences that also study human behaviour? The underpinning theme of economics is limited resources or **scarcity**. From the economist's perspective, coping with scarcity to achieve given ends is the essence of the human condition. Our world is a finite place where people, both individually and collectively, face the problem of scarcity. Scarcity is the condition in which human wants are forever greater than the available supply of resources. Because of scarcity, it is impossible to satisfy every desire. Pause for a moment to think of some of the ways in which scarcity affects our everyday lives. Perhaps you wish you could have more restaurant meals, the latest model of iPhone, a pair of designer shoes and so on.

Scarcity
The condition in which human wants are forever greater than the available supply of resources.

You may think your scarcity problem would disappear if you were rich. Surely wealth does go some way to improve the situation. However, no matter how affluent an individual is, the wish list continues to grow. We are familiar with the 'rich and famous' who never seem to have enough. Although they live well, they still desire finer homes, faster planes and more Jimmy Choo shoes. They would like more material goods and more leisure time in which to use them. It is true that human beings are not only concerned with their own wellbeing. Perhaps they want better life for others – shelter for the homeless, jobs for the unemployed, reducing global inequalities,

saving endangered species, and devoting personal time to campaign for the environment. These other-regarding desires are more intense for some people than others. Nonetheless, as individuals we just do not have resources to achieve all of these ends.

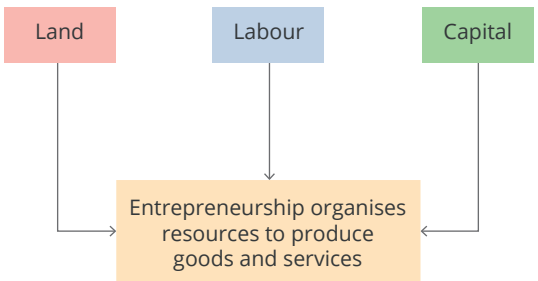
What is true for individuals also applies to governments, which are constrained by limited resources and at the same time face numerous competing demands for government funding from different parts of the society. The community want more hospital beds to reduce the waiting list for elective surgeries; schools want more teachers to reduce class sizes to improve quality of education; universities want more funding for research; automobile associations lobby for more roads, bridges and tunnels to relieve traffic congestion; social welfare groups petition for more public housing for the homeless, more unemployment benefits for the jobless and more pension for the aged. The list goes on.

Of course, scarcity is a fact of life throughout the world. In many parts of South America, Africa and Asia the problem of scarcity (e.g., clean water and medical supplies) is often life-threatening. On the other hand, in North America, Europe, Australasia and much of Asia there has been substantial economic growth and development. Although life is much less gruelling in the more advanced countries, the problem of scarcity exists since these developed countries do not have unlimited funding to satisfy the ever-increasing aspirations brought about by rising living standard, as discussed earlier. In Australia (often called the 'lucky country') state governments search for innovative ways to raise taxes for the funding of schools and hospitals. The federal government's desire to spend on the poor and on higher education, highways and defence exceeds the tax revenue it receives to pay for these programs. So, not even the Australian government escapes the problem of scarcity.

Because of the economic problem of scarcity, no society has enough resources to produce all the goods and services necessary to satisfy all human wants. Production is the process of converting resources (or inputs) into goods and services. These inputs are also called **factors of production**. Economists divide factors of production into three categories: *land*, *labour* and *capital* (see **Exhibit 1.1**).

Factors of production
Inputs used to produce goods and services. Economists divide factors of production into three categories: land, labour and capital.

Exhibit 1.1 The three factors of production



Factors of production are the inputs organised by entrepreneurship (a special type of labour) to produce goods and services. Factors of production can be categorised into three types: land, labour and capital.

Land

Land
A shorthand expression for any resource provided by nature.

Land is a shorthand expression for any resource provided by nature. Land includes those resources that are gifts of nature available for use in the production process. Land includes anything natural above or below the ground, such as forests, minerals, oil, wildlife and fish. Other examples are rivers, lakes, oceans, the atmosphere, the sun and the moon. Pursuits such as farming, fishing, manufacturing and retailing all use land to a greater or lesser extent. Land in its usual sense is an important input to the housing industry. Two broad categories of natural resources are *renewable resources* and *non-renewable resources*. Non-renewable resources are basic inputs that nature cannot replace; for example, there is only so much coal, oil, copper and iron ore in the world. Renewable

resources are basic inputs that nature can replace with or without help from human beings, such as forests, crops, fish and clean air.

Labour

Labour is the mental and physical capacity of workers to produce goods and services. The services of farmers, factory workers, lawyers, professional footballers and economists are all *labour*. Both the number of people available for work and the skills or quality of workers measure the labour resource. One reason why nations differ in their ability to produce is that human characteristics, such as the education, experience, health and motivation of workers, differ among nations. And it is this fundamental difference in labour productivity that explains the wide disparity of income and living standard between nations.

Labour
The mental and physical capacity of workers to produce goods and services.

Not only do the characteristics of labour determine its productivity but this productivity is also affected by the way in which labour is organised in the workplace. During the Industrial Revolution of the 18th and 19th centuries, the productivity of labour increased enormously as workers took on specialised tasks in factories. This specialisation was a development that was discussed by the father of modern economics, Adam Smith, in his example of the division of labour in a pin factory, where each of the tasks involved in making a pin – drawing the wire, forming the point, etc. – was carried out by a different worker.

Entrepreneurship is a special type of labour, and is the creative ability of individuals to seek profits by combining resources to produce new or existing products. The *entrepreneur* is a motivated person who seeks profits by undertaking risky activities such as starting new businesses, creating new products or inventing new ways of accomplishing tasks.

Entrepreneurship
The creative ability of individuals to seek profits by combining resources to produce new or existing products.

Entrepreneurs are the agents of change who help bring material progress to society. The origins of the world's largest pharmaceutical group, GlaxoSmithKline, can be traced to New Zealand entrepreneur Joseph Nathan, a poor Jewish immigrant from London's East End who developed interests in shipping, railways and eventually dried milk. This dried milk, which was first produced at Bunnythorpe near Palmerston North, was given the trade name Glaxo. By the 1930s it had become the pre-eminent dried milk for babies, and Glaxo was a household word. At the end of the 20th century the trade name that Nathan had registered in 1906 became GlaxoSmithKline, after Glaxo had been involved in mergers with leading pharmaceutical companies. The Bunnythorpe dried milk factory was closed in 1974 but remains as a historic building – a lasting reminder of the entrepreneurial spirit of Joseph Nathan.

Incidentally, GlaxoSmithKline was fined $492 million by the Chinese authorities in 2014 when it was found to have bribed doctors and other health professionals to encourage them to prescribe their products. In response Glaxo declared that it would change its incentive system that had encouraged this behaviour. Throughout this book we will look at ethical issues such as this. As you will see, it is not uncommon for inappropriate incentive systems to encourage unethical behaviour.

Capital

Capital is the physical plant, machinery and equipment used to facilitate the production of other goods. Capital goods, also known as producer goods, are human-made goods that do not directly satisfy human wants. Capital before the Industrial Revolution meant a tool, such as a hoe, an axe or a bow and arrow. In those days, these items served as capital to build a house or to provide food for the dinner table. Today, capital also consists of factories, office buildings, warehouses, robots, trucks and distribution facilities. University buildings, computers and software are also examples.

Capital
The physical plant, machinery and equipment used to produce other goods. Capital goods are human-made goods that do not directly satisfy human wants.

The term *capital* as it is used in the study of economics can be confusing. In everyday conversation, capital often means money or the money value of assets, such as stocks, bonds or the deeds to real estate. This is actually *financial capital*. Money is not capital as defined in economics because it does not *directly* facilitate production; although it does have important functions of facilitating transactions and storage of purchasing power, a topic we will discuss at length in **Chapter 15**.

In summary

- **Economics** is the study of how individuals and society choose to allocate scarce resources in order to satisfy unlimited wants.
- **Scarcity** is the fundamental economic problem that human wants exceed the availability of resources used to satisfy our wants.
- **Factors of production** are inputs used to produce goods and services, and are classified as land, labour and capital.

2 Macroeconomics and microeconomics

People often think that economics is about money, the stock market, or banking and finance. While economists do study the roles of money in the economy and working of the financial system, neither money nor the financial system can adequately define economics as a way of thinking.

Economics
The study of how human beings deal with the problem of scarcity to satisfy unlimited wants.

Economics is the study of how human beings deal with the problem of scarcity to satisfy unlimited wants. The basis for this definition is the perpetual problem of scarcity forcing people to make choices. We will discuss choice and the related concept of cost in detail in **Chapter 2**. It is important to note that scarcity is a *relative* concept. Take sand as an example. It is in plentiful supply if you live near a desert, and its availability would far exceed the want for it, meaning it is not a scarce resource in a desert. But it is mostly likely a scarce if you live in a modern city and want a Zen sand garden in your backyard.

Now we will have a preview of the two main branches in economics: *microeconomics* and *macroeconomics*.

Microeconomics

Microeconomics
The branch of economics that studies decision-making by a single individual, household, firm or industry.

Microeconomics is the branch of economics that studies decision-making by a single individual, household, firm or industry. The focus is on the behaviour of individual economic units, be they consumers, businesses or government departments. In microeconomics, some of the questions we ask concern how prices of goods and services are determined; how consumers and businesses respond to changes in prices; the impacts of government policies, such as price control, taxes and subsidies on the performance of individual markets; and how to instil empathy to foster more compliance with new rules and restrictions to contain coronavirus. More specifically, is it a good idea for the government to protect tenants by imposing rent control? Why does the government provide measles vaccinations free of charge to parents? In the first half of this book we will look at many of these questions.

You're the economist

What is the price of economic freedom in Cuba or China?

Because most of us have grown up and lived in market economies, we are not really aware of the ways in which these economies differ from centrally controlled economies like those of the communist countries of Vietnam, Cuba and even China, where markets still have only a limited role to play.

To give you some idea of these differences we will first take a very brief look at Cuba, where recent initiatives designed to modernise the economy give a stark picture of the conditions that Cuban residents experience in their daily lives. For example, in 2010 the government released a list of 178 activities that would now be available to people who wanted to be self-employed rather than being employed by the state. Nonetheless, these newly self-employed people would not be allowed to accumulate property. These recent changes followed an earlier, but limited liberalisation in the 1990s that allowed people to let out rooms and to run private restaurants. When one considers that all but the tiniest of businesses are run by the state, and that there are continuing widespread controls on prices and availability of goods, it can be seen that economic freedom is a scarce commodity in Cuba.[1]

Even with the widespread liberalisation that has occurred in China in recent decades, the state still has a huge role to play in the economy – a much larger role than it plays in the west. For example, it has been estimated that roughly

half of China's non-farm output is generated by state-owned enterprises (SOEs). According to an Australia Treasury report, in terms of assets, SOEs still owned well above 40 per cent in 2010. And there are signs in more recent times (around 2020) that the state has taken over the control of more and more big private corporations.[2] China's biggest banks are all government owned and the government directs their lending towards economic activities it has chosen rather than allowing free markets to determine which enterprises should be given access to loans. In these and in more subtle ways, the Chinese government is able to direct resources to areas of the economy of its choice.[3] Knowing that scarce commodities attract prices, can you think of any phenomenon that indicates the price that Cubans (and to a lesser extent Chinese citizens) place on the scarcity of economic freedom in their country?

Alamy Stock Photo/Phanie

1 Anon, 'Trying to make the sums add up', *The Economist*, 13 November 2010, pp. 43–4.
2 D. Zhang and O. Freestone. 'China's unfinished state-owned enterprise reforms'. *Economic Round-up 2*, 2013, 77–99, https://treasury.gov.au/sites/default/files/2019-03/5-China-SOE-reforms.pdf
3 For a recent discussion of the extent of market reform in China, see W. M. Morrison, 'China's economic rise: History, trends, challenges, and implications for the United States', *Congressional Research Service*, 9 October 2014, https://fas.org/sgp/crs/row/RL33534.pdf, accessed 10 November 2017.

Macroeconomics

Macroeconomics is the branch of economics that studies the performance and impacts of government policies on the economy as a whole. It applies an overview perspective to an economy by examining economy-wide variables, such as inflation, unemployment, the money supply and the flows of exports, imports and international financial capital. Macroeconomics studies the effect that balancing the federal budget may have on unemployment, investment decision and income distribution, the unintended consequence of raising unemployment benefits on unemployment rate during the pandemic, the effect of increasing the money supply on prices, interest rate and output, and the effect of economic growth on income distribution. Every day in the media, reference is made to the macro economy. Whether it is commentary on the performance of the government in delivering jobs and growth, discussion of the latest data releases pertaining to inflation and the value of the Australian dollar, or concern that growth in wages is getting out of hand, there is always some facet of the macro economy requiring our attention.

Macroeconomics
The branch of economics that studies decision-making for the economy as a whole.

We have described macroeconomics and microeconomics as two separate branches, but they are related. Because the overall economy is the sum or aggregation of its parts, micro changes affect the macro economy, and macro changes produce micro effects.

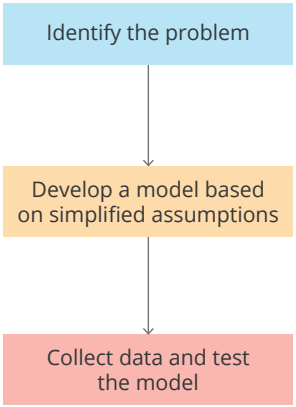
In summary

- **Macroeconomics** applies an economy-wide perspective that focuses on such issues as inflation, unemployment, the growth rate of the economy and international trade.
- **Microeconomics** examines individual decision-making units within an economy.

3 The methodology of economics

Economists use the same *scientific method* used in other disciplines such as criminology, biology, chemistry and physics. The scientific method is a step-by-step procedure for solving problems by developing a theory, gathering data and testing whether the data are consistent with the theory. **Exhibit 1.2** summarises the model-building process.

Exhibit 1.2 **The steps in the model-building process**



The first step in developing a model is to identify the problem. The second step is to select the critical variables necessary to formulate a model that explains the problem under study. Eliminating other variables that complicate the analysis requires simplifying assumptions. In the third step, the researcher collects data and tests the model. If the evidence supports the model, the model is accepted for the time being. If not, the model is rejected or modified.

Identifying the problem

The first step in applying the scientific method is to define the problem. Suppose an economist wishes to investigate the microeconomic question of why motorists have cut back on petrol consumption in the second quarter of 2019.

Developing a model

Model
A simplified description of reality used to understand and predict the relationship between variables.

The second step in our example towards finding an explanation is for the economist to build a **model**, which is a simplified description of reality used to understand and predict the event or phenomenon of interest. A model is built on the foundation of an underlying theory. It looks at the factors, often called *variables*, that explain the event. However, a model emphasises only those variables that are most important to explaining the event. In this respect, models are similar to their underlying theories, which, according to Albert Einstein, 'should be as simple as possible, but not any simpler'. Paring a model down to its simplest possible form is sometimes described as the pursuit of *parsimony*. The purpose of a model is to construct an abstraction from real-world complexities and make events understandable. For example, a map of New Zealand's largest city, Auckland, is far from a precise duplication of this beautiful place (see **Exhibit 1.3**). Just substitute the *satellite* view for the *map* view on your smartphone to see the difference! But a map of the city does help a visitor understand the best way to get about by leaving out the clutter of detail.

Exhibit 1.3 **A map of Auckland CBD**

Getty Images/RUSSELLTATEdotCOM

This map of Auckland CBD is a model because it is an abstraction from the actual beauty of the city.

A model requires simplified assumptions in order to be useful. Someone must decide, for example, whether a map will include only symbols for the major highways or the details of every minor road. In our petrol consumption example, several variables may be related to the quantity of petrol consumed, including consumer incomes, the price of goods other than petrol, the price of petrol and the weather conditions. Because we wish to focus only on the main or critical variables, the economist must be a Sherlock Holmes and use a keen sense of observation to form a model. Using their expertise, the economist must select the key variables that are related to petrol consumption and reject variables that have only a slight or no relationship to petrol consumption. In this simple case, the economist removes the cloud of complexity by formulating the theory that increases in the price of petrol *cause* the quantity of petrol consumed to decrease during the time period.

Testing the model

An economic model can be formulated using verbal arguments, numerical tables, graphs or mathematical equations. You will soon discover that a major part of this book is devoted to building and using economic models. The purpose of an economic model is to enable us to *forecast* or *predict* the results of various changes in key variables. An economic model is useful only if it yields accurate predictions. When the evidence is consistent with the prediction that a change in *A* causes outcome *B*, all other factors remaining constant, there is confidence in the model's validity. This confidence in the model is maintained unless there is evidence that contradicts its prediction. So, if repeated tests indicate that the evidence is inconsistent with the prediction that a change in *A* causes outcome *B*, the researcher rejects the model.

Returning to our falling petrol consumption problem in the second quarter of 2019, the economist gathers data to test the hypothesis that if the price of petrol rises, then petrol purchases fall – all other relevant factors held constant. If there was a sharp rise in the price of petrol in April to June of 2019, then the data were consistent with the hypothesis that the quantity of petrol consumed per month falls when its price rises, assuming no other relevant factors change. When we examine the petrol price data, this was found not to be the case. The petrol price actually dropped to below $1 a litre around that time – an all-time low for the past two decades

in Australia. Does the observation of falling petrol price and falling petrol consumption refute the theory (or model) economists proposed? What was then the cause for the fall in petrol consumption? We will discuss this case in detail in **Chapter 4**.

In summary

- **Models** are simplified descriptions of reality used to understand and predict economic events.

4 Hazards of the economic way of thinking

Models help us to understand and predict the impact of changes in economic variables. A model is an important tool in the economist's toolkit, but it must be handled with care. The economic way of thinking seeks to avoid reasoning mistakes. Two of the most common obstacles to clear thinking are: (1) failing to understand the *ceteris paribus* assumption; and (2) confusing *association* and *causation*.

The *ceteris paribus* assumption

Ceteris paribus
A Latin phrase that means that while certain variables change, all other things remain unchanged.

As you work through a model, try to think of a host of relevant variables assumed to be 'standing still', or 'held constant'. ***Ceteris paribus*** is a Latin phrase that means that while certain variables change, all other things remain unchanged. As in the petrol example discussed earlier, a key simplifying assumption of the model is that changes in consumer incomes and certain other variables do not occur and complicate the analysis. The *ceteris paribus* assumption holds everything else constant and therefore allows us to concentrate on the study of the relationship between two key variables: changes in the price of petrol and the quantity of petrol purchased per month.

As another example, suppose an economist wishes to explain a model for the price and quantity purchased of Rihanna concert tickets. Assume the theory is: 'If the price increases, then the quantity of concert tickets purchased decreases, *ceteris paribus*'. A pitfall in reasoning might occur if you observed that the price of concert tickets increased one year and more tickets were actually bought and not fewer. On the basis of this real-world observation, you declare that the model does not work. Think again! The economist responds that the model is valid, based on the assumption of *ceteris paribus*, and that your observation gives you no reason to reject the model. The reason why the model appeared flawed is that another factor, the screening of a television documentary about Rihanna that occurred just a few months before the concert, *caused* people to buy more tickets in spite of their higher price. If the amount of publicity given to Rihanna and all other factors are held constant as the price of concert tickets rises, then people will indeed buy fewer tickets, as the model predicts.

Association versus causation

Another of the most common errors in reasoning is to confuse *association* (or correlation) and *causation* between variables. Stated differently, you err when you read more into a relationship between variables than is actually there. A model is valid only when there is a genuine cause-and-effect relationship. The model will be questionable when its validity relies solely on an association between variables. Suppose that the hole in the ozone layer increases in size during three different months, and Indonesia's exports to Australia increase during each of these months. The change in the ozone layer is *associated* with the increase in exports, but this does not mean the ozone layer change *caused* the event. Even though there is a statistical relationship between two variables in a number of observations, economists would not be concerned if further increases in the size of the hole in the ozone layer were associated with a fall in Indonesian exports to Australia. The reason is that there is no true economic relationship between the ozone layer and Indonesian exports. A more likely explanation for the increase in Indonesian exports to Australia would be that currency movements have reduced the cost to Australians of buying these exports from Indonesia.

Throughout this book, you will study economic theories and models that include variables linked by stable cause-and-effect relationships established under a sound theoretical framework, in both microeconomics and macroeconomics.

In summary

- ***Ceteris paribus*** means that all other factors that might affect a particular relationship remain unchanged. Reasoning will be unsound if this assumption is violated. Another reasoning pitfall is to think that *association* means *causation*.

5 Why do economists disagree?

Economists disagree for many reasons. Sometimes their arguments are about the facts of the matter and sometimes they are about opinions. Of course, disagreements about opinions are commonplace in all disciplines and are to be expected. For example, one economist might assert that a disadvantaged minority group should be given assistance whereas another may argue that there is no good reason for such positive discrimination. However, when it comes to matters of fact, we have every reason to believe that there should be agreement among practitioners in a rigorous discipline. Thus, engineers are likely to agree on the necessary size of component parts of a road bridge, and physicists are likely to agree on the amount of thrust required to send a rocket into orbit. Even two chefs are likely to agree on the likely effect of a change to the mix of ingredients in a traditional recipe. Yet, in economics there sometimes can be significant disagreements even in this factual realm. For example, one economist might declare that the introduction of a payment for the giving of blood to a blood bank would increase the supply of blood because people react positively to monetary incentives, while another may say that a sense of responsibility of individuals towards the community in which they live can lead to a blood donor scheme that gives better results than a system of monetary rewards.

Recently, some economists have been pushing for performance-related pay for teachers while other economists, and teachers themselves, are worried that such a move may diminish teachers' sense of commitment to their calling. In the 'Global perspective' box that follows we will see why such disagreements might occur. An important reason for such differences of opinion lies in the assumptions economists make about human nature. As you will discover in the following chapters, mainstream economics assumes that individuals are motivated almost exclusively by the pleasure/pain principle. That is, in deciding how to react to a given set of circumstances, individuals weigh up the costs and benefits of a particular action (the pain and the pleasure) in a straightforward way and then decide what to do.

Observing differences in the way in which different economists who use this technique have come up with differing predictions about human reactions to a given set of circumstances, some members of the profession, sometimes assisted by psychologists and other social scientists, have decided to investigate the use of more elaborate assumptions about human nature and what motivates us. The outcome of these investigations has been the development of a branch of economics known as *behavioural economics*.

Global perspective

Applicable concepts: why economists disagree, *ceteris paribus* assumption, and behavioural economics

Economists disagree about how to get the best from teachers

Quite a few economists around the world study the economics of education, and some of them argue that the best way to improve our schools is to introduce performance pay for teachers. Under such a system, teachers would be paid higher wages if they could show that they were better teachers than their peers. This might be demonstrated, for example, by the teacher's classes receiving higher grades on objective tests or by their pupils gaining more university entrance places at the end of their schooling. In contrast to this view, other economists, and teachers themselves, argue that performance pay would reduce the quality of schooling. They say that not only is measuring teaching performance notoriously difficult (*ceteris paribus* does not hold) but that, more importantly, the introduction of performance pay would destroy the sense of obligation that teachers have towards the school community. They say that it is this sense of obligation to fellow teachers, pupils and parents that motivates teachers to teach well.

A similar argument was put forward many years ago by Englishman Richard Titmuss in relation to rewarding individuals for supplying blood for use in medical procedures. Comparing the UK, where the donation system is used, with the US, where payment for blood is made, Titmuss argued that the donation system was superior to the payment system. The Australian economist Henry Ergas, although he did not refer specifically to Titmuss, succinctly put the case for a blood donation system in a 2009 article in *The Weekend Australian* about limits to the usefulness of markets in contemporary society.[1] The following extract from that article makes the case for donation clear:

Alamy Stock Photo/Takatoshi Kurikawa

> [B]lood is scarce but it is collected on a voluntary basis, at least in Australia. If we paid donors, surely they would be better off, as they would get the satisfaction associated with the supply of blood and an increased income (which, if they wanted, they could then donate to worthwhile causes, including the Red Cross). This should induce even existing donors to give more frequently and attract new donors.
>
> But what if injecting a price crowded out less selfish motives, so that paying for what would otherwise be provided on a voluntary basis could reduce supply? May it be that motivations, unlike cocktails, are not readily mixed?
>
> Indeed it may. When we give blood, we do so as an expression of membership of a broader community; we value the act of giving to strangers. Were we paid, the gift relationship would be undermined, even if the money were promptly recycled to other worthy causes.
>
> Setting a price, in other words, would increase the disutility [unpleasantness] of giving blood rather than making it more attractive …
>
> Admitting an ounce of motivation that is self-serving or otherwise tainted can [in some aspects of human interaction] annihilate the credibility of claims that one has the other person's interests at heart.

A similar argument to the one above about the advantages of donating blood rather than selling it can be applied to teaching. If teachers, like blood donors, are motivated by higher ideals than would be satisfied by payment alone, then teachers may give more to the school community when they are motivated by this commitment rather than by performance pay.

So, will consensus be reached among economists as to the best way to deal with the issue of how teachers should be rewarded? This is a question that it is impossible to answer. However, we do know that until consensus is reached, economists will continue to disagree. If a consensus is reached, it could well be that behavioural economics, with its consideration of a broad range of explanatory factors, will have a significant part to play in this outcome.

What do you think?

1. Performance pay for schoolteachers is a topic that has been hotly debated at both the state and national level. Do you agree or disagree that performance pay for teachers will improve schooling outcomes?
2. Can you think of other jobs where concern for community is a very strong motivator? What about jobs where community concerns seem to be of little or no significance? Is there some aspect of the former that is not shared by the latter, or vice versa?

[1] H. Ergas, 'Giving freely as a premium', *The Weekend Australian*, 31 October–1 November 2009, *Inquirer*, p. 8.

Behavioural economics

Behavioural economics investigates the outcome of a particular set of circumstances with a more comprehensive view of human behaviour than is usually used in mainstream economics by incorporating elements of human psychology, such as cognition, emotion and social norms. Thus, according to the behavioural economics view, in addition to the simple pleasure/pain principle, a series of other factors such as ethical and moral beliefs, social mores and class relationships may also determine outcomes. Thus, people may not react to monetary rewards for ethical or religious reasons. A lower-class person in dire need may take a demeaning job whereas an upper-class person in similar circumstances may be unwilling to do so. Citizens may happily pay their taxes because they see that, although not paying tax may make them better off in the short run, in the long run the basis for civil society may be compromised if everyone were to engage in tax avoidance. As we will see shortly (as well as in detail in **Chapter 10**), even differences in the way that simple choices are presented to individuals (known as framing) can lead to very different behavioural outcomes. While it is not fair to say that mainstream economics has neglected these issues, it is true to say that they are given much more emphasis in behavioural economics.

Behavioural economics Study of human behaviour by combining both economic and psychological factors.

Often described as the father of behavioural economics, psychologist and 2002 Nobel Prize winner in economics Daniel Kahneman has researched the myriad ways in which our behaviour can differ from what would be predicted by applying the simple pleasure/pain principle of conventional economics. For example, *The Economist*, in its review of Kahneman's popular book on behavioural economics *Thinking, Fast and Slow,* provided the following example of behavioural outcomes that might not ordinarily be predicted by economists:

> In one experiment described by Mr Kahneman, participants asked to imagine that they have been given £50 behave differently depending on whether they are then told they can 'keep' £20 or must 'lose' £30 – though the outcomes are identical.[1]

In 2017, the Nobel Prize was again won by a behavioural economist – Richard Thaler. Famously, he predicted that behavioural economics would be so successful that it would eventually disappear – mainstream economics would become as behavioural as was required.[2]

[1] Anon, 'Not so smart now: The father of behavioural economics considers the feeble human brain', *The Economist*, 29 October 2011, from the print edition available at https://www.economist.com/books-and-arts/2011/10/29/not-so-smart-now, accessed 15 December 2017.

[2] B. Appelbaum, 'Nobel in economics is awarded to Richard Thaler', *The New York Times*, 9 October 2017, https://www.nytimes.com/2017/10/09/business/nobel-economics-richard-thaler.html, accessed 11 November 2017.

On another research direction, in order to help verify some of the assumptions they make about human behaviour and about the extent to which people respond to incentives, disincentives and opportunities to cooperate with others, economists, and behavioural economists in particular, have developed techniques for conducting controlled experiments designed to mimic real-world scenarios. These activities, which go by the name of **experimental economics**, are proving to be fruitful areas for research.

Experimental economics
A branch of economics in which experiments designed to mimic real-world scenarios are used to explore aspects of human behaviour relevant to economics.

Earlier in this section we looked at the issue of disagreements between economists on matters of opinion as opposed to differences relating to matters of fact. Economists have special terms to describe each of these possibilities: they say that matters of fact are the realm of *positive economics* while matters of opinion involve *normative economics*. This is a very important distinction, which we discuss in the following.

Positive economics

Positive economics deals with facts and therefore addresses 'what is' questions. Positive economic analysis is limited to statements that are verifiable. Positive statements can be proved either true or false by appealing to objective facts. Often a positive statement is expressed 'If *A*, then *B*'. For example, it might be stated that if the national unemployment rate is 7 per cent, then youth unemployment exceeds 80 per cent. This is a positive 'if–then' prediction, which may or may not be correct. The accuracy of the statement is not the criterion for being a positive statement. The key consideration for a positive statement is whether the statement is *testable* and not whether it is true or false. In fact, this positive statement is false. (In Australia in recent times, the youth unemployment rate has been approximately three times the overall unemployment rate.)

Positive economics
An analysis limited to statements that are verifiable.

As we said earlier, the more rigorous a science, including economics, the more likely it is that its practitioners would agree on positive issues.

Normative economics

Instead of using objective statements, an argument can be phrased subjectively. **Normative economics** is concerned with 'what ought to be'. Normative economics is an analysis based on value judgements. Normative statements express an individual or collective opinion on a subject and cannot be proved by facts to be true or false. Certain words or phrases tell us clearly that we have entered the realm of normative economics. These include the words *good*, *bad*, *need*, *should* and *ought*.

Normative economics
An analysis based on value judgements.

The point here is that different people interpret the same facts differently. Each of us has individual subjective preferences that we apply to a particular topic. An economist trained in the US may argue that Asian nations *should* adopt western values and institutions. Or one member of parliament argues, 'We *ought to* see that every teenager who wants a job has one'. Another counters by saying, 'Keeping inflation under control is *more important* than teenage unemployment'.

When considering a debate, make sure you separate the arguments into their positive and normative components. This distinction allows you to determine if you are choosing a course of action related to factual evidence or to value judgement. The material presented in this textbook, like most of economics, takes pains to stay within the boundaries of positive economic analysis. In our everyday lives, however, politicians, business executives, relatives and friends often use normative statements when discussing economic issues. Economists also may associate themselves with a political position and use normative arguments for or against some economic policy. When using value judgements, an economist's normative judgements may have no greater validity than those of others. As is the case with all human beings, an economist's own personal values or preconceptions can influence his or her thinking about many things, ranging from deficit spending to whether there should be a tax on carbon emissions.

Analyse the issue

The myriad ways in which economists disagree

Fairfax Syndication/Steven Siewert

Ross Gittins

It is common for economists to comment on and undertake research into matters that have an ethical dimension. So, for example, they often have views about the ethical aspects of issues, such as the setting of minimum wages, access to welfare payments or environmentally sensitive development proposals. As we have seen, economics as a discipline has developed a neat way of distinguishing between these ethical aspects of important issues and the factual aspects of these issues. This involves use of the distinction between their normative and positive characteristics.

An example of this distinction can be found in the discussion of the need of a tax reform, and merits of increasing the goods and services tax (GST) to restore fiscal balance. In response to the economic downturn during the COVID-19 pandemic, the federal government sharply increased financial support for affected businesses and their workers to keep them in work. The unemployment benefits were also boosted to stimulate spending. While this had prevented the economy slipping further into recession, the 2020–21 underlying budget deficit would balloon to $214 billion.[1] How to return the federal budget to balance has been a hotly debated topic.

The pandemic presented not only a need but also an opportunity to reform the tax system in Australia. The NSW Treasurer proposed that the tax base be broadened and the GST be increased from 10 to 15 per cent, replacing some outdated state taxes (e.g. payroll tax and stamp duties). Many economists and economic think tanks (e.g., the Grattan Institute) support such an increase because their research shows that such a tax is hard to avoid and is thus likely to be a more efficient way of collecting government revenue. It is also likely to involve less of a disincentive to work and investment than the alternative of using income taxes. Economists also understand that because poorer consumers use a bigger proportion of their income for consumption than do richer consumers, the tax falls more heavily on the poor in percentage terms. These are some of the factual, or *positive*, aspects of the GST.

On the other hand, economists will often have opinions about the desirability of increasing the GST if it adversely affects the poor, creating more income inequality. This is one of the *normative* aspects of the GST.

While there will generally be agreement among economists about the positive aspects of increasing the GST, there can be very marked divergences of opinion about its normative implications. And this is true of many other issues that economists comment upon. Whether it be a decision to develop a new port, change the health system or alter the welfare payments regime, there will generally be agreement on the positive effects of such proposals. When it comes to the desirability of these proposals, however, there can be a great divergence of views among economists. As you study more economics and listen to the views of different economists – including those of your lecturers – this will become more and more apparent.

Another dimension of disagreement between economists that has lately been more widely discussed is the concern that economists will even disagree about the positive aspects of an economic proposal. This concern arises from the fact that economists are often paid to analyse issues in the knowledge that their findings will have a significant effect on the fortunes of their client – whoever pays the piper calls the tune. This tendency is, for example, particularly important when it comes to development proposals where developers are required to use economic consultants to determine the extent of potential adverse effects of their proposal. As *Sydney Morning Herald* economics commentator Ross Gittins has put it:

> For some reason, when the independent [economic] consultants run their economic models they invariably reach conclusions that support their paying customer's proposal. Remarkable.[2]

In response to the tendency for objectivity to be lost in striving to please the client, more and more economists and their professional organisations are calling for a code of ethics for economists. Indeed, when such codes have for decades been an integral part of practice for professions as diverse as the law, accounting and medicine, it is surprising that it has taken so long for economists to see the light.

What do you think?

Using the information provided, answer the following questions:

1. Explain the difference between positive and normative analysis. Which of these is the stock in trade of economists?
2. If a tribunal is considering whether a development proposal should be allowed, should it give more weight to the findings of an analysis of the proposal by government economists or the findings of private sector economists employed by the developer?
3. Explain your own position on the issue of economists being more ethically responsible. Identify positive and normative reasons for your decision. How could economists try to ensure that members of their profession behave ethically?

[1] Australian Government, 'Budget overview 2020', at https://budget.gov.au/2020-21/content/download/glossy_overview.pdf

[2] R. Gittins, 'Economists face flak over ethics', *Sydney Morning Herald*, 23 June 2014, p. 26, at https://www.smh.com.au/business/economists-facing-flak-over-ethics-20140622-3am9r.html, accessed 15 December 2017.

In summary

- **Behavioural economics** is a branch of economics that combine economics and psychology in its study of human behaviour.
- **Experimental economics** is a branch of economics in which experiments designed to mimic real-world scenarios are used to explore aspects of human behaviour.
- **Positive economics** deals with statements can be proved true or false by appealing to objective facts.
- **Normative economics** is based on value judgements or opinions.

6 Practical applications of economics

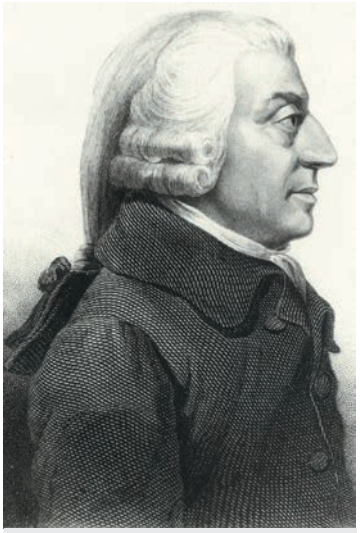
Now that you've seen what economics is about, it's time to have an overview of the practical ways in which economics is used in contemporary society.

Today, there is hardly a human endeavour that is not subject to the calculus of economics. Thus, when new infrastructure projects are planned or new government policy measures are put in place, an economic analysis of costs and benefits is usually undertaken. Most environmental impact studies have their foundation in economic analysis of the advantages and disadvantages of the proposed project. Similarly, health and social welfare expenditure is increasingly being directed to where the economic benefits are greatest. Furthermore, as you will see in later chapters, economic analysis of costs and revenues that has always been part and parcel of big business is increasingly being used to inform decision-making by small- and medium-sized enterprises.

Does economics promote greed?

Economics and ethics

As you study this book you will discover that the economic model is based on the assumption that economic agents pursue their own self-interest. This self-interest is said to be enlightened (i.e., **enlightened self-interest**) in that individuals pursue their own ends subject to the set of laws and social mores that prevails in society. Critics of economics sometimes argue, however, that the economic model promotes unenlightened self-interest – sometimes described as greed. Indeed, some individuals have used the notion of economic self-interest to justify their own greed – as did Gordon Gekko, played by actor Michael Douglas, when he famously declared that 'greed is good' in the film *Wall Street*. With the onset of the global financial and economic crisis (GFEC), greed again reared its ugly head when a number of individuals and organisations were accused of unethical behaviour and some were even convicted for criminal activities.



Alamy Stock Photo/Pictorial Press Ltd

Adam Smith

In spite of some opportunistic interpretations of the relevance of the assumption of self-interested behaviour, the mainstream economic model does not set about promoting greed. It is concerned to promote the efficient working of free markets subject to conformity to an underlying set of ethical or moral principles – a set of principles that is not to be traded for increased material welfare. The founder of modern economics, Adam Smith, pointed out over two centuries ago that the gaining of economic advantage by nonconformity to these principles does not promote the economic welfare of the community. Management guru Charles Handy, in *The Empty Raincoat*, sums up the case by referring to Smith's 18th-century writings:

> Adam Smith, who was professor of moral philosophy not of economics, built his theories on the basis of a moral community. Before he wrote *A Theory of the Wealth of Nations* he had written his definitive work – *The Theory of Moral Sentiments* – arguing that a stable society was based on 'sympathy', a moral duty to have regard for your fellow human beings. The market is a mechanism for sorting the efficient from the inefficient, it is not a substitute for responsibility.[1]

Determination of the set of ethical principles that underpin capitalist society is, of course, a normative exercise. The appropriateness of these principles cannot be tested objectively and often changes when the society changes. Nonetheless, these principles evolve and constitute the foundation for today's successful capitalist economies.

[1] C. Handy, *The Empty Raincoat*, Hutchinson, London, 1994, p. 15.

Each of the above examples involves microeconomics. At the macroeconomic level, the extent to which economics shapes our everyday lives is even more obvious. The GFEC and its aftermath made this abundantly clear. More recently, in relation to the Australian government's stimulus response to the COVID-19 pandemic in the 2020 budget, business and academic economists debated on the pros and cons of different aspects of the stimulus spending. Watching the nightly television news or taking a quick glance at a newspaper reveals the importance placed on economists' views of the state of the national and world economies, the rate of unemployment, share prices and whether interest rates are likely to rise or fall. The economists who comment on these issues are typically employed by large financial institutions, consulting firms or as academics in universities.

Enlightened self-interest This involves members of the community respecting the laws and social mores of society while they pursue their own individual goals.

Not only can practical applications of economics be observed in daily events that unfold around us but they can also be seen in the ways in which many leading professions use economic principles as part of their toolkits. Whether it be accountants, valuers, finance personnel, marketers or engineers, you can be sure that economic principles have an important role to play in the services they provide. The rest of this book is devoted to the study and practical applications of these economic principles.

In summary

- Practical applications of economics occur in the context of enlightened self-interest, which involves members of the community respecting the laws and social mores of society while they pursue their own individual goals.

Key concepts

Scarcity
Factors of production
Land
Labour
Entrepreneurship
Capital
Economics
Microeconomics
Macroeconomics
Model
Ceteris paribus
Behavioural economics
Experimental economics
Positive economics
Normative economics
Enlightened self-interest

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand why the problem of scarcity is central to economics

Economics is the study of how individuals and society choose to allocate scarce resources in order to satisfy unlimited wants. **Scarcity** is the fundamental economic problem that human wants exceed the availability of resources. **Resources** are factors of production, which are classified as **land**, **labour**, and **capital**. **Entrepreneurship** is a special type of labour. An entrepreneur combines resources to produce products and take risk in running a new business or creating a new product.

2. Explore the differences between macroeconomics and microeconomics

Macroeconomics applies an economy-wide perspective that focuses on such issues as inflation, unemployment, the growth rate of the economy and international trade. It studies the overall performance of an economy. **Microeconomics** examines individual decision-making units. It studies individual markets for goods, services and resources.

3. Discover the methodology of economics

Models are simplified descriptions of reality used to understand and predict economic events. A good model must be testable by examining factual evidence.

4. Understand the hazards of the economic way of thinking

Ceteris paribus means that all other factors that might affect a particular relationship remain unchanged. Reasoning will be unsound if this assumption is violated. Another reasoning pitfall is to think that *association* means *causation*.

5. Understand why economists disagree

Positive economics deals with statements that are testable by appealing to facts, while **normative economics** is based on value judgements or opinions. Although economists generally are in agreement with respect to factual (or positive) economic

statements, they can have different views and value judgements in relation to desirability of certain market outcomes or public policies (normative economics). Some of these differences arise out of their different assumptions about human nature, which is more rigorously investigated in behavioural economics and experimental economics research.

Behavioural economics is a branch of economics in which more comprehensive assumptions about human behaviour are employed by combining psychology with economics. **Experimental economics** is a branch of economics in which experiments designed to mimic real-world scenarios are used to explore aspects of human behaviour.

Study questions and problems

1 Does scarcity affect all individuals and all nations? Why or why not? Would monks, who often choose to live the most austere lives, be free from the problem of scarcity? What about Graeme Wood, founder of Wotif, who gives large amounts of money away to various causes including environmental groups?

2 Are computer software programs an example of land, labour or capital? Explain your answer.

3 What is financial capital? Is financial capital (e.g., shares and bonds) included in the economist's definition of capital? Why or why not?

4 Explain the difference between macroeconomics and microeconomics. Give examples of some areas investigated by each of these branches of economics.

5 What are three normative issues of interest in each of microeconomics and macroeconomics?

6 Discuss some aspects of human behaviour that you think might be chosen by economists as the subject of an exercise in experimental economics.

7 Discuss the way in which behavioural economics might take a different approach to remuneration of nurses from that taken by mainstream economics.

8 Analyse the positive versus normative arguments in the following case. Which are the positive statements used and which are normative?

Should the wearing of flotation devices be compulsory on all pleasure craft?

Advocates of the compulsory wearing of flotation devices (i.e., life jackets) on pleasure craft argue that they will save lives and that governments should make wearing them compulsory on *all* pleasure craft – big or small. A sound flotation device can be bought for less than $100. Opponents argue that, because flotation devices give a false sense of security, wearing them encourages riskier behaviour on the part of boat owners. They therefore believe the government should leave the decision about whether to wear a flotation device to individuals. They say the role of the government should be limited to providing information on the risks of wearing or not wearing a flotation device.

Answer to 'You're the economist'

What is the price of economic freedom in Cuba or China?

Just as Eastern Europeans risked their lives to escape to the west before the fall of the Berlin Wall, so Cubans have, for generations, endeavoured to flee Cuba for a better life elsewhere – especially in the US. Although their quest is for both political and economic freedom, there is no doubt that the risks to their lives that these people are willing to take are a measure of the value they place on the economic freedom that they seek in capitalist countries. Although the Chinese situation is somewhat different, we do observe wealthy Chinese citizens who are endeavouring to find a home in free-market economies where they often buy real estate and arrange for the education of their children. If you said that the scarcity of economic (and political) freedom that residents of Cuba and China experience is indicated by their willingness to take risks and forgo contact with friends and family to gain economic freedom in another country, YOU ARE THNKING LIKE AN ECONOMIST.

Multiple-choice questions

1 Which of the following is NOT a scarce resource?
- **a** Clean water
- **b** Pollution
- **d** Petrol
- **d** iPhones.

2 Which of the following best reflects the economist's view of the pursuit of individual self-interest in society?
- **a** Governments and families both should endeavour to minimise it.
- **b** In order to maximise social welfare, pursuit of self-interest should be unconstrained.
- **c** Individual self-interest should be constrained by the social mores of society.
- **d** Pursuit of self-interest is bad since it is the same as greed.

3 Which of the following would economists be most likely to disagree about?
- **a** The figure for the current inflation rate.
- **b** The rise of unemployment rate during pandemic.
- **c** Whether or not free public transport should be available to university students.
- **d** The effect of home mortgage rates on purchase of residential housing.

4 Which of the following are microeconomic issues? Which are macroeconomic issues?
- **a** How will a decrease in the price of tattoo removal affect the quantity of tattoos applied?
- **b** What will cause the rate of growth in consumer spending to fall?
- **c** How does a tax on the sale of houses affect the quantity of houses sold?
- **d** Would a move by the G20 to raise infrastructure spending in member countries help promote economic growth?

5 Behavioural economics is:
- **a** concerned with whether an individual's behaviour is bad or good.
- **b** about the role of enlightened self-interest in the economy.
- **c** a branch of macroeconomics.
- **d** involved with the refinement of assumptions about how humans behave.

6 'The government should collect higher taxes from the rich and use the additional revenues to provide greater benefits to the poor.' This statement is an illustration of a:
- **a** testable statement.
- **b** basic principle of economics.
- **c** statement of normative economics.
- **d** statement of positive economics.

7 When it comes to positive economics, economists could disagree because:
- **a** some economists are men and others are women.
- **b** they are being paid by different parties who are hoping for different outcomes from the analysis they have commissioned.
- **c** the assumptions of their models differ.
- **d** both b and c are correct.

8 Which of the following is a statement of positive economics?
- **a** Poverty is more harmful than inflation.
- **b** An unemployment rate of 7 per cent is a serious social problem.
- **c** If the overall unemployment rate is 7 per cent, the youth unemployment rate will be around 20 per cent.
- **d** It is fair that there is more income equality by using taxation and transfer (i.e., increasing income taxes and social benefits).

APPENDIX 1

Applying graphs to economics

Learning objectives

In this appendix, you will be exploring the following key ideas:

1 Distinguish between direct and inverse relationships.
2 Interpret the slopes of lines in a graph.
3 Discover how to introduce a third variable to the graph.
4 Understand the importance of index numbers.

Remember the old adage that 'a picture is worth a thousand words'? Economists do and they are well known for their use of graphs. Graphs are used throughout this text to present economic models. By drawing a line in a designated space, you can create a two-dimensional illustration to analyse the effects of a change in one variable on another. You could describe the same information using other devices, such as verbal statements, tables or equations. But the graph provides one of the simplest ways to present and understand relationships between economic variables.

Don't worry if graphs are unfamiliar to you. This appendix explains all the basic graphical language you will need for the economic analysis in this text.

1 Direct and inverse relationships

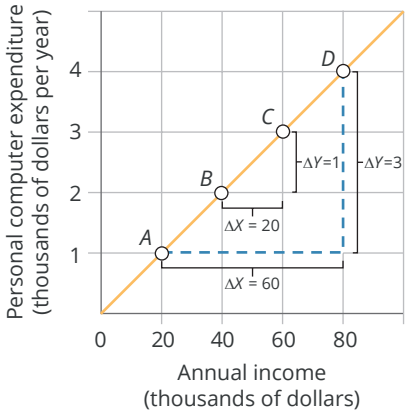
Relationships between two variables can be such that they move in the same direction (direct) or in opposite directions (inverse). We look first at direct relationships.

A direct relationship

Basic economic analysis typically concerns the relationship between two variables, both of which have positive values. Thus, we can confine our graphs to the upper right-hand (north-east) quadrant of the coordinate number system. In **Exhibit A1.1**, notice that the scales on the horizontal axis (x-axis) and the vertical axis (y-axis) do not measure the same variables.

Exhibit A1.1 A direct relationship between variables

Expenditure for a personal computer at different annual incomes		
Point	Personal computer expenditure (thousands of dollars per year)	Annual income (thousands of dollars)
A	$1	$20
B	2	40
C	3	60
D	4	80



The line with a positive slope shows that the expenditure per year on a personal computer has a direct relationship to annual income, *ceteris paribus*. As annual income increases along the horizontal axis, the amount spent on a personal computer also increases, as measured by the vertical axis. Along the line, each 20-unit increase in annual income results in a 1-unit increase in expenditure on a personal computer. Because the slope is constant along a straight line, we can measure the same slope between any two points. Between points B and C or between points A and D, the slope is ΔY/ΔX = +3/+60 = 1/20.

The horizontal axis in **Exhibit A1.1** measures annual income, and the vertical axis shows the amount spent per year on a personal computer (PC). The intersection of the horizontal and the vertical axes is the *origin* and the point where both income and expenditure are zero. In **Exhibit A1.1**, each point is a coordinate that matches the dollar value of income and the corresponding expenditure on a PC. For example, point *A* on the graph shows that people with an annual income of $20 000 spent $1000 per year on a PC. Other incomes are associated with different expenditure levels. For example, at $60 000 per year (point *C*), $3000 will be the annual amount spent on a PC.

Direct relationship A positive association between two variables. When one variable increases, the other variable increases, and when one variable decreases, the other variable decreases.

The straight line in **Exhibit A1.1** allows us to determine the direction of change in PC expenditure as annual income changes. This relationship is *positive* because PC expenditure, measured along the vertical axis, and annual income, measured along the horizontal axis, move in the same direction. PC expenditure increases as annual income increases. As income declines, so does the amount spent on a personal computer. Thus, the straight line representing the relationship between income and PC expenditure is a **direct relationship**. A direct relationship is a positive association between two variables. When one variable increases, the other variable increases, and when one variable decreases, the other variable decreases. In short, both variables change in the *same* direction.

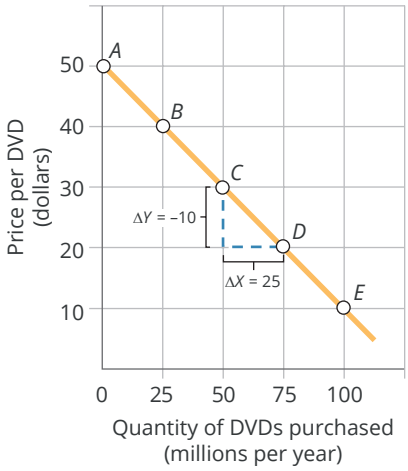
Finally, a two-variable graph, like any model, isolates the relationship between two variables and holds all other variables constant under the *ceteris paribus* assumption. In **Exhibit A1.1**, for example, other possible causal factors, such as the prices of PCs and the education level of the individual, are held constant by assumption. In **Chapter 3**, you will learn how to deal with changes in these variables as well.

An inverse relationship

Now consider the relationship between the price of DVDs and the quantity consumers will buy per year, shown in **Exhibit A1.2**. These data indicate a *negative* relationship between the price variable and the quantity variable. When the price is low, consumers purchase a greater quantity of DVDs than when the price is high.

Exhibit A1.2 An inverse relationship between variables

Quantity of DVDs consumers purchase at different prices		
Point	Price per DVD	Quantity of DVDs purchased (millions per year)
A	$50	0
B	40	25
C	30	50
D	20	75
E	10	100



The line with a negative slope shows an inverse relationship between the price per DVD and the quantity of DVDs consumers purchase, *ceteris paribus*. As the price of a DVD rises, the quantity of DVDs purchased falls. A lower price for DVDs is associated with more DVDs purchased by consumers. Along the line, with each $10 decrease in the price of DVDs, consumers increase the quantity purchased by 25 units. The slope = ΔY/ΔX = –10/+25 = –1/2.5.

In **Exhibit A1.2** there is an **inverse relationship** between the price per DVD and the quantity consumers buy. An inverse relationship is a negative association between two variables. When one variable increases, the other variable decreases, and when one variable decreases, the other variable increases. Stated simply, both variables move in *opposite* directions. Again, we are dealing with only two variables, holding constant all other causal factors, such as consumer income and sellers' expenditures on advertising DVDs.

Inverse relationship A negative association between two variables. When one variable increases, the other variable decreases, and when one variable decreases, the other variable increases.

The line drawn in **Exhibit A1.2** is an inverse relationship. By long-established tradition, economists put price on the vertical axis and quantity on the horizontal axis. In **Chapter 3**, we will study in more detail the relationship between price and quantity demanded, which is called the *law of demand*.

Finally, what does the *intercept* at point *A* in the exhibit mean? The intercept in this case means that at a price of $50 no consumer is willing to buy a single DVD.

In summary

- A **direct relationship** is one in which two variables change in the *same* direction.
- An **inverse relationship** is one in which two variables change in *opposite* directions.

2 Slopes of lines in a graph

Plotting numbers gives a clear visual expression of the relationship between two variables, but it is also important to know how much one variable changes as another variable changes. This can be determined by calculating the slope.

The slope of a straight line

Slope
The ratio of the changes in the variable on the vertical axis (the rise or fall) to the change in the variable on the horizontal axis (the run).

The **slope** of a line is the ratio of the change in the variable on the vertical axis (i.e., the rise or fall) to the change in the variable on the horizontal axis (i.e., the run). Algebraically, if Y is on the vertical axis and X on the horizontal axis, the slope is expressed as follows (the delta symbol, Δ, means 'change in'):

$$\text{slope} = \frac{\text{rise}}{\text{run}} = \frac{\text{change in vertical axis}}{\text{change in horizontal axis}} = \frac{Y}{X}$$

Consider the slope between points B and C in **Exhibit A1.1**. The change in expenditure on a PC, Y, is equal to +1 (from $2000 up to $3000 per year) and the change in annual income, X, is equal to +20 (from $40 000 up to $60 000 per year). The slope is therefore +1/+20. The sign is positive because computer expenditure is directly or positively related to annual income. The steeper the line, the greater the slope because the ratio of ΔY to ΔX rises. Conversely, the flatter the line, the smaller the slope. **Exhibit A1.1** also illustrates that the slope of a straight line is constant; that is, the slope between any two points along the line, such as between points A and D, is equal to +3/+60 = 1/20.

What does the slope of 1/20 mean? It means that a $1000 increase (decrease) in PC expenditure each year occurs for each $20 000 increase (decrease) in annual income. The line plotted in **Exhibit A1.1** has a *positive slope*, and we describe the line as 'upward-sloping'.

On the other hand, the line in **Exhibit A1.2** has a *negative slope*. The change in Y between points C and D is equal to –10 (from $30 down to $20), and the change in X is equal to 25 (from 50 million up to 75 million DVDs purchased per year). The slope is therefore –10/+25 = –1/2.5, and this line is described as 'downward-sloping'.

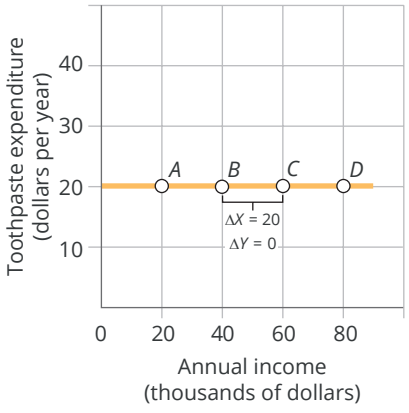
What does this slope of –1/2.5 mean? It means that raising (lowering) the price per DVD by $1 decreases (increases) the quantity of DVDs purchased by 2.5 million per year.

Suppose we calculate the slope between any two points – say, points B and C in **Exhibit A1.3**. In this case, there is no change in Y (expenditure for toothpaste) as X (annual income) increases. Consumers spend $20 per year on toothpaste regardless of annual income. It follows that $\Delta Y = 0$ for any ΔX, so the slope is equal to 0. When the relationship between two variables is indicated by a horizontal line (or a vertical line) there is an **independent relationship**. An independent relationship means there is no association between two variables. When one variable changes, the other variable remains unchanged.

Independent relationship
No association between two variables. When one variable changes, the other variable remains unchanged.

Exhibit A1.3 An independent relationship between variables

Expenditure for toothpaste at different annual incomes		
Point	Toothpaste expenditure (dollars per year)	Annual income (thousands of dollars)
A	$20	$20
B	20	40
C	20	60
D	20	80

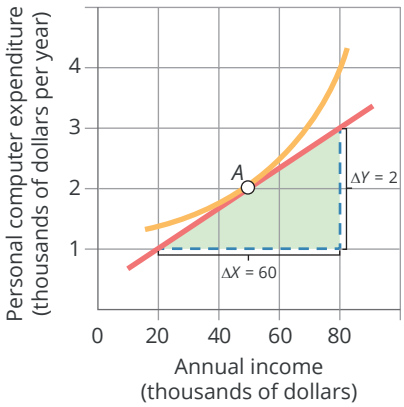


The horizontal line with a zero slope shows that the expenditure per year for toothpaste is unrelated to annual income. As annual income increases along the horizontal axis, the amount spent each year for toothpaste remains unchanged at 20 units. If annual income increases by 20 units, the corresponding change in expenditure is zero. The slope = ΔY/ΔX = 0/+20 = 0.

The slope of a curve

So far, we have looked at straight-line or linear relationships. Now we examine relationships between two variables that result in a curved-line or non-linear relationship. The slope of a curved line changes from one point to another. Suppose the relationship between the expenditure on a PC per year and annual income is not a straight line, but an upward-sloping curved line, as drawn in **Exhibit A1.4**. The slope of the curve is *positive*, but it changes as we move along the curve. To calculate the slope of a given point on the curve requires two steps. For example, at point *A*, the first step is to draw a tangent line that just touches the curve at this point without crossing it. The second step is to determine the slope of the tangent line. In **Exhibit A1.4**, the slope of the tangent line, and therefore the slope of the curve at point *A*, is +2/+60 = 1/30. What does this slope of 1/30 mean? It means that at point *A* there will be a $1000 increase (decrease) in PC expenditure each year resulting from a $30 000 increase (decrease) in annual income.

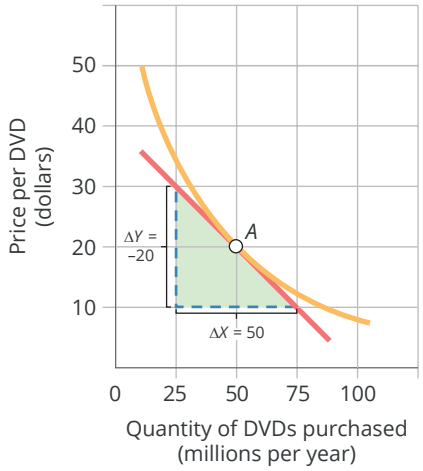
Exhibit A1.4 **The slope of an upward-sloping, non-linear curve**



The slope of a curve at any given point, such as point A, is equal to the slope of the straight line drawn tangent to the curve at that point. The tangent line just touches the curve at point A without crossing it. The slope of the upward-sloping curve at point A is +2/+60 = +1/+30 = 1/30.

Now consider that the relationship between the price per DVD and the quantity demanded by consumers per year is the downward-sloping, non-linear curve shown in **Exhibit A1.5**. In this case, the slope of the curve is *negative,* but again it changes as we move along the curve. To calculate the slope at point *A*, draw a line tangent to the curve at point *A*. Thus, the slope of the curve at point *A* is –20/+50 = –1/2.5.

Exhibit A1.5 **The slope of a downward-sloping, non-linear curve**



In this exhibit, the negative slope changes as one moves from point to point along the curve. The slope at any given point, such as point A, can be determined by the slope of the straight line tangent to that point. The slope of the downward-sloping curve at point A is –20/+50 = –1/+2.5 = –1/2.5.

In summary

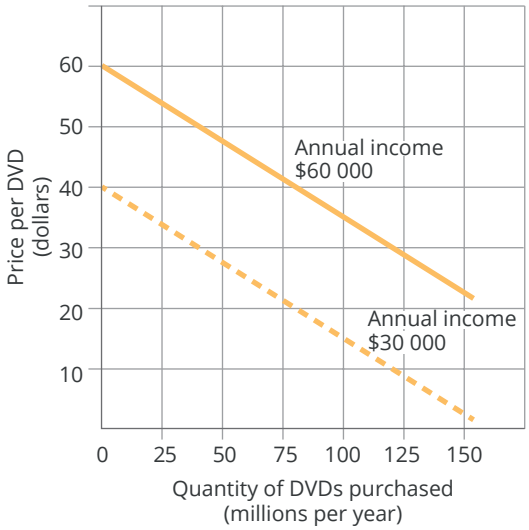
- **Slope** is the ratio of the vertical change (the rise or fall) to the horizontal change (the run). The slope of an *upward-sloping* line is *positive*, and the slope of a *downward-sloping* line is *negative*.

3 Introducing a third variable to the graph

The two-variable relationships drawn so far conform to a two-dimensional flat piece of paper. For example, the vertical axis measures the price per DVD variable, and the horizontal axis measures the quantity of DVDs purchased variable. All other factors, such as consumer income, that may affect the relationship between the price and quantity variables are held constant by the *ceteris paribus* assumption. But reality is frequently not so accommodating. Often, a model drawn on a two-dimensional piece of graph paper must take into account the impact of changes in a third variable (e.g., consumer income).

The method used to depict a three-variable relationship is shown in **Exhibit A1.6**. As explained earlier, the cause-and-effect relationship between price and quantity of DVDs determines the downward-sloping curve. A change in the price per DVD causes a movement downward along either of the two separate curves in **Exhibit A1.6**. As the price falls, consumers increase the quantity of DVDs demanded. The location of each curve on the graph, however, depends on the annual income of consumers. As the annual income variable increases from $30 000 to $60 000 and as consumers choose to purchase more at each possible price, the price–quantity demanded curve shifts rightward. Conversely, as the annual income variable decreases and as consumers choose to buy less at each possible price, the price–quantity demanded curve shifts leftward.

Exhibit A1.6 **Changes in price, quantity and income in two dimensions**



Economists use a multi-curve graph to represent a three-variable relationship in a two-dimensional graph. A decrease in the price per DVD causes a movement downward along each curve. As the annual income of consumers rises, there is a shift rightward in the position of the demand curve.

This is an extremely important concept you must understand: throughout this book, you must distinguish between *movements along* and *shifts in* a curve. Here is how you tell the difference. A change in one of the variables shown on either of the coordinate axes of the graph causes *movement along* a curve. On the other hand, a change in a variable not shown on one of the coordinate axes of the graph causes a *shift in* a curve's position on the graph.

A helpful study hint using graphs

Don't be the student who tries to memorise the graphs and then wonders why he or she failed economics. Instead of memorising graphs, you should use them as a valuable aid to learning the economic concepts they illustrate. After studying a chapter, go back to the graphs one by one. Hide the brief description accompanying each graph and describe to yourself or other students what the graph means. Next, uncover the description and check your interpretation. If you still fail to understand the graph, read the text again and correct the problem before proceeding to the next chapter.

In summary

- **Shifts** in a curve occur when there is a change in a third variable that is not shown on the coordinate axes of the graph – that is, when the *ceteris paribus* assumption is relaxed.
- **Movements** along a curve occur when there is a change in one of the variables shown on the coordinate axes of the graph.

4 The importance of index numbers

Index numbers Indicate changes in the value of a variable over time where the value of the variable in the base (or reference) year is set at 100.

In many areas of the physical and social sciences and particularly in economics, there is a need to make comparisons between the values of variables over time. For example, we may want to compare the overall price level in the economy from year to year, or over a number of years, so that we can adjust for the effects of inflation. Similarly, changes over time in the overall level of consumer expenditure or business investment can help to paint a picture of the way in which the national economy is performing. Undertaking tasks such as these is made so much easier if the relevant values are expressed in the form of **index numbers**.

The practice with index numbers is to choose an initial period – for example, a particular year or month in a year – that is described as the base period and for which the value of the index is designated as 100. As the quantity of the variable being measured changes in subsequent periods, so the value of the index would rise above or fall below the base year level of 100. For example, if there were 500 new homes built in a particular suburb in 1990, and 1990 is our base (or reference) year, we would say that the value of the index of home building was 100 in that year. If, by the year 2000, the number of homes built had fallen to 300 per year, we would calculate the value of the index to be 60 (300/500 × 100) in that year. Alternatively, if the number of homes built had risen to 650, the value of the index would be 130 (650/500 × 100).

Not only is information such as this useful on its own but it is also invaluable for making comparisons. Thus, if we learnt that over the same period the index of home building in a particular suburb grew from 100 to 130, and the index of factory construction grew from 100 to 150, we would know that factory construction was growing more rapidly than home building in that suburb. We would know this without having to know what the absolute levels of home building or factory construction were in each of the relevant years. Similarly, if the index of factory construction in a neighbouring suburb had grown from 100 to 180 over the same period that it grew from 100 to 150 in our particular suburb, we would know that this neighbouring suburb had experienced higher growth in factory construction – regardless of whether this neighbouring suburb had a smaller or larger factory construction sector than the suburb with which it is being compared.

Index numbers – especially as they are applied to changes in the overall price level in the economy over time – are discussed and explained further in the macroeconomics section of this book.

In summary

- **Index numbers** indicate changes in the value of a variable over time where the value of the variable in the base (or reference) year is set at 100.

Key concepts

Direct relationship Inverse relationship Slope Independent relationship Index numbers

Summary

Here we summarise the key ideas we have discussed under each of this appendix's learning objectives.

1. Distinguish between direct and inverse relationships

- A **direct relationship** is one in which two variables change in the *same* direction.
- An **inverse relationship** is one in which two variables change in *opposite* directions.

2. Interpret the slopes of lines in a graph

- **Slope** is the ratio of the vertical change (the rise or fall) to the horizontal change (the run).
- The slope of an *upward-sloping* line is *positive*, and the slope of a *downward-sloping* line is *negative*.

3. Discover how to introduce a third variable to the graph

- Shifts in a curve occur when there is a change in a third variable that is not shown on the coordinate axes of the graph – that is, when the *ceteris paribus* assumption is relaxed.
- Movements along a curve occur when there is a change in one of the variables shown on the coordinate axes of the graph.

4. Understand the importance of index numbers

- **Index numbers** indicate changes in the value of a variable over time where the value of the variable in the base (or reference) year is set at 100.

Study questions and problems

1 Without using specific data, draw a graph for the expected relationship between the following variables, *ceteris paribus*. In each case, state whether the expected relationship is *direct* or *inverse*. Explain one of the factors that would be included in the *ceteris paribus* assumption but which, if it were allowed to change, might affect the expected relationship.

a the number of motor vehicles and the number of motor vehicle accidents
b life expectancy and consumption of methamphetamines
c height above sea level and mean summer temperature
d mean summer temperature and sales of ice creams.

2 Assume a research firm collects survey data that reveal the relationship between the income of consumers and their purchases of overseas holidays, all other factors remaining constant. The research shows that consumers earning $40 000 p.a. purchase one overseas holiday each decade; those earning $80 000 p.a. purchase eight overseas holidays per decade while those earning $120 000 p.a. purchase 10 overseas holidays per decade. Based on these data, describe the relevant relationship between income and the quantity of overseas holidays consumers are willing to purchase, using:

a a verbal statement
b a numerical table
c a graph.

Which device do you prefer and why?

Multiple-choice questions

1 Straight line *CD* in **Exhibit A1.7** shows that:
- **a** decreasing the value of *X* is associated with a decrease in the value of *Y*.
- **b** an increase in the value of *X* is associated with an increase in the value of *Y*.
- **c** there is a direct relationship between *X* and *Y*.
- **d** all of the above are true.

2 In **Exhibit A1.7**, the slope of straight line *CD* is:
- **a** 3.
- **b** –1.
- **c** 1.
- **d** ½.

3 In **Exhibit A1.7**, the slope of straight line *CD* is:
- **a** variable.
- **b** zero.
- **c** negative.
- **d** positive.

4 Straight line *AB* in **Exhibit A1.8** shows that:
- **a** there is an inverse relationship between *X* and *Y*.
- **b** decreasing the value of *X* is associated with an increase in the value of *Y*.
- **c** increasing the value of *X* is associated with a reduction in the value of *Y*.
- **d** all of the above are true.

5 A shift in a curve represents a change in:
- **a** the variable on the horizontal axis.
- **b** the variable on the vertical axis.
- **c** a third variable that is not on either axis.
- **d** none of the above.

6 In **Exhibit A1.8**, the slope of straight line *AB:*
- **a** decreases with increases in *X*.
- **b** increases with increases in *X*.
- **c** increases with decreases in *X*.
- **d** remains constant with changes in *X*.

7 In **Exhibit A1.8**, the slope for straight line *AB* is:
- **a** 1.
- **b** 4.
- **c** –5.
- **d** –1.

Exhibit A1.7 Straight line *CD*

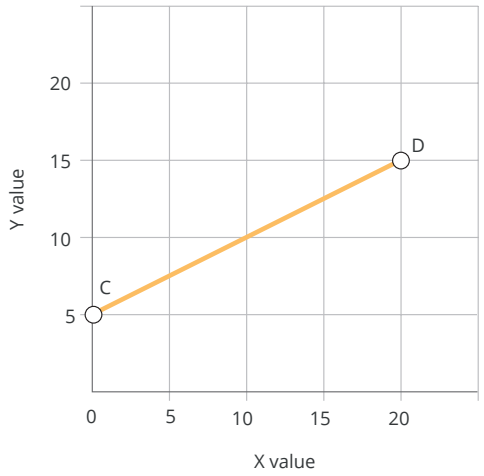
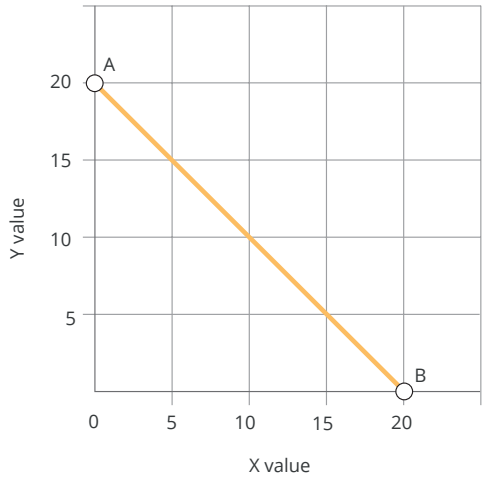


Exhibit A1.8 Straight line *AB*



Production possibilities and opportunity cost

2

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Investigate the three fundamental economic questions.
2 Understand opportunity cost and its central role in economic thinking.
3 Know what marginal analysis is and how it is applied.
4 Describe and graphically present the production possibilities frontier.
5 Identify the factors shifting the production possibilities frontier.
6 Examine the impact of present investment on the future production possibilities frontier.
7 Understand the notion of gains from trade.

Shutterstock.com/g-stockstudio

This chapter begins by examining the three basic economic questions faced by society: *what*, *how* and *for whom to produce*? To fully understand human behaviours in their attempt to address these questions, we will introduce two other key building blocks in the economic way of thinking – *opportunity cost* and *marginal analysis*. For example, will you think that during time of high unemployment it would be easier for businesses to hire staff? We will discuss this and other related questions in this Chapter. Once you understand these important concepts, it will be easier to interpret our first formal economic model, the *production possibilities frontier*. This model illustrates many of the most important concepts in economics, including scarcity, opportunity cost, efficiency, investment and economic growth. The chapter includes a section dealing with the vexed question of whether it is a good thing to decrease the number of foreign workers issued with visas allowing them to work in Australia.

1 The three fundamental economic questions

Whether rich or poor, every nation must answer these same three fundamental economic questions: (1) *What* goods and services will be produced? (2) *How* will they be produced? (3) *For whom* will they be produced?

What to produce?

Should the government allocate its funds to producing more schools and fewer hospitals? Should society devote its limited resources producing more mobile phones and fewer books? Should more bicycles and fewer cars be produced, or should more bus and train services be provided instead of cars? The problem of scarcity imposes a restriction on the ability to produce everything we want during a given period, so the choice to produce 'more' of a good requires producing 'less' of another good. By and large, in a free market economy, *what* to produce is determined on the basis of individual (or family) preferences expressed by consumers in the marketplace, and on the basis of collective preferences expressed through governments.

How to produce?

After deciding which goods and services to produce, the second question for society to decide is how to mix technology and scarce resources in order to produce them. For instance, a scarf can be sewn primarily by hand (labour), partially by hand and partially by machine (labour and capital), or primarily by machine (capital). In short, the *how* question asks whether a production technique will be more or less capital-intensive.

The producers answer this question by consider the relative availability and relative costs of various factors of production. If labour is in abundant supply and cheap, firms will tend to choose a method that is more labour-intensive (or less capital intensive). Education and training also play an important role in answering the *how* question. Education improves the ability of workers to perform their work. The quality and quantity of education and training vary among nations – this is one reason why economies differ in their capacities to apply resources and technology to answer the *how* question. For example, while many Asian countries are still catching up with the US and Australia, which are among the world's biggest users of advanced technology, these Asian countries have an enviable reputation for low-cost manufacturing. Regardless of whether it is advanced technology or low-cost manufacturing that is expanding, an appropriate mix of capital, technology, skilled and unskilled labour is required.

***What, how* and *for whom* questions**
Fundamental questions about the composition of a nation's output, the ways in which this output should be produced and the way in which the benefits of this output should be distributed among the nation's citizens.

For whom to produce?

Once the ***what*** and ***how*** questions are resolved, the third question is ***for whom***. Of all the people who desire to consume the goods and services produced, who actually receives them? Who dines out at high-end restaurants, and who patronises fast-food outlets? Who travels business class, and who travels economy? Who lives in houses with water views? In a market economy, this will depend largely on the way in which incomes are distributed among different members of the community. In the course of economic growth, a country will enjoy a higher living standard. However, the benefits of growth are likely to be unevenly distributed, creating haves and have-nots. Therefore, the *for whom* question also means that society must ask whether governments should override market outcomes to achieve more income equality in society and to ensure that the social welfare system guarantees a minimum standard of living for all.

In summary

- Three fundamental economic questions facing any economy are ***what*** to produce, ***how*** to produce it and ***for whom*** to produce it.
- In a market economy, the price system is commonly used to address these questions.
- Regardless of its political system, the government often comes in to override the market to determine the answers to some of these economic questions.

2 Opportunity cost

Because of scarcity of resources, the three basic economic questions cannot be fully answered without considering alternative uses of a resource. Consider a common situation facing every consumer. You have $100 to spend in a weekend and there are a few options; for example, attending a concert, getting a tattoo, going to a football match or keeping the money. Since you cannot use the money to do all of the above, you have to make a *choice*. Suppose the concert is your best alternative, and getting a tattoo done is the next best alternative. When you choose the concert, you give up the opportunity to buy time at a tattoo studio. When you make a choice, the next best alternative that you give up is called the **opportunity cost** in economics. In the above scenario the opportunity cost of purchasing the concert ticket is the tattoo you have to give up.

Opportunity cost The best alternative sacrificed for a chosen alternative.

Examples of opportunity cost are endless, but let's consider a few. Suppose you have a passion for smashed avocado on toast purchased regularly at the local cafe. On what else could you have spent your hard-earned cash? Well, according to *The Betoota Advocate*, millennial Toby Campbell, who has given up smashed avocado on toast, is now able to afford an iPhone and will soon be able to fulfil the Australian dream of owning his own home! Although Toby is aware that the opportunity cost of his newly acquired phone and the soon-to-be-purchased home is his favourite snack, he's willing to forgo the avocado on toast for his preferred alternative – a phone and a home.[1] Opportunity cost also applies to national economic decisions. Suppose the federal government decides to spend more tax revenues on discouraging asylum seekers from trying to come to Australia. The opportunity cost depends on the next best program *not* funded. Assume new roads and hospitals are the highest-valued projects not undertaken as a result of the decision to increase expenditure on discouraging asylum seekers. Then new roads and hospitals are the opportunity cost of the decision to devote resources to the asylum seeker issue.

If we ask an ordinary person what the term *cost* means. The common response would be to say that the purchase price is the cost. A weekend concert ticket *costs* $100, or a shirt *costs* $50. While this answer is not wrong, it does not tell the whole story. Let us look at the concept of opportunity cost more closely. Consider tertiary education in Australia. If you attended university in the early 1980s, you paid no tuition fees – the price was zero. Does that mean the cost of attending university was zero? Think about the alternative uses of the time you spent on university. If earning income was your best alternative of studying, then the cost of your university education was the income you sacrifice. Even though the tuition fee was zero, the cost of university education was not zero. Free university education is not really *free*.

The concept of opportunity cost can be applied to understanding many human behaviours. In mid-2020, due to the pandemic-induced recession, the unemployment rate rose to 10 per cent. However, during this period it was reported that businesses were struggling to find staff.[2] Why? To prevent the economic downturn from getting worse, the federal government doubled the JobSeeker Payment (i.e., unemployment benefit). The opportunity cost of working – the unemployment benefit and leisure forgone – was very high and staying unemployed was the best option for the many unemployed.

Scarcity, choice and opportunity cost are the unifying theme in economics. As mentioned, time is a scarce resource to most of us. Knowing the relationship between time and opportunity cost also helps understand decision-making by young school leavers during the mining boom in the first decade of this century. During this time, unskilled and semi-skilled workers in the mining industry could earn much more than recent university

---

1 Toby Campbell is a presumably fictional character written about in the spoof newspaper *The Betoota Advocate.* Read Clancy Overell's story about Toby, and the origins of the idea that foregoing smashed avocado on toast is the key to housing affordability, in C. Overell, 'Local millennial experiencing unimaginable financial security after quitting smashed av', *The Betoota Advocate*, 21 October 2016, at https://www.betootaadvocate.com/breaking-news/local-millennial-experiencing-unimaginable-financial-security-after-quitting-smashed-av, accessed 8 March 2021.

2 'Jobless opt for dole over employment', *The Australian*, 30 June 2020, https://www.theaustralian.com.au/nation/jobless-opt-for-dole-as-businesses-struggle-to-find-workers-despite-unemployment-surge/news-story/0ebe5e226775055acdfa90a1bd88751b

graduates. This encouraged some school leavers to work for mining companies rather than study at university. The opportunity cost of study, in terms of income forgone, was therefore very high.

You're the economist

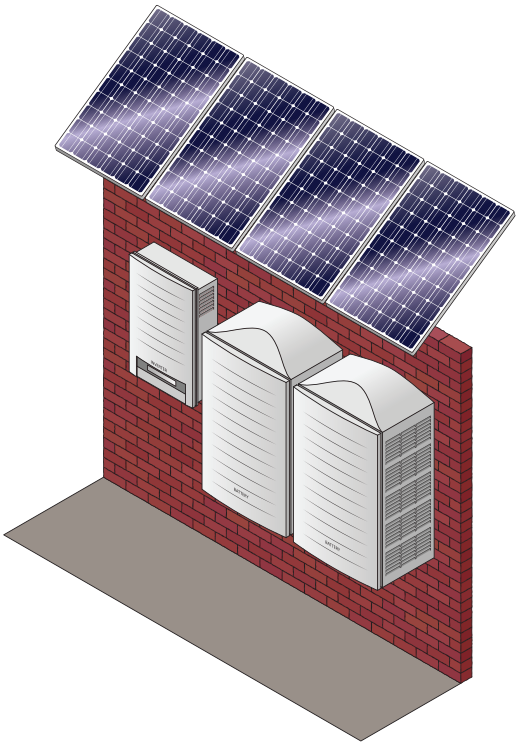
Are consumers and governments always conscious of opportunity cost?

In recent years, public officials and elected representatives in cities and towns across Australia have leapt at the opportunity to reduce electors' power bills while at the same time helping them to clean up the environment. They have done this by subsidising the cost of special equipment that performs these tasks. This situation has been brought about by rising energy prices and increasing concern about the carbon emissions caused by the use of fossil fuels.

The subsidised equipment that reduces consumers' power bills generally does so by either reducing energy consumption directly or by providing opportunities for householders to generate their own power at home. In some cases energy produced in the home in excess of the amount required at the time can be sold to energy supply companies. The solar hot water system is a good example of a device that reduces energy consumption directly. On the other hand, banks of solar panels, usually installed on the roof of the home, are the most common way for householders to make their own power.

In assessing these opportunities for householders to obtain subsidies for energy-saving and generating equipment, both governments and consumers have a tendency to forget the concept of opportunity cost. Let us look specifically at the case of solar panels. We will assume, for the sake of simplicity, that they reduce householders' power bills but do not allow the sale of home-generated electricity to the grid. We also assume that the overall cost of a particular installation is $20 000 with the consumer paying $10 000 and the government subsidy covering the remaining $10 000.

While consumers are likely to be very conscious of the opportunity cost of their $10 000 contribution in terms of what else could be bought with this money, most of them will not consider the income forgone on this amount as a cost of the project. But the interest forgone on the $10 000 is not an insignificant amount. Deposited in a bank account at 3 per cent it could earn $300 a year; if used to pay down the mortgage on the family home it could save $600 per annum if the home loan interest rate is 6 per cent. Put simply, the opportunity cost of putting $10 000 into a subsidised bank of solar panels includes between $300 and $600 each year in interest costs. Quite a lot of power can be bought from the grid for this amount.



Shutterstock.com/Zern Liew

Similarly, if the government was subsidising the solar panels to the tune of $10 000 there is an opportunity cost of this expenditure that it should consider. This opportunity cost would be discovered by looking at other ways in which the government could have used these funds, such as on increased family payments or on painting householders' roofs white, which would also cut back power consumption. Alternatively, if the government has to borrow funds to pay its solar panel subsidies then it should also count the interest paid on this $10 000 as part of the opportunity cost of this expenditure – an important consideration at a time when many governments around the world are trying desperately to reduce debt created as response to the Global Financial and Economic Crisis (GFEC) and its aftermath.

Increased use of solar panels and sales of electricity generated by them to the grid has recently had an interesting effect on the electricity market that was largely unforeseen: as more and more electricity has been generated by solar panels, so the sale of electricity from large-scale power stations has been decreasing. Increasing viability of battery storage of electricity in the home,

which allows solar power generated during daylight hours to be used at night, has exacerbated this problem. This contrasts with previous decades when there had been ever-increasing demand for power from power stations. These lower sales of power mean that the large overhead costs of power stations are being spread over smaller and smaller amounts of output. In turn this has led to significant increases in per unit power prices to householders. From the point of view of the owners of solar panels, these price increases mean that their decision to 'go solar' and use less power from the grid looks like a good one. On the other hand, governments are copping lots of flak from householders (including those who have installed solar panels but who still purchase some of their power from the grid) who find that they are subject to higher and higher grid prices over time. Solar panel policy has become a real headache for politicians who must be thinking about how they could have used the funds expended on solar panel subsidies in other ways – ways that might have increased their popularity with a bigger group of voters. Put simply, one of the opportunity costs of government expenditure on solar panel subsidies is the opportunity forgone to undertake these expenditures in other areas that might have been electorally more popular. Remember that opportunity cost is the cost of the *next best* alternative forgone.

Suppose governments no longer subsidise solar panel installation. If interest rates rise, are more or fewer people likely to install solar panels? What are some of the opportunity costs that environmentally conscious consumers should take into account when considering whether or not to install electricity-saving devices?

In summary

- Scarcity, choice and cost is the unifying theme in economics.
- **Opportunity cost** of a decision is the best alternative forgone for a chosen option.
- Time is an important resource. To determine the cost of a decision, we must consider not only monetary but also non-monetary cost, for example, time cost.

3 Marginal analysis

In most situations, we seldom make all-or-nothing decisions. For example, in deciding how to use your scarce time, you will usually not decide to spend all your time on a particular activity, for example, watching YouTube 24 hours a day. Instead you ask, 'Should I spend an extra hour on YouTube?' The answer depends on a comparison of the additional benefit obtained from watching YouTube for an extra hour and the benefit from its next best alternative (i.e., its opportunity cost). Economists call this additional benefit the marginal benefit, or more commonly, *marginal utility*. This method of analysis is known as **marginal analysis** and is at the heart of many important decision-making techniques used throughout this text. Marginal analysis examines the effects of additions to or subtractions from a current situation. This is a very valuable tool in the economic-way-of-thinking toolkit because it considers the effects of changes. Let us re-visit the 'how to use your scarce time' problem. Should you devote an extra hour to reading this book, going to a movie, watching YouTube, talking on the phone or sleeping? There are many ways to spend your time. Which option do you choose? The answer depends on marginal analysis. If you decide that the benefit of a higher grade in economics exceeds the benefit of, say, watching YouTube, then you allocate the extra hour to studying economics. Excellent choice!

Marginal analysis
An examination of the effects of additions to or subtractions from a current situation.

Similarly, producers use marginal analysis. The farmer asks if they should add fertiliser to improve the wheat yield. Using marginal analysis, the farmer estimates that the wheat revenue yield will be about $75 per hectare without fertiliser and about $100 per hectare using fertiliser. If the cost of fertiliser is $20 per hectare, marginal analysis tells the farmer to fertilise. The additional fertiliser will increase profit by $5 per hectare because fertilising brings an additional yield of $25 at an added cost of $20. We will examine the producer's problem in detail in **Chapter 7**.

Marginal analysis is an important concept when considering changes in the overall composition of output produced within an economy. For example, as demonstrated in the next section, it is useful to know that an increase in the production of one good can result in the production of less of another, and its implications for the present and future welfare of the society.

In summary

- **Marginal analysis** examines the impact of changes from a current situation. The basic approach is to compare the additional benefits of a change with the additional costs of the change.

4 The production possibilities frontier

Production possibilities frontier
Shows the maximum combinations of two outputs that an economy can produce, given its available resources and technology.

Technology
The body of knowledge applied to how goods and services are produced.

The economic problem of scarcity means that society's capacity to produce combinations of goods is constrained by its limited resources. This condition can be represented in a model called the **production possibilities frontier** (PPF), which shows the maximum combinations of two outputs that an economy can produce, given its available resources and technology. Three basic assumptions underlie the PPF model:

1. *Fixed resources* – the quantities and qualities of all resource inputs remain unchanged during the time period. But the 'rules of the game' do allow an economy to shift any resource from the production of one output to the production of another output. For example, an economy may shift workers from producing consumer goods to producing capital goods. Since the number of workers remains unchanged, this transfer of labour will produce fewer consumer goods and more capital goods.
2. *Fully employed resources* – the economy operates with all its factors of production fully employed without waste or mismanagement.
3. *Technology unchanged* – **technology** is the body of knowledge applied to the way goods and services are produced. Holding existing technology fixed creates limits, or constraints, on the amounts of goods and services an economy can produce.

In modern economies an enormous range of goods and services is produced. Goods, which are tangible, include capital goods sold to the business sector and consumer goods, such as food, clothing and furniture sold to consumers. Services, which are intangible, are sold to both businesses and consumers, such as, advice from accountants and lawyers, a trip to the movie, electricity and education. **Exhibit 2.1** shows a hypothetical economy that has the capacity to manufacture any combination of consumer goods and consumer services per year along its PPF, including points *A*, *B*, *C* and *D*. For example, if this economy uses all its resources to make consumer goods, it can produce a *maximum* of 160 units of consumer goods and zero units of consumer services (combination *A*). Another possibility is for the economy to use all its resources to produce a *maximum* of 100 units of consumer services and zero units of consumer goods (point *D*). Between the extremes of points *A* and *D* lie other production possibilities for combinations of goods and services. If combination *B* is chosen, the economy will produce 140 units of consumer goods and 40 units of consumer services. Another possibility (point *C*) is to produce 80 units of consumer goods and 80 units of consumer services.

At this stage it is important to point out that production is a *flow* of output that occurs over a particular time period – in this case one year. Whenever a variable is expressed as a quantity per time period, we are dealing with a flow concept. By contrast, a variable such as the amount of capital used to produce such flows of output is described as a *stock*. It is simply expressed as a quantity without reference to a relevant time period. For example, we might say that a nation's stock of capital produces an annual flow of output of goods and services.

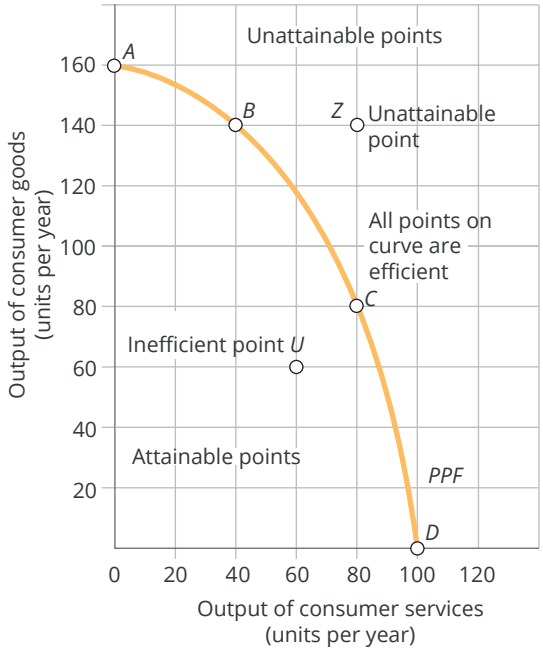
What happens if the economy does not use all its resources to their capacity? For example, some workers may not find work, or plant and equipment may be idle during a recession. The result is that our hypothetical economy fails to reach any of the combinations along the PPF. In **Exhibit 2.1**, point *U* illustrates an *inefficient* output

combination for an economy operating without all of its resources fully employed. At point *U*, our hypothetical economy is producing only 60 units of consumer goods and 60 units of consumer services per year. Such an economy is under-producing because it could satisfy more of society's wants if it were producing at some point along the PPF, say point C, at which more of both consumer goods and consumer services can be produced.

Given its resource constraints, this economy cannot produce beyond its PPF. The society may prefer Point *Z*, for example, enjoying 140 units of consumer goods and 80 units of consumer services, but the economy cannot reach this point with its existing resources and technology. Any point outside the PPF is *unattainable*.

Because all the points along the frontier represent full utilisation of given resources and technology, they are all called *efficient* points. If the economy wishes to produce more of one output, it has to produce less of the other

Exhibit 2.1 **The production possibilities frontier for consumer goods and consumer services**



The production possibilities for consumer goods and consumer services per year

Output (billions of units per year)	Production possibilities			
	A	B	C	D
Consumer goods	160	140	80	0
Consumer services	0	40	80	100

All points along the PPF are maximum possible combinations of consumer goods and consumer services. One possibility, point *A*, would be to produce 160 units of consumer goods and zero units of consumer services each year. At the other extreme, point *D*, an economy uses all its resources to produce 100 units of consumer services and zero units of consumer goods each year. Points *B* and *C* are obtained by using some resources to produce each of the two outputs. If the economy fails to use its resources fully, the result is the inefficient point *U*. Point *Z* lies beyond the economy's present production capabilities and is unattainable.

output. For example, in **Exhibit 2.1**, moving from point *A* to point *B* produces 40 additional units of consumer services per year, but only at a cost of sacrificing 20 units of consumer goods. A movement between any two points on the PPF graphically illustrates the old adage that 'there is no such thing as a free lunch'.

The law of increasing opportunity costs

If you observe the PPF more closely, you will notice it bows out to the right. Why is the PPF shaped the way it is? **Exhibit 2.2** will help us answer this question. It presents a PPF for a hypothetical economy that must choose between producing houses (i.e., goods) and producing university degrees (i.e., the services that comprise university education). This choice has particular contemporary relevance given that Australian governments are committed to increasing participation in tertiary education while at the same time endeavouring to solve the housing affordability crisis by encouraging an increased supply of housing. Looking now at **Exhibit 2.2**; let's consider expanding the production of university degrees in 20 000-unit increments. Moving from point *A* to point *B*, to increase university degrees by 20 000 units, we have to give up 1000 houses. In other words, the *opportunity cost* of 20 000 units of university degrees is 1000 houses. Between point *B* and point *C* the *opportunity cost* is 2000 houses; and between point C and point D the *opportunity cost* is 5000 houses.

Law of increasing opportunity costs
The principle that the opportunity cost increases as the production of one output expands.

Exhibit 2.2 illustrates the **law of increasing opportunity costs**. The law of increasing opportunity costs states that the opportunity cost of producing a good increases as production of this good expands, holding the stock of resources and technology constant (*ceteris paribus*). The law of increasing opportunity costs causes the PPF to display a *bowed-out* shape (concave to the origin).

Why must our hypothetical economy sacrifice larger and larger amounts of output of houses in order to produce each additional batch of 20 000 university degrees? The reason is that the factors of production (i.e., primarily labour and capital) are not equally suited to producing all goods. To expand the output of labour-intensive university degrees requires a large amount of highly educated labour and some suitably equipped buildings and computers, whereas the expansion of output of capital-intensive houses requires a large amount of purpose-built machinery and computers and workers with specialised technical skills.

Suppose our hypothetical economy has a range of capital goods and labour resources, some of which are best suited to the production of houses and some of which are best suited to producing university degrees. Further, suppose that it currently produces no university degrees (point *A*). Now imagine that a decision is made to expand the economy's production of university degrees.

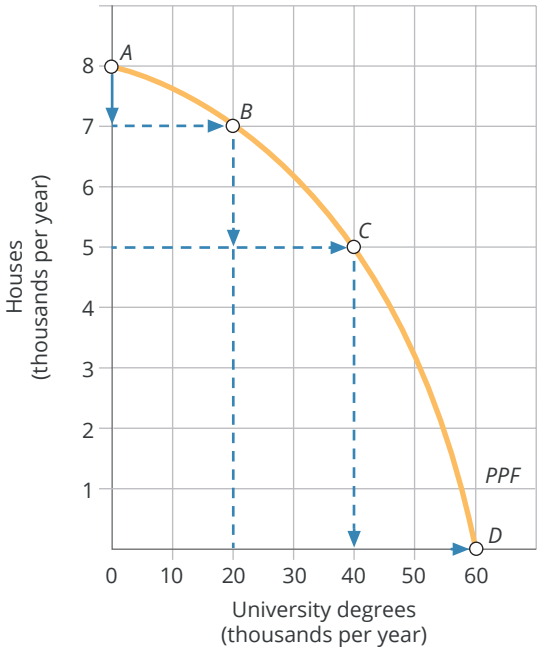
At first this expansion is straight forward by transferring labour and capital resources from construction industry to the university sector to produce 20 000 university degrees. When this occurs, only 1000 houses are sacrificed. At this stage, few houses are sacrificed because the resources that are transferred are those best suited to production of university degrees and least suited for construction. This transfer will have relatively small impact on housing output.

Now suppose the economy wants to further expand university education by moving from point *B* to point *C*. At point B, we have already used up resources most suitable for the education sector. To produce more degrees we have to utilise labour and capital resources that are less suited for the education sector. On the other hand, home builders are now losing labour and capital that are better suited to building houses. **Exhibit 2.2** shows that in order to produce another 20 000 degrees, the economy now has to give up 2000 houses. So, the opportunity cost rises from 1000 to 2000 houses.

Finally, the economy can decide to increase another 20 000 university degrees from 40 000 to 60 000. In the process of moving from point *C* to point *D* the home builders have lost their most productive labour and capital and the output of houses will fall by a lot. We can see the opportunity cost of the last 20 000 degrees rises further to 5000 houses. This example illustrates the law of increasing opportunity costs.

Although this model may seem to you to be somewhat simplistic, it can nonetheless be fruitfully applied to the real world. Let's look at a real situation to illustrate the important idea that labour resources are not perfectly transferable across industries. Over the past 30 years the US has seen a marked decline in the share

Exhibit 2.2 **The law of increasing opportunity costs**



Production possibilities for houses and university degrees per year

Output (thousands per year)	Production possibilities			
	A	B	C	D
Houses	8	7	5	0
University degrees	0	20	40	60

A hypothetical economy produces equal increments of 20 000 university degrees per year as we move from point *A* to point *D* on the PPF. If the hypothetical economy moves from point *A* to point *B*, the opportunity cost of 20 000 university degrees is a reduction in output of houses of 1000 per year. This opportunity cost rises to 2000 houses by selecting point *C* instead of point *B*. Finally, production at point *D*, rather than point *C*, results in an opportunity cost of 5000 houses per year. The opportunity cost rises because factors of production are not equally suited to building houses and conferring university degrees.

of its national output contributed by motor vehicle manufacturing. At the same time there has been a very significant increase in the volume of software produced. For many years now unemployment in the US has been very high in areas where motor vehicle manufacturing is concentrated. In his successful 2016 presidential campaign Donald Trump made a concerted effort to attract the votes of these, and other US 'rust belt' workers. These unemployed Americans, who are unlikely to ever find jobs in software firms, represent a resource that was well suited to the manufacturing of motor cars but is entirely unsuited to software development. See if you can plot these changes using a PPF.

If the factors of production employed in an economy were equally suited to the production of its two hypothetical outputs, then a situation of constant opportunity cost would prevail, and the PPF would be a straight line. We will revisit this model in **Chapter 18** when dealing with international trade and finance.

In summary

- A **production possibilities frontier** illustrates an economy's capacity to produce goods and services, subject to the constraint of scarcity.
- The **law of increasing opportunity costs** explains the *bowed-out* shape of the production possibilities frontier.

5 Shifting the production possibilities frontier

Economic growth The ability of an economy to produce greater levels of output, represented by an outward shift of its production possibilities frontier.

The economy's production capacity is not permanently fixed. If the resource base increases, workers become more educated or technology advances, the economy experiences **economic growth**, and the PPF shifts outward. Economic growth is the ability of an economy to produce greater levels of output. In **Chapter 12**, theories of economic growth are discussed at length. At this stage, however, we will undertake a basic explanation of the factors that determine economic growth represented by an outward shift of an economy's PPF. The three major factors are: resource base, investment, and technological change.

Analyse the issue **Applicable concepts: change in resources and sustainability**

Immigration and the PPF

Getty Images/Michael Masters

Over the past decade the world has seen larger migrant and refugee flows than have been witnessed since the great post-war migrations of the late 1940s and 1950s. Whether it is North African and Middle Eastern migrants seeking to live in Europe, Mexicans migrating to the US, Eastern Europeans migrating to the UK, or refugees and migrants wanting to live in Australia, it seems that all around the world people are on the move.

One of the most important consequences of these events has been the worldwide increase in the popularity of political leaders who have promised to halt or at least significantly restrict migrant inflows (think of the likes of Donald Trump or Pauline Hanson). These events are also reflected in Brexit – the decision of a majority of British voters to support Britain's exit from the European Union (EU). When it is no longer part of the EU the UK will be in a better position to control immigration from European countries.

The headlines dealing with the consequences of large-scale people movements typically emphasise their social and political aspects. For example, in New Zealand, net migration rates continued to rise since 2012, peaking at 1.32 per cent of its population of just under 4.7 million at end of 2016. Would the new migrants flood the country with Asians and take away jobs from Kiwis? These were the views expressed by politicians during the 2017 general election campaign.

Reflecting increasing concerns about the adverse effects of migration on unemployment, in April 2017 the New Zealand government announced restrictions on work visas granted to unskilled migrant workers. At the same time, the Australian government announced a tightening of work visas for both skilled and unskilled workers including discontinuation of long-established 457 residency visas.

Fast forward to 2020. In Australia underemployment soared past 9 per cent in June, the highest in nearly two decades. While migrant numbers plummeted due to the border closure, some politicians were concerned if our economy started to recover after the pandemic, the return of migrants would add to our labour force, depressing wages and taking jobs away from Australians. In a press article published in the *Sydney Morning Herald* (May 3, 2020) Senator Kristina Keneally called on the government to redesign its migration policy and 'put Australian workers first'. She warned against importing cheap labour from overseas to lower labour cost for businesses while our youth unemployment rate was double digit.

Whereas an increasingly vocal segment of the population sees immigration as a threat to jobs and quality of life, economists, for good or ill, are more likely to emphasise the increased national output that can accompany an immigration-fuelled increase in the labour force. As the economy expands more jobs will be created.

What do you think?

Using the information provided, undertake the following tasks:

1 Construct a PPF for a country that has no net migrant inflow. Show what happens to the frontier when there is significant immigration of unskilled workers into the country. What about when the immigration is of highly skilled workers?

2 Now think about a country that has net emigration outflows as in the period between 2011 and 2013 in New Zealand. What is the likely effect of these outflows on its PPF?

4 A feeling for the extent to which each of the world's countries has experienced immigration can be found in the Pew Research Center's table 'International Migrants by Country', which can be found at http://www.pewglobal.org/interactives/migration-tables. This table, which includes entries for both Australia and New Zealand, shows the total number of persons who have been living in a particular country for a year or more but who were not born there.

Changes in resources

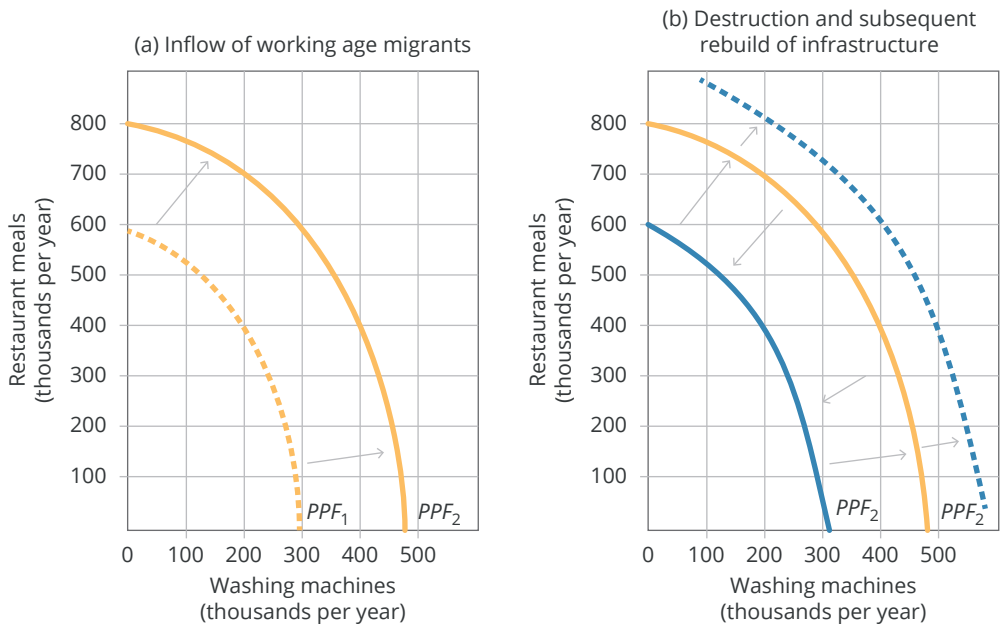
One way to accelerate economic growth is to gain additional resources. Any increase in resources – for example, more natural resources, increased immigration or more infrastructure – will shift the PPF outward. In part (a) of **Exhibit 2.3**, PPF_1 represents New Zealand's production possibilities for restaurant meals and washing machines in a given year. Suppose a series of severe weather events increases the threat that climate-change-induced sea level rises pose to Pacific Island nations. This results in an influx of migrants into New Zealand, in particular, immigration of persons of working age. Such new resources will give New Zealand an expanded capacity to produce more of both goods, pushing its PPF_1 frontier outward to PPF_2.

Reductions in resources will cause the PPF to shift inward. Now consider the frontier PPF_1 – part (b) of **Exhibit 2.3**, which describes Germany's economy before the Second World War. After the destruction of its factors of production in the war, Germany's frontier shifted inward to frontier PPF_2. Over the years Germany trained its workforce, built new factories and equipment, and developed new technology to shift its frontier outward and surpass its pre-war production capacity at frontier PPF_1.

Investment

Population growth can increase a nation's resource base, but it is a double-edged sword. Larger population results in larger labour force capable of producing greater output. However, the increased output will now spread over a larger population. Therefore, the amount of output per person, which determines the general living standard, may or may not increase, depending on whether the rate of output increase is greater than that of population growth. Gross domestic product (GDP) measures an economy's output. Economic growth rate, typically measured by per capita GDP growth, is greatly influenced by labour productivity. (We will discuss how GDP is measured in **Chapter 11**.) Workers with more education and training tend to be more productive. Historical data shows that the numbers of years of education can explain much of the differences in economic growth across countries.

Exhibit 2.3 **Shifting of the production possibilities frontier due to changes in resource base**



In (a) New Zealand begins with the capacity to produce combinations along production possibilities frontier PPF_1. Inflows of working age population shifts the PPF outward from PPF_1 to PPF_2. This outward shift permits the economy to produce greater quantities of output. In (b) The production possibility frontier of Germany shifts inward from PPF_1 to PPF_2 after the WWII due to destruction of factors of production. Not long after the war, it successfully rebuilt its infrastructure, pushing its PPF outward beyond its original frontier PPF_1.

Therefore, a more effective way to raise the general living standard is to increase expenditure on education. Economists call this expenditure *investment in human capital*. A more educated labour force means more skilled and adaptable workers, able to produce more output with the same resource base.

In any economy, farmers, mining workers, accountants, sergeants will not be very productive, however highly educated and trained they may be, without the assistance of modern-day farm and mining machineries, tools, computers and hospital equipment. As discussed in **Chapter 1**, all these made goods are called capital (or physical capital, to distinguish it from human capital). Like investment on human capital, *investment in physical capital* will increase labour productivity and as a result the productive capability of an economy.

Investment
The process of producing human and physical capitals, such as education, factories and machines. Increase in capital stock will shift a country's production possibility frontier to the right.

In short, **investment** in human and physical capital is required to build up the capital stocks of a nation, if the nation wishes to push its PPF outward.

Technological change

Another way to achieve economic growth is through research and development of new technologies. The knowledge of how to transform timber or a block of stone into a wheel vastly improved the standard of living in prehistoric times. Technological change also makes it possible to shift the PPF outwards by producing more from the same resources base. One source of technological change is *invention*. The light bulb, the transistor, the computer chip and antibiotics are all examples of technological advances resulting from the application of science and engineering knowledge.

Technological change also includes the creation and development of new ways of doing things. A century ago Henry Ford revolutionised manufacturing by pioneering the use of the assembly line for making cars. Australian Dr David Warren developed the black box flight recorder in Melbourne in the 1950s, and despite early scepticism about its usefulness, it is now fitted to the aircraft of all commercial airlines. These and a myriad other success stories illustrate the fact that since time immemorial new ideas have been transformed, often by entrepreneurs, into production and practical use. The link between technological change and growth will be analysed more fully in **Chapter 12**.

Social corporate responsibility

Economics and ethics

Economic growth is represented by the outward shifting of the production possibility frontier. This growth involves the use of a range of resources, many of which are destroyed or damaged in the process of production. When this destruction or damage has adverse effects on the wellbeing of individuals, the extent to which the ensuing growth in output is beneficial to humankind is diminished. So, for example, when production of electricity, the harvesting of fish or the production of timber from forests damages the environment by creating carbon emissions, depleting fish stocks and destroying wildlife habitats, there are costs to the community as well as benefits. Increasing awareness of these adverse effects has led to the questioning of the appropriateness of economic growth. Some commentators have gone so far as to suggest that, given these adverse effects, it may be unethical to pursue economic growth at all cost.

Growth in knowledge, the arts, technology and in enlightenment is fine, but growth in these areas doesn't necessarily mean that we must also have material growth. If what is considered to be within the realm of ethical behaviour today has the effect of destroying the wellbeing of future generations, society must rethink its contemporary ethical values. Indeed, many of the cherished beliefs of our forebears, such as the virtue of hard work and the desirability of taming nature, are already seen by many as ethically questionable and inconsistent with a sustainable future.

Corporations undertaking a major infrastructure project, such as the Western Sydney Airport would assess and mitigate the environmental impacts of its construction and subsequent operation on the surrounding communities. There are some investors whose concern for society and the environment has led them to undertake so-called socially responsible investment – investment that seeks good financial returns while, at the same time, promoting social and environmental benefits. Recently, each of Australia's 'big four banks' has declared that it will not help fund the contentious Adani coal mine planned for Queensland's Galilee Basin. In 2014 the Australian National University decided to change its investment portfolio by divesting itself of shares in fossil fuel companies. This move was followed by a few other universities, the latest of which is University of New South Wales in 2020. These moves were lauded by environmentalists but criticised by many conservative politicians and captains of industry.

In summary

- The PPF can shift outward as the result of an increase in resources, **investment** in human and physical capitals or an advance in technology. Any such shift provides the potential for **economic growth**.

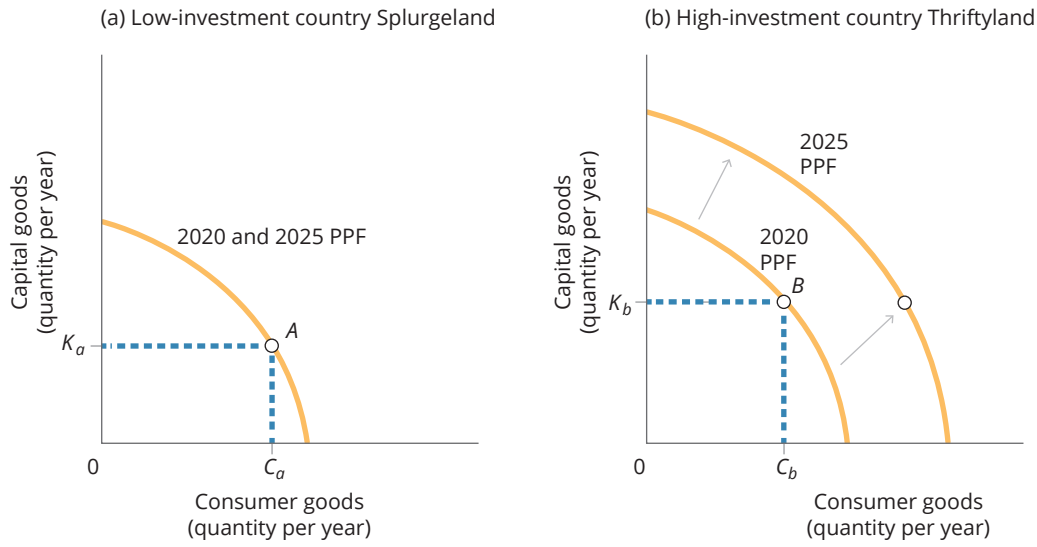
6 Present investment and the future production possibilities frontier

Our earlier discussion on investment in physical and human capitals also raises the issue of inter-temporal trade-offs. Consider the decision for an economy involving a choice between two types of output: capital and

consumer goods. As we will see the output combination chosen for the present period can determine the future production capacity of a country.

Exhibit 2.4 compares two countries producing different combinations of capital and consumer goods. Suppose both countries, Splurgeland and Thriftyland have the same resource base and technology. Diagram (a) shows the PPF for the low-investment economy of Splurgeland. This economy was producing combination *A* in 2020, which consists of a high level of consumer goods, C_a and a low level of capital goods, K_a. Let's assume for simplicity K_a is just enough capital output to replace the capital being worn out each year (i.e., to cover depreciation). As a result, Splurgeland fails to increase its net amount of factories and equipment (i.e., its capital stock). This means that there can be no expansion of its PPF outwards in future years. Why wouldn't Splurgeland simply move up along its production frontier by shifting more resources to capital goods production? The fact is that sacrificing consumer goods for capital formation causes the current standard of living to fall. Perhaps Splurgeland prefers to have more present consumption.

Exhibit 2.4 **Splurgeland's and Thriftyland's present and future production possibilities frontiers**



In part (a), in 2020 Splurgeland produces only enough capital (K_a) to replace existing capital being worn out. Without greater capital and assuming other resources remain fixed, Splurgeland is unable to shift its PPF outward in future. In part (b), Thriftyland produces K_b capital, which is more than the amount required to replenish its depreciated capital. In the year 2020, this expanded capital provides Thriftyland with the extra production capacity to shift its PPF to the right in future.

On the other hand, Thriftyland operated in 2020 at point *B*, which consists of a lower output of consumer goods C_b and a higher level of capital goods K_b, as shown in diagram (b). Assuming K_b is more than enough to replace worn-out capital, Thriftyland is adding to its capital stock. Economists call this activity *capital accumulation*. Comparing Splurgeland to Thriftyland illustrates the importance of capital accumulation. The larger capital stock will create greater production capacity, pushing the PPF outward, not in the present period, but in the future.

Note that both Splurgeland and Thriftyland utilise their resources to produce capital goods, so they both undertake investment. In Splurgeland's case, where the investment just makes good the depreciation of the existing capital stock, we say that there is zero net investment and zero capital accumulation. In Thriftyland's

case, where the investment is in excess of that required to cover depreciation, we say there is positive net investment, resulting in capital accumulation.

Practical examples of the power of high levels of positive net investment to produce high growth rates can be found in many Asian countries, including Singapore, China and South Korea. These Asian countries tend to have high saving and investment rates. By contrast, a number of 'developing' countries have failed to grow in recent decades because they have been unable to achieve positive net investment.[3] Later, in **Chapter 12**, the role that saving and investment plays in promoting growth will be discussed at length.

In summary

- If investment in physical capital is great enough to more than compensate for wear and tear in the existing capital stock (i.e., depreciation), the country experiences capital accumulation.
- A country that devotes more of its resources to increase its capital stock *now* will reap the benefit of higher productivities, expanding its PPF in the future.

7 Trade and production possibility frontier

Up to this point the production possibility frontier shows the way in which an economy can choose between production of *two* different outputs given its endowment of resources and its production technology. This analysis can, of course, be extended to the full range of possible outputs that an economy might produce. The important point to remember is that production of more of any given output must be accompanied by less production of one or a number of other outputs. A little reflection reveals that this analysis can also be applied to an individual's decision about what outputs to produce. So, for example, if an individual spends more time on production of home-based outputs, such as meals and bicycle maintenance, this must be accompanied by less output of other home-based or work-based outputs.

So far in this chapter we have been making an important, but highly unrealistic, assumption – that nations (or individuals) do not engage in trade. Imagine in a community without trade, where a household grows its own foods, builds its shelter and makes its own clothes, the amounts and range of goods and services they can produce and enjoy will be very limited. The reality is that trade is a core activity for both individuals within a community, and between nations. Trade allows us to specialise. As individuals we tend to specialise in certain productive activities from which we earn income that can then be used to purchase goods and services produced by others. As a nation we do the same thing. By engaging in trade, nations and individuals are able to benefit from levels of output that lie outside the confines of their own PPFs. They do this by specialising in production of certain outputs for which they are well suited and then trading the output that is surplus to their own needs with other nations or other individuals. Thus, the hairdresser specialises in cutting, shaping and colouring hair, using the income earned to buy cleaning services from a specialist cleaner, jewellery from jewellers, clothing from clothing stores, food from supermarkets and so on. Similarly, Australia specialises in activities such as primary production, tourism and education; using the proceeds of the export of these outputs to pay for imports such as manufactured goods, software and clothing.

[3] James Stewart argues that a key factor in holding back growth rates in the poorest developing countries is their infrastructure investment deficits – deficits that he says result from debt, the direction of aid away from infrastructure towards social expenditure, and from the high cost of infrastructure in relation to these countries' low levels of income. See J. Stewart 'Growth for low-income countries?', United Nations University, 17 May 2013, https://unu.edu/publications/articles/growth-for-low-income-countries.html, accessed 7 November 2017. The United Nations University is the academic and research arm of the United Nations.

In short, specialisation and trade greatly expand the productive capacities and consumption possibilities of a nation. The theory that enables us to analyse the nature and extent of specialisation undertaken by individuals and nations is called the *theory of comparative advantage*.

Economists agree that specialisation and trade according to our comparative advantages, which makes mass production and mechanisation possible and highly profitable, which in turn promotes technological advancement, are the most significant driver of economic growth in human history. Recently, however, many politicians and some economists have questioned the benefits of trade if it makes a nation better off in aggregate but at the same time reduces incomes of a significant group of citizens. Thus Donald Trump, observing the hollowing out of America's middle class, has said the US will place restrictions on trade with many nations, including Mexico and China. Brexit – the UK's decision to leave the EU – resulted in part from the unpopular effects of free movement of resources, as well as goods and services, between the rest of Europe and Britain. This and other issues related to the theory of comparative advantage, free trade and globalisation will be discussed at length in **Chapter 18**.

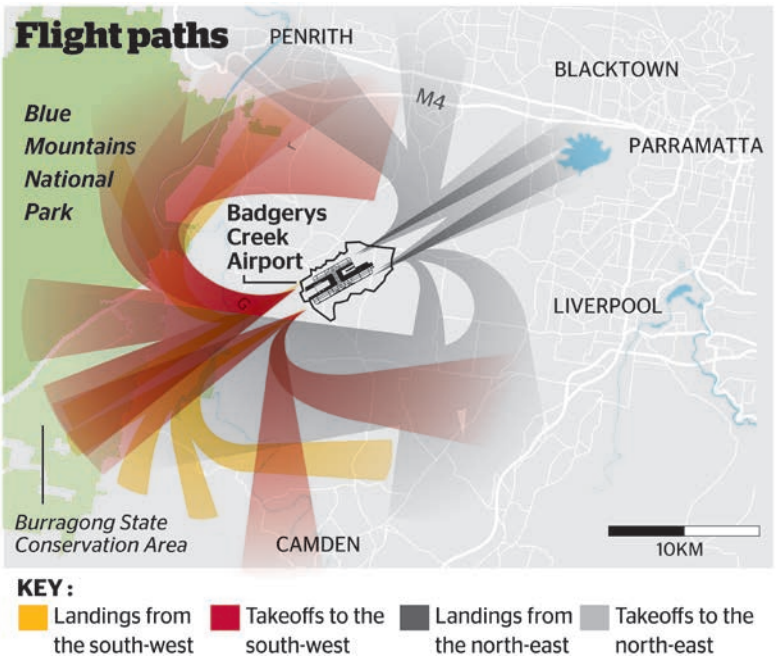
Global perspective — Applicable concept: production possibilities frontier

How can the PPF be used to explain the long-term effects of decisions to increase public capital expenditure?

Earlier in this chapter we looked at the difference in growth between two countries – one with a marked preference for consumer goods, the other with a preference for capital goods. We saw that the country with a strong preference for capital goods (Thriftyland) had higher long-term growth prospects than the country that emphasised production of consumer goods (Splurgeland) for present enjoyment.

When we discussed these differences, you may have been thinking of capital goods produced by the private sector, such as motor vehicles, factories, machinery, office buildings and computers. While private sector capital growth is an important contributor to outward shifts in the PPF, the role of public sector capital is no less important. Thus, government investment in assets such as roads, railways, schools and hospitals can contribute significantly to economic growth.

In 2017, the Turnbull government sought to make political mileage from its announcements of a number of new public investment initiatives. These included a feasibility study of a new Snowy Hydroelectric Scheme (known as Snowy 2.0) designed to raise the efficiency and reliability of Australia's national electricity grid. It resulted in falls in prices and improvements in reliability of electricity supply, which significantly increased the productivity of businesses, which were big users of electricity. A second major public investment initiative announced by the government in 2017 was the decision to build Sydney's second major airport at Badgerys Creek in Sydney's west. Estimated to cost $6 billion, the new airport expected to be in operation in



2026 would reduce congestion at Sydney's existing international airport while reducing travel times and costs for the rapidly growing western Sydney region. A decision to construct an inland rail line from Melbourne to Brisbane was also a feature of the 2017–18 federal budget brought down in May 2017.

For a number of reasons public investment – especially investment involving large infrastructure projects – has recently become more popular with governments around the world. In his inauguration speech US President Donald Trump announced that '[w]e are going to fix our inner cities and rebuild our highways, bridges, tunnels, airports, schools, hospitals. We're going to rebuild our infrastructure, which will become, by the way, second to none. And we will put millions of our people to work as we rebuild it.'[1] Not only would such an initiative put more Americans to work but it would also raise the overall productivity of the US economy.

What do you think?

Construct a PPF that represents a country's goal of producing both consumer goods and public capital. Locate a point on this PPF that shows current production of these alternative outputs. Now show a movement along the PPF resulting from a decision by government to increase public investment funded by new taxes that reduce demand for, and production of, consumer goods. Show how this PPF would move out in subsequent years if (a) the old combination of public capital and consumer goods had been continued, and (b) if it was decided to adopt the new combination with fewer consumer goods and more public investment goods.

[1] 'The Inaugural Address', 20 January 2017, at https://www.whitehouse.gov/inaugural-address, accessed 7 November 2017.

In summary

- By engaging in trade, nations and individuals are able to benefit from levels of output that lie outside the confines of their own PPFs. They do this by specialising in production of certain outputs for which they are well suited and then trading the output that is surplus to their own needs with other nations or other individuals.

Key concepts

What, how and for whom to produce
Opportunity cost
Marginal analysis
Production possibilities frontier
Technology
Law of increasing opportunity costs
Economic growth
Investment

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Investigate the three fundamental economic questions

- Three fundamental economic questions facing any economy:
  - ***what*** to produce: what goods and in what quantities

- ***how*** to produce it: requires firms to decide the resource mix used to produce goods and services
- ***for whom*** to produce it: concerns the rationing of output among society's citizens.

2. Understand opportunity cost and its central role in economic thinking

- **Opportunity cost** is the value of best alternative forgone for a chosen option.
- The cost of an action includes *monetary* and *non-monetary* costs.

3. Know what marginal analysis is and how it is applied

- **Marginal analysis** examines the impact of incremental changes from a current situation. The basic approach is to compare the additional benefits of a change with the additional costs of the change.

4. Describe and graphically present the production possibilities frontier

- A **production possibilities frontier** (PPF) illustrates an economy's capacity to produce goods and services, subject to the constraint of scarcity. The PPF is a graph showing the maximum possible combinations of two outputs that can be produced in a given period of time, subject to three conditions: (1) all resources are fully employed; (2) the resource base is fixed during the time period; and (3) **technology** remains constant. Production that occurs at any point inside the PPF implies resources are not fully utilised. Each point on the frontier represents a maximum output possibility. Points outside the frontier is unattainable due to fixed resources.

5. Identify the factors shifting the production possibilities frontier

- The possibilities frontier can shift outward as the result of an increase in resources, increase in labour productivity or an advance in technology. Any such shift provides the potential for **economic growth**.

6. Examine the impact of present investment on the future production possibilities frontier

- **Investment** occurs when an economy produces new capital. Investment in physical capital involves production of factories, machines and inventories. If physical capital investment is great enough to more than compensate for wear and tear in the existing capital stock (depreciation), then the PPF will shift outward in the future. Investment in human capital will have the same effect on the PPF.

7. Understand the notion of gains from trade

- By engaging in trade, nations and individuals are able to benefit from higher levels of output that lie outside the confines of their PPFs. They do this by pursuing their **comparative advantage**, which involves specialising in production of certain outputs for which they are well suited and then trading the output that is surplus to their own needs with other nations or other individuals.

Study questions and problems

1 Less than half a century ago in many Australian cities it was possible to park cars on streets in the city centre without paying a fee, moor a boat in the local waterway, and dispose of rubbish by incinerating it in the back yard. By contrast, today, parking meters are the norm in city centres, the mooring of boats is closely regulated, and back yard incinerators are illegal. Explain why these changes have come about.

2 Draw a production possibilities frontier showing production of capital goods and consumer goods. Show how the frontier changes when:
- **a** there is a technological breakthrough in production of capital goods
- **b** when there is a breakthrough in production of consumer goods
- **c** when both capital goods and consumer goods benefit from technological change.

3 It is often said that international migration benefits the communities that take these migrants. Yet most of these migrants prefer to settle in bigger cities. More people living in capital cities means that some resources and some goods and services become scarcer. List some examples and provide brief explanation for each. (*Hint*: rises in relative prices may provide some clues as to increased scarcity.)

4 What are some of the opportunity costs of a decision by the New South Wales government to construct a new ring road through an established suburb in which there are many heritage buildings?

5 In recent times there has been a significant increase in immigration to Australia. Explain how this would affect the production possibilities frontier if:
- **a** most of the immigrants are highly skilled people of working age
- **b** most are children and elderly individuals.

6 Which of the following decisions has the greater opportunity cost? Why?
- **a** The decision to use undeveloped land in Beijing for a park.
- **b** The decision to use a square kilometre in the Simpson Desert for a service station.

7 The following is a set of hypothetical production possibilities for a nation:

Combination	Tanks (thousands)	Food (thousands of tonnes)
A	0	10
B	2	9
C	4	7
D	6	4
E	8	0

- **a** Plot these production possibilities data. What is the opportunity cost of the first 2000 tanks produced? The second 2000 tanks? The third 2000 tanks, and the fourth 2000 tanks?
- **b** Does this production possibilities frontier reflect the law of increasing opportunity costs? Explain.
- **c** During peaceful times, which combination of the two goods would the nation normally select?
- **d** Suppose military crisis with a neighbouring country escalates. What combination of the two goods will more likely be chosen?

8 For some years now Australian governments have had a policy of increasing the proportion of young people who attend university. While this reduces the number of people available for work it also increases the skills of these people when they eventually graduate from university and enter the workforce.
- **a** Draw a production possibility frontier of two outputs: university education and all other goods. Illustrate the policy of increasing university education with a movement along the frontier.
- **b** Draw a new production possibilities frontier model to show the effects of such a policy on the output of the nation *over time*.

9 Do you think that environmental assets (e.g., clean air, wilderness areas and wildlife) are becoming more or less abundant over time? Is there some simple economic way to make such a judgement?

Answer to 'You're the economist'

Are consumers and governments always conscious of opportunity cost?

If you said a rise in interest rates means that installation of solar panels involves an increase in interest income forgone, and thus, if consumers are conscious of opportunity cost, fewer installations would occur, YOU ARE THINKING LIKE AN ECONOMIST. In making a decision whether or not to install solar panels, the environmentally conscious consumer would need to consider the opportunity cost of damage to the environment caused by using the solar panels as compared to the damage caused by using electricity from the grid. If you said the environmental damage caused by the manufacture, transport and installation of the solar panels would need to be compared to the environmental damage that goes with providing electricity from power stations that burn fossil fuels, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 A truck driver gives up her $100 000 per annum job in the city to earn $160 000 working at a remote mine in the outback. The opportunity cost of this decision is:
- **a** $60 000 per annum only.
- **b** $100 000 per annum only.
- **c** $160 000 per annum only.
- **d** $100 000 per annum, plus the loss of the benefits of city living.

2 The answer to the *for whom* question is influenced by which of the following factors?
- **a** The educational attainment of individuals in the community.
- **b** The work ethic of individuals in the community.
- **c** The health status of individuals in the community.
- **d** The wealth owned by individuals in the community.
- **e** All of the above are factors in determining who gets what.

3 The law of increasing opportunity cost as applied to a PPF implies that:
- **a** if technology improves, the opportunity cost is the loss of old technology.
- **b** the PPF is a straight line.
- **c** the opportunity cost increases as the production of one output expands.
- **d** capital goods are preferred to consumer goods.
- **e** none of the above.

4 Johnny used to work at Bunnings on the weekend, earning $100 per week. His father, John operates a newsagent. Recently, he asked Johnny to give him a hand in the newsagent instead of working at Bunnings. And Johnny is paid $60 per week for his work at his father's newsagent. What is the opportunity cost of using Johnny's labour in the newsagent?
- **a** zero
- **b** $40
- **c** $60
- **d** $100

5 A band leader is deciding whether or not to take on a keyboard player. If the band adds one keyboard player, the price it can charge for a gig is expected to rise from $800 to $900. According to marginal analysis, the band leader should add a keyboard player if, *ceteris paribus*, he costs less than:
- **a** $800.
- **b** $900.
- **c** $100.
- **d** $1700.

6 An outward shifting of a country's production possibilities frontier can be due to:
- **a** discovery of new mineral resources.
- **b** increase in skilled migrants.
- **c** advancement of technology.
- **d** all of the above.

7 Using a production possibilities frontier, unemployment is represented by a point located:
- **a** near the middle of the frontier.
- **b** at the top of the frontier.
- **c** at the bottom of the frontier.
- **d** outside the frontier.
- **e** inside the frontier.

8 A nation that decided to become self-sufficient by no longer engaging in trade with other nations would:
- **a** have a lower level of output than if it engaged in trade.
- **b** be unable to move outside the production possibilities frontier that describes its efficient output combinations without trade.
- **c** not be profiting from opportunities to pursue comparative advantage.
- **d** all of the above.

9 A nation can accelerate its economic growth by:
- **a** reducing the number of skilled migrants allowed into the country.
- **b** adding to its stock of capital.
- **c** reducing expenditure on research and development (R&D).
- **d** imposing tariffs and quotas on all imported goods.

3 Market demand and supply

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand how consumers make purchase decisions – the law of demand.
2 Explore non-price determinants of demand and their effects on the position of the demand curve.
3 Understand how suppliers make production decision – the law of supply.
4 Explore non-price determinants of supply and the way in which they affect the position of the supply curve.
5 Understand how interaction of consumers and suppliers determines market price – demand and supply analysis.

Shutterstock.com/Alp Aksoy

Markets are the essence of capitalist economies. Economists marvel at the way in which the multitude of markets in a modern economy are able, without any external direction or interference, to provide individuals with jobs, incomes and a continual stream of goods and services. In a *market economy*, buyers and sellers trade millions of goods and services, including homes, alcohol, food, bikes, phones, cosmetics, medical services and marijuana. How can markets coordinate the behaviours of hundreds and thousands of consumers and producers in the economy? This chapter introduces the supply and demand framework that underpins the price system. This framework will prove invaluable because it is applicable to a multitude of real-world events – both large and small. For example, like the legal market for medicinal cannabis, the illegal market for recreational use of cannabis, as we will discuss, is subject to the same forces of demand and supply.

Demand represents the choice-making behaviour of consumers, while *supply* represents the decisions of producers. This chapter begins by looking closely at demand and then at supply. Finally, it combines these concepts to see how prices and quantities are determined in the marketplace.

1 The law of demand

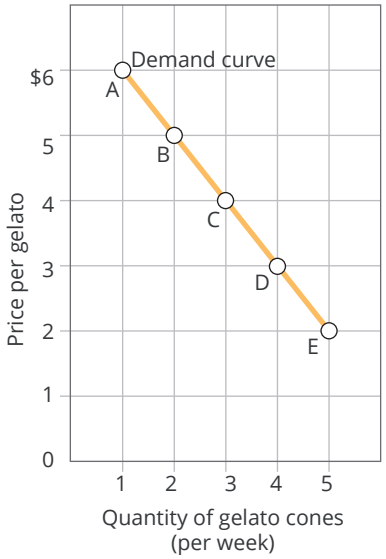
Demand for goods and services in capitalist economies reflects a high degree of **consumer sovereignty**, which is the freedom of consumers to make their own choices about which goods and services to buy. If consumer sovereignty prevails, these choices are made without coercion on the part of government, which can pass laws regarding what we may or may not consume. Let us look at the market of gelato cones in a small town.

Consumer sovereignty The freedom of consumers to make their own choices about which goods and services to buy.

Exhibit 3.1 shows the number of gelato cones Sophie purchases at different prices in tabulated and graphical formats. Each point on the demand curve represents a quantity demanded (measured along the horizontal axis) at a particular price (measured along the vertical axis). For example, at a price of $6 per gelato (point A), Sophie will purchase 1 cone per week. At the lower price of $5, she will purchase 2 cones per week (point B). We can see that when the price of gelato cone falls, she will purchase more per week. The table reveals an important 'law' in economics – the **law of demand**. The law of demand states that there is an inverse relationship between the price of a

Law of demand The principle that there is an inverse relationship between the price of a good or service and the quantity buyers are willing to purchase in a defined time period, *ceteris paribus*.

Exhibit 3.1 **Sophie's demand curve for gelato cones**



Sophie's demand schedule for gelato

Point	Price per gelato	Quantity demanded (per week)
A	$6	1
B	5	2
C	4	3
D	3	4
E	2	5

Sofie's demand curve shows how many gelato cones she is willing to purchase at different prices. As the price of gelato cone declines, the quantity demanded increases and Sofia purchases more per week.

good or service and the quantity buyers are willing to purchase in a defined time period, *ceteris paribus* (i.e., 'other things being constant'). The law of demand makes good sense; for example, consumers buy more when a department store has a sale; more mangoes are purchased at the height of the season when their price is lower than at other times.

Economists are able to verify the law of demand from first principles by looking at human behaviour more closely. When an individual consumes more of a good or service in a given time period, each successive unit provides less additional satisfaction than did the previous unit consumed. Economists call the additional satisfaction *marginal utility* – 'utility' being the economist's preferred synonym for satisfaction.

Law of diminishing marginal utility The principle that the marginal utility obtained a good or service falls when more of it is consumed.

To facilitate discussion let us assume satisfaction (utility) can be measured in monetary terms. Part (a) of **Exhibit 3.2** presents the marginal utilities Sophie obtains from the different numbers of cones consumed per week. When Sophie eats one cone per week, her satisfaction (utility) is $6. When she eats two cones per week, the *second* cone will give her additional satisfaction (marginal utility) of only $5, less than the marginal utility from the first cone. Similarly, the marginal utility from the third is less than that of the second. Her marginal utility falls when Sophie increases her consumption of gelato cones per week. This characteristic of human behaviour is known as the **law of diminishing marginal utility**, which states that when more of a good or service is consumed, the marginal utility to the consumer falls.

Exhibit 3.2 Deriving Sophie's demand for gelato cones

(a) Sophie's marginal utility schedule for gelato

Quantity consumed (per week)	Marginal utility
1	$6
2	5
3	4
4	3
5	2

(b) Sophie's demand schedule for gelato

Price per gelato	Quantity demanded (per week)
$6	1
5	2
4	3
3	4
2	5

From the marginal utility schedule in part (a) of **Exhibit 3.2**, we can derive Sophie's demand for gelato cones. Here is the thinking based on the idea of willingness to purchase. We are willing to make a purchase only if the additional satisfaction is equal to or greater than its price. If the price is high, say, $6 per cone, Sophie is willing to buy one cone per week, but not the second as it gives her only additional satisfaction (marginal utility) of $5, less than the price of $6. She is willing to buy the second cone only if the price drops to $5. Using similar reasoning, Sophie will purchase three cones per week if the price is $4, four cones per week if the price is $3, and so on. Part (b) of **Exhibit 3.2** presents the numbers of cones Sophie purchases per week at different prices.

Demand curve The line connecting the possible price and quantity purchasing responses of an individual consumer.

If we plot Sophie's demand schedule for gelato in a graph, the result is **Exhibit 3.1** – Sophie's **demand curve** for gelato. The demand curve is formed by the line connecting the possible price and quantity purchasing responses of an individual consumer. The demand curve therefore allows you to find the quantity demanded by a buyer at any possible selling price by moving along the curve.

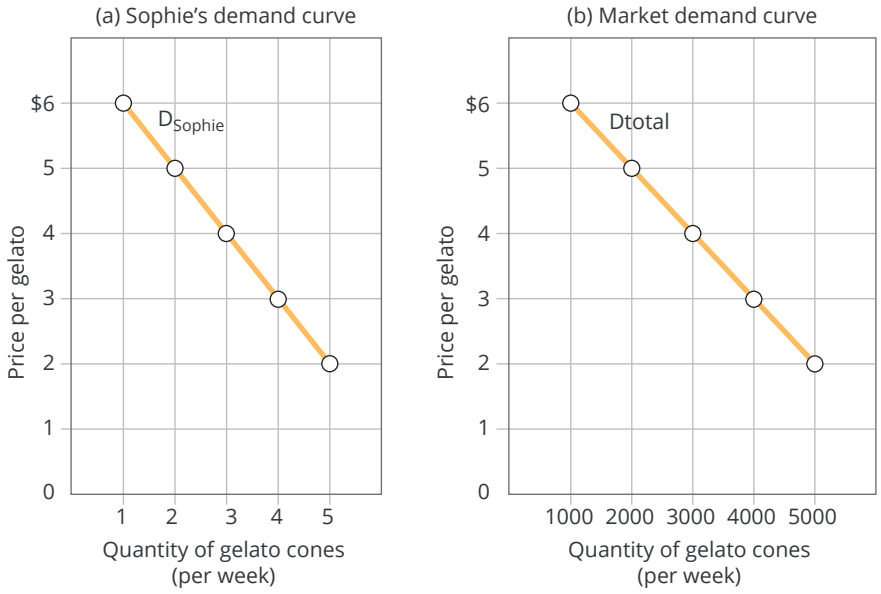
If we examine **Exhibit 3.2** carefully, it is not difficult to see that the two tables are almost identical, apart from their labels. From the law of diminishing marginal utility, economists have derived the law of demand which is consistent with the idea that consumers will purchase additional units of a good or service only if its price decreases.

Note that, until we know the actual price, we do not know how many gelato cones Sofie will actually purchase per week. The demand curve is a summary of her buying intentions. Once we know the market price, a quick look at the demand curve tells us how many cones Sofie will buy.

Market demand

We can now move from an *individual* demand curve to a *market* demand curve. The market demand curve is derived by summing *horizontally* the individual demand curves at each possible price. For simplicity, we assume there are 1000 buyers in the gelato market, and they all have the same tastes. Since their tastes are identical, their demand for gelato are the same. At a price of $6, for example, we multiply Sophie's one gelato demanded per week by 1000 to obtain the total quantity demanded at this price. So, at $6 per cone the total quantity demanded is 1000 cones per week. At $5 the total quantity demand is 2000 per week. Repeating the same process for other prices generates the market demand curve D_{total}. **Exhibit 3.3** presents the market demand for gelato in tabular and graphical form.

Exhibit 3.3 **The market demand curve for gelato cones**



Price per gelato	Quantity demanded (per week)	
	Sophie	Total demand
$6	1	1000
5	2	2000
4	3	3000
3	4	4000
2	5	5000

Since we assume the 1000 consumers have the same taste, we obtain the market demand by multiplying Sophie's quantities demanded at each price by 1000.

The distinction between changes in quantity demanded and changes in demand

Price is not the only variable that determines how much of a good or service consumers will buy. Recall from **Chapter 1** that the price and quantity variables in our demand model are subject to the *ceteris paribus* assumption. If we relax this *ceteris paribus* assumption and allow other variables to change, a variety of factors can influence the position of the demand curve. Because these factors are not the price of the good itself, these variables are called *non-price determinants*. The major non-price determinants include: (1) the number of buyers, (2) tastes and preferences, (3) income, (4) expectations and (5) prices of related goods.

Change in quantity demanded
A movement between points along a stationary demand curve.

Before discussing these non-price determinants of demand, we must pause to explain an important distinction in terminology. We have been referring to a **change in quantity demanded**, which results solely from a change in the price. This is represented by a movement between points along a stationary demand curve. In part (a) of **Exhibit 3.4**, at the price of $5, the quantity demanded is 2000 gelato cones per week. This is shown as point *A* on the market demand curve. At the lower price of, say, $4, a larger quantity demanded of 3000 gelato per week occurs, shown as point *B*. Verbally, we describe the impact of the price decrease as an increase in the quantity demanded. We show this relationship on the demand curve as a movement down along the curve from point *A* to point *B*.

Change in demand
An increase or decrease in the demand at each possible price. An increase in demand is a rightward shift in the entire demand curve. A decrease in demand is a leftward shift in the entire demand curve.

By contrast, a **change in demand** is an increase or a decrease in the demand at each possible price. If one of the five non-price factors changes, *ceteris paribus* no longer applies, the demand curve shifts – to the right if demand increases, to the left if demand decreases.

Part (b) of **Exhibit 3.4**, illustrates an increase in demand for gelato when there is an increase in the number of buyers from 1000 to, say, 1600. The increase in demand is represented by a shift of the demand curve to the right from D_1 to D_2. This means that at *all* possible prices a larger quantity of gelato than before are demanded. At $5, for example, the number of cones demanded increases from 2000 (point *A*) to 3200 cones per week (point B).

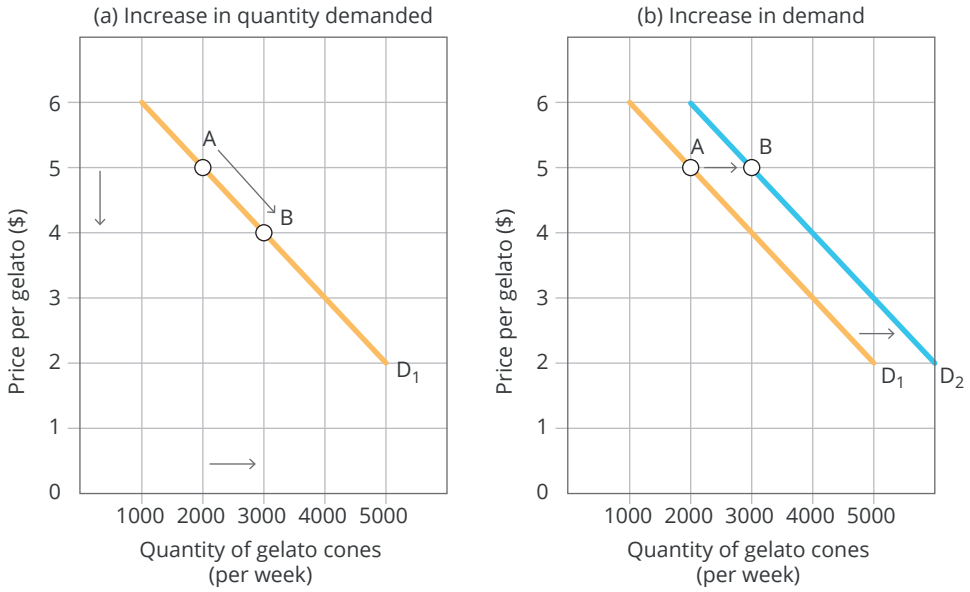
Now suppose a change in some non-price factor (e.g., cooler summers as an unexpected result of climate change) causes demand curve D_1 to shift leftward (a decrease in demand). The interpretation in this case is that at *all* possible prices consumers will buy a smaller quantity than before the shift occurred.

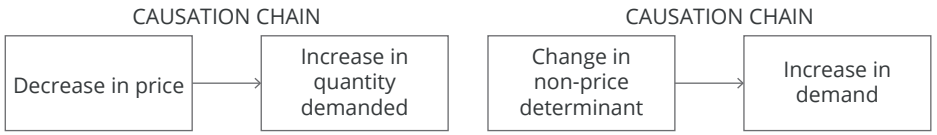
Exhibit 3.5 summarises the terminology for the effect of changes in price and non-price determinants on the demand curve.

In summary

- The **law of demand** states that there is an inverse relationship between the price and the quantity demanded, *ceteris paribus*.
- The **law of diminishing marginal utility states** that that the marginal utility obtained from a good or service falls when more of it is consumed; it explains why consumers will buy more of a good only if its price decreases, *ceteris paribus*.
- The **market demand curve** is the horizontal summation of individual demand curves.
- Under the **law of demand**, any change in price will cause an inverse **change in quantity demanded**, which is described by a movement along the demand curve.
- Changes in **non-price determinants** will produce a shift in the **demand curve**.
- A **change in demand** can involve either an increase in demand (rightward shift) or a decrease in demand (leftward shift).

Exhibit 3.4 Movement along a demand curve versus a shift in demand





Part (a) shows the market demand curve D_1 for gelato cones. If the price is \$5, the quantity demanded by consumers will be 2000 gelato (point A). If the price decreases to \$4, the quantity demanded will increase to 3000 gelato per week (point B).

Part (b) illustrates an increase in demand. An increase in number of buyers can cause an increase in demand. The demand curve shifts from D_1 to D_2. At a price of \$5 on D_1 (point *A*), 2000 cones would be demanded per year. At this price on D_2 (point *B*), the demand increases to 3200 per week.

Exhibit 3.5 Terminology for changes in price and non-price determinants of demand

Change	Effect	Terminology
Price increases	Upward movement along the demand curve	Decrease in quantity demanded
Price decreases	Downward movement along the demand curve	Increase in quantity demanded
Non-price determinant	Leftward or rightward shift in the demand curve	Decrease or increase in demand

2 Non-price determinants of demand

After distinguishing between a change in quantity demanded and a change in demand, we can now discuss how each non-price variable affects demand.

Number of buyers

As we have illustrated previously, if there are more buyers in the market, at all possible prices, there is extra quantity demanded by the new customers, and the market demand curve for gelato cones shifts rightward (an increase in demand). Population growth therefore tends to increase the number of buyers, which shifts the market demand curve for a good or service rightward. Conversely, a population decline shifts most market demand curves leftward (a decrease in demand).

Tastes and preferences

Notwithstanding the importance of consumer sovereignty, we know that fads, fashions, advertising and societal changes can influence consumer preferences for a particular good or service. E-books became popular some years ago, and the demand curve for them shifted to the right. Now that people are tiring of this product, the demand curve is shifting back to the left. In a survey study of impacts of social isolation on alcohol consumption, some ANU academics found that people reported an increase in both frequencies and levels of consumption of alcohol during the COVID-19 pandemic compared to before the pandemic.[1]

Income

Normal good
Any good or service for which there is a direct relationship between changes in income and its demand.

Inferior good
Any good or service for which there is an inverse relationship between changes in income and its demand curve.

As you might expect, changes in income can have very significant effects on demand. For most goods and services, an increase in income causes buyers to purchase more at any possible price. These goods are known as **normal goods**; that is, any good or service for which there is a direct relationship between changes in income and its demand. As buyers receive higher incomes, the demand shifts rightward for *normal goods*, such as airline travel, new cars, restaurant meals, expensive jewellery and designer clothes. A decline in income has the opposite effect, and demand shifts leftward.

In contrast, there are goods or services, the demand for which falls when our income rises. We call them **inferior goods**; that is, any good or service for which there is an inverse relationship between changes in income and its demand. Examples of *inferior goods* include interstate bus travel, second-hand cars, generic-brand groceries and cheap jewellery. Instead of buying these inferior goods, higher incomes allow consumers to buy airline tickets, new cars, brand-name groceries and expensive jewellery. During the Global Financial and Economic Crisis (GFEC) around 2008, many households in Australia experienced a big fall in income, and we observed a rise in consumption of fast food. This shows that fast food was an inferior good.

---

1 N. Biddle, B. Edwards, M. Gray & K. Sollis, 'Alcohol consumption during the COVID-19 period: May 2020', Australian National University,10 June 2020, https://csrm.cass.anu.edu.au/research/publications/alcohol-consumption-during-covid-19-period-may-2020

Expectations of buyers

What is the effect on demand in the present when consumers anticipate future changes in prices, incomes or other factors? What happened in 2000 in the lead-up to the July introduction of the Australian government's goods and services tax (GST)? Because the GST was expected to raise home building costs there was a boom in housing construction in the months before the introduction of GST. On the other hand, suppose car tariffs will be phased out six months from now. An expected fall in new motor car prices would lead to a slump in motor vehicle sales *now*. Suppose new graduates expect their incomes to grow rapidly in a few years after graduation. On this basis, they are likely to commit to substantial levels of borrowing to finance the purchase of a home, a motor car or overseas travel.

Can housing become an exception to the law of demand?

You're the economist

At the end of 2019 and beginning of 2020, house prices across major capital cities in Australia had continued to surge. In the December quarter, median house prices were up by 5.6 per cent, or by more than 22 per cent compared to January 2019.[1] The Reserve Bank of Australia (RBA) predicted the Australian housing market would experience another 'crazy ride' in 2020. (This did not eventuate due to COVID-19.) A number of factors were responsible for these price rises at that time. Most importantly, demand was growing as a result of immigration-fuelled population growth and record low interest rates. These low interest rates encouraged more borrowing by investors and owner occupiers. At the same time construction activity had been slow and the supply of new housing had been falling behind because builders and developers had been cautious about supplying more housing in the middle of the trade tension between China and the US. With demand outstripping supply, an increase in house prices occurred and was expected to continue into 2020. Such a price increase is often a signal for investors and speculators to enter the market. When they do, their actions often result in a full-blown housing boom characterised by increased home sales, as more and more of them enter the market hoping to buy before prices rise even further. This can result in a situation where the quantity of houses sold is increasing *at the same time* that house prices are increasing. Is this an exception to the law of demand?

Getty Images/James Braund

[1] S. Wright, Sydney, 'Melbourne house values surge through start of the year', *Sydney Morning Herald*, 3 February 2020.

Prices of related goods

Remember that in the law of demand the *ceteris paribus* assumption holds all other factors constant, including the prices of other goods. So, what happens if we relax the assumption and allow prices of related goods to change? When we draw the demand curve for an Optus mobile plan, for example, we assume that the prices of a Telstra mobile plan and other similar plans remain unchanged. If we relax the *ceteris paribus* assumption and the Telstra price rises, many Telstra buyers will switch to an Optus plan and the demand curve for the Optus plans shifts rightward (an increase in demand). An Optus mobile plan and a Telstra mobile plan are one type of related goods, called **substitute goods** (or in this case, services). A substitute good competes with another good for consumer purchases. As a result, there is a direct relationship between a price change for one good and the demand for its 'competitor' good. Other examples of substitutes include overseas and domestic holidays, new cars and second-hand cars, and theatre movies and Netflix movies.

Substitute goods Goods that compete with each other for consumer purchases. As a result, there is a direct relationship between a price change for one good and the demand for its 'competitor' good.

Complementary goods
Goods that are jointly consumed. As a result, there is an inverse relationship between a price change for one good and the demand for its 'complementary' good.

Game consoles and games illustrate a second type of related goods, called **complementary goods**, which are goods that are jointly consumed. As a result, there is an inverse relationship between a price change for one good and the demand for its complementary good. Although in many instances buying a game console and buying an electronic game are seen as separate decisions, these two purchases are often related. What happens if the price of game consoles falls sharply from, say, $250 to $50? The market demand curve for games shifts rightward (an increase in demand) because new owners of game consoles add their individual demand curves to those of current owners. Similarly, if the price of games were to fall, more people would consider purchasing a console. The fall in the price of games thus results in an increase in demand for game consoles involving a shift to the right of the demand curve for consoles.

Exhibit 3.6 summarises the relationship between changes in the non-price determinants of demand and the demand curve, accompanied by examples for each type of non-price factor change.

Exhibit 3.6 **Summary of the impact of changes in non-price determinants of demand on the demand curve**

Non-price determinant of demand	Relationship with demand	Examples
1 Number of buyers	Direct	Due to COVID-19 pandemic, travel restrictions reduced overseas tourists in Australia by over 95%. As a result, demand for hotel accommodation fell sharply A decline in the birth rate reduces the demand for baby clothes
2 Tastes and preferences	Direct	For no apparent reason, consumers want fidget spinners, and demand increases, but after a while the fad dies and demand declines
3 Income:		
a Normal goods	Direct	There is an increase in consumers' income, and the demand for designer label clothes increases A decline in income decreases the demand for air travel
b Inferior goods	Inverse	There is an increase in consumers' income, and the demand for rice decreases A decline in income increases the demand for bus travel
4 Expectations of buyers	Direct or inverse	At the start of COVID-19 crisis, consumers expected that supermarkets would soon run out of many daily necessities, such as toilet paper. Consequently, many consumers rushed to buy toilet paper to try to beat the anticipated shortage, and demand for toilet paper increased. (In Australia, such a shortage luckily did not occur.)
5 Prices of related goods:		
a Substitute goods	Direct	A rise in the price of Domino pizza increases the demand for Pizza Hut pizza A fall in the price of imported cars causes lower demand for domestically produced cars
b Complementary goods	Inverse	A decline in the price of laser printers increases the demand for laser printer cartridges A higher price for rum decreases the demand for Coke

Applicable concepts: law of demand, complements and substitutes

Global perspective

'It's brilliant – I am being rewarded not to use my car'

These are the words of Londoner Michele Chung who has taken out an insurance policy that is tailored to where she drives, when she drives and how far she drives. Economics is all about understanding how incentives and disincentives affect human behaviour. In this chapter, among other things, we are looking at how demand for a good or service is affected by a range of factors, the most important of which is price. This price relationship gives rise to the law of demand that tells us that quantity demanded moves in the opposite direction to price. So what is it about Michele's insurance policy that relates to the law of demand?

Michele's car is fitted with a GPS device that tells her insurance company when, where and how far she drives every day. This information, taken in conjunction with the insurer's knowledge of the risks of having an accident in different places at different times, enables the company to charge Michele a different insurance tariff for every day depending on how she is using her car. Michele then adapts to these charges accordingly.

So, if Michele can cycle to work on some days, she pays a very low premium on those days because the only risk of damage to her car is as a result of it being stolen or run into while it is stationary. This is why Michele says she's being rewarded for not using her car. From the economist's perspective she is simply responding to the incentive to save on her insurance – responding to price according to the law of demand. Similarly, Michele may choose not to drive on a notoriously unsafe stretch of road if she knows that she will be paying a significantly higher premium to do so.

In an article in the *Financial Times*, journalist John Reed points out that such a system of charging for insurance has enormous benefits for older, infrequent drivers and people who use their cars only on the weekends.[1] In relation to younger drivers who have very high accident rates at night, the high night-time premium discourages their use of cars at this time, resulting in fewer deaths and injuries among this accident-prone group. And why wouldn't younger drivers respond to this incentive when the insurance rate is 20 times higher at around $2.60 per kilometre at night compared to 13 cents per kilometre in the day?

There is the potential for everybody to benefit from time-of-day pricing for insurance. With accident rates at morning peak times being 50 per cent greater than on weekends and in the evening, the higher premium for this time would result in some drivers avoiding morning peak travel, so reducing the amount of congestion on already busy roads.

Insurance policies like Michele's, which are sometimes called pay as you drive (PAYD) policies, are now available in many countries including the US, Australia, South Africa, Japan and Canada. This type of policy is continually evolving. For example, some PAYD insurers now require the on-board, or smartphone, installation of telematics devices which can measure driving habits relating to vehicle speed, acceleration, braking and cornering.[2]

You can see that driving behaviour can be made to respond to the risk of accidents at different times of day and in different places. But, of course, this will only happen if demand for these types of policies grows. At the present time most car insurance policies involve a flat-rate premium that is charged regardless of when and how often the policyholder uses their car. It is easy to see that once the annual flat-rate premium has been paid, the cost of insurance per kilometre goes down the more kilometres you drive. In fact, the additional insurance cost for an additional kilometre driven – what economists call marginal cost – is zero. When the price paid for additional time spent driving on roads is zero it is no wonder that once the annual premium has been paid, there is no incentive, in terms of insurance cost, to switch to cycling as Michele does because of her use-related insurance policy.

Getty Images/FREDERIC J. BROWN

What do you think?

1 If Michele uses her car less at night because the insurance premium at this time is higher, does this involve a change in demand for car travel or a change in quantity demanded?
2 If drivers such as Michele find that their annual insurance cost is significantly lower as a result of the implementation of the new system, is there likely to be an increase or increase in demand for motor cars as far as these drivers are concerned?
3 If the price of pushbikes were to increase tenfold, would people like Michele be likely to use their car more or less, *ceteris paribus*?

[1] J. Reed, 'Drivers take the money and the box as Big Brother tailors insurance', *The Financial Times*, reprinted in *The Australian*, 3 August 2007, p. 23.

[2] For more information about the use of telematics devices, an example of which is pictured above, and the possible downside of driver information collection, see Jeff Zurschmeide, 'Auto insurers desperately want your driving data, but should you give it to them?', *Digital Trends*, 27 January 2016, https://www.digitaltrends.com/cars/how-telematics-may-affect-your-auto-insurance-rates/, accessed 19 June 2017.

In summary

- *Non-price determinants of demand* are the number of buyers, tastes and preferences, income, expectations, and prices of related goods (substitutes and complements).

3 The law of supply

In everyday conversations, the term *supply* refers to a specific quantity. So, if strawberries are in limited 'supply' at the local fruit and vegetable market, this means only that the quantity available is lower than at other times. This interpretation of supply is *not* the economist's definition. To economists, supply is the relationship between ranges of possible prices and quantities supplied, which is stated as the **law of supply**. The law of supply states that there is a direct relationship between the price of a good and the quantity a seller is willing to offer for sale in a defined time period, *ceteris paribus*. Gela is one of the many sellers of gelato cones in the town. **Exhibit 3.7** shows the number of gelato cones Gela is willing to supply at different prices in tabulated and graphical formats. We can see that while the demand curve is downward sloping, the supply curve is upward sloping to the right. Gela's supply curve shows that when the price of gelato rises, Gela will supply more for sale.

Law of supply
The principle that there is a direct relationship between the price of a good and the quantity sellers are willing to offer for sale in a defined time period, *ceteris paribus*.

Why are sellers willing to sell more at a higher price? An intuitive explanation is that at a higher price, the seller makes more profit. Therefore, they have an incentive to produce more at a higher price. The following provides a more formal explanation.

Supply represents the seller's willingness to produce at various prices. A seller is willing to supply an extra unit of output only if the price covers the additional cost of producing it. In economics this additional cost is known as the **marginal cost**. Marginal cost is the increase in cost of production when an additional unit of output is produced. Part (a) of **Exhibit 3.8** shows Gela's marginal cost of producing gelato cones at various levels of output. For example, the marginal cost of producing the first 100 cones is $2 per cone. If Gela increases its production rate to 200 per week, the marginal cost for the second 100 cones is $3 per cone. Note that the marginal cost rises as Gela increases its production rate.

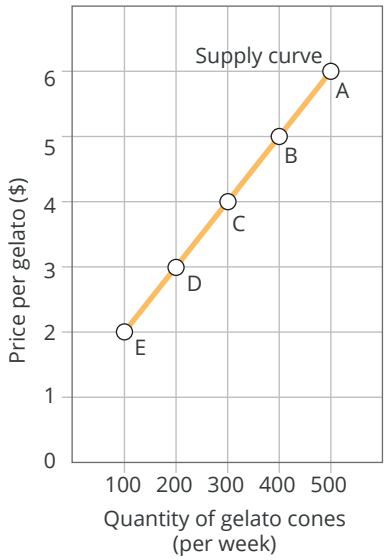
Marginal cost
The increase in cost of production when an additional unit of output is produced.

The perceptive reader may ask: Why does the marginal cost rise as Gela supplies more cones? To answer this question, we assume that Gela is currently producing at the optimal level of 100 cones per week. If it is required

Exhibit 3.7 Gela's supply curve for gelato cones

Gela's supply schedule for gelato

Point	Price per gelato	Quantity supplied (per week)
A	$6	500
B	5	400
C	4	300
D	3	200
E	2	100



The supply curve for an individual seller, such as Gela, shows the quantity of gelato offered for sale at different possible prices. As the price of gelato rises, Gela has an incentive to increase the quantity of gelato supplied.

Exhibit 3.8 An individual seller's marginal cost of producing gelato cones

(a) Gela's marginal costs schedule for gelato

Quantity supplied (per week)	Marginal cost
500	$6
400	5
300	4
200	3
100	2

(b) Gela's supply schedule for gelato

Price per gelato	Quantity supplied (per week)
$6	500
5	400
4	300
3	200
2	100

to substantially increase its production rate by 100 per cent from 100 to 200 or cones per week, it will experience the bottleneck problem. When this occurs, the marginal cost of production rises with output. **Chapter 6** will discuss this phenomenon (diminishing returns) in detail.

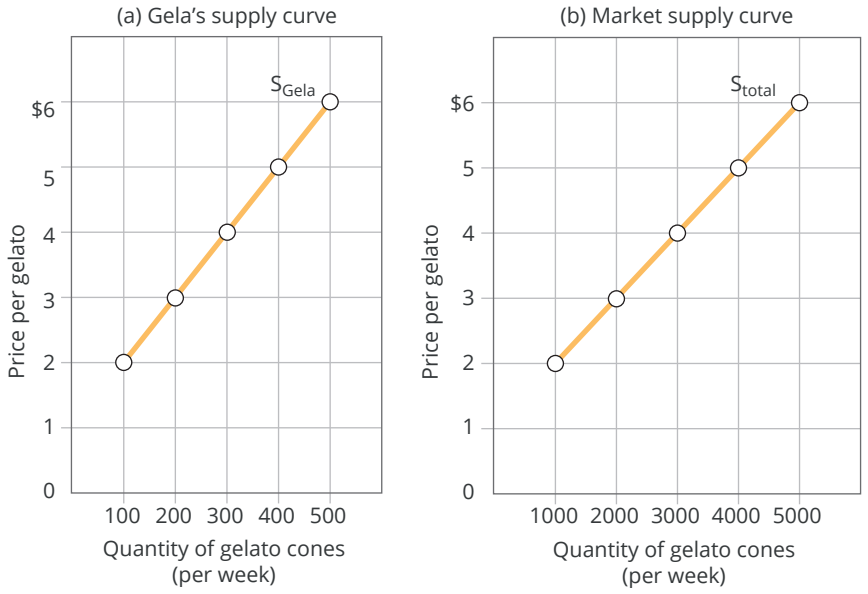
Let's now continue to examine how to obtain the firm's supply curve. As an economic agent, we make decision by comparing the cost and benefit of an action. This applies to Gela, as well. When the price is $2 per cone, Gela is willing to supply 100 cones per week because the price can cover the marginal cost of the first 100 cones but not the second 100. Now if the price is $3, Gela is willing to produce 200 cones. Similarly, based on the marginal cost information in part (a) of **Exhibit 3.8**, we can derive the rest of Gela's supply schedule as presented in part (b) of **Exhibit 3.8**. Comparing parts (a) and (b), we can see that Gela's marginal cost schedule and supply schedule are essentially the same. Therefore, a firm's supply is determined by its marginal cost of production.

Market supply

To construct a *market* supply curve, we follow the same procedure used to derive a market demand curve; that is, we *horizontally* sum all the quantities supplied at various price levels.

Let's assume again for simplicity there are 10 identical firms in the gelato market. Gela's supply curve is presented in part (a) of **Exhibit 3.9**. At a price of $2, Gela would supply 100 gelato cones per week. Since there are 10 identical firms, the market would supply 1000 cones per week. Similar calculations at other prices along the price axis generate the other points on the market supply curve, S_{total}. Part (b) of **Exhibit 3.9** shows that like the individual firm's supply curve S_{Gela}, the market supply curve S_{total} is upward sloping to the right.

Exhibit 3.9 **The market supply curve for gelato cones**



The individual seller and the market supply schedules for gelato

(a) Gela	
Price per gelato	Quantity supplied (per week)
$6	500
5	400
4	300
3	200
2	100

(b) Market	
Price per gelato	Quantity supplied (per week)
$6	5000
5	4000
4	3000
3	2000
2	1000

Gela is one firm selling gelato. If there are 10 identical firms in the gelato market, the market supply schedule can be derived by multiplying the firm's quantity supplied by 10 at each price level.

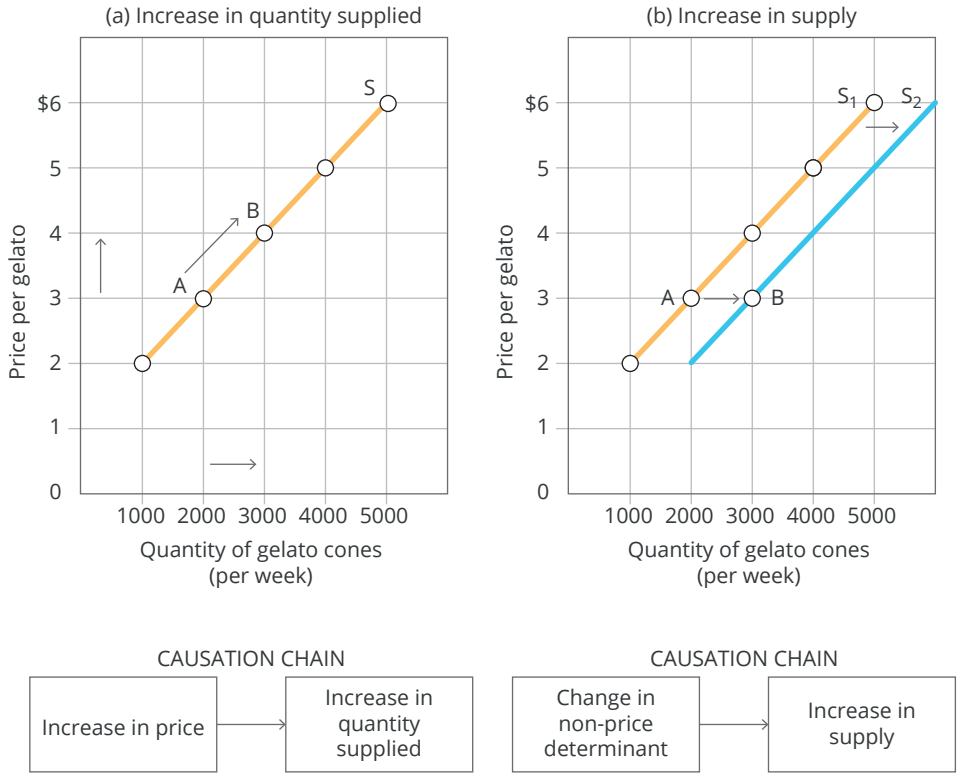
The distinction between changes in quantity supplied and changes in supply

The price of a product is not the only factor that influences how much sellers offer for sale. Once we relax the *ceteris paribus* assumption, there are five principal *non-price determinants* that can shift the supply curve's position: (1) the number of sellers, (2) technology, (3) input prices, (4) taxes and subsidies and (5) expectations of producers. We will discuss these non-price determinants in more detail soon, but first we must distinguish between a change in quantity supplied and a change in supply.

Change in quantity supplied A movement between points along a stationary supply curve, *ceteris paribus*.

A **change in quantity supplied** due to a price change is a movement between points along a stationary supply curve. In part (a) of **Exhibit 3.10**, when the price increases from $3, to $4, sellers increase quantity supplied from 2000 to 3000 gelato cones per week (point *A* to point *B*).

Exhibit 3.10 **Movement along a supply curve versus a shift in supply**



Part (a) presents the market supply curve, *S*, for gelato. If the price is $3 the quantity supplied by firms will be 2000 gelato per week (Point A). If the price increases to $4 the quantity supplied will increase to 3000 gelato per week (Point B).

Part (b) illustrates an increase in supply. A change in some non-price determinant (e.g. an increase in number of firms) can cause an increase in supply. The supply curve shiftsfrom S_1 to S_2. At a price of $3 on S_1 (point *A*), 2000 gelato would be supplied per year. At this price on S_2 (point *B*), the market supply increases to 3000.

Change in supply
An increase or decrease in the supply at each possible price. An increase in supply is a rightward shift in the entire supply curve. A decrease in supply is a leftward shift in the entire supply curve.

A **change in supply** is due to a change of non-price factors, and is represented by a shift of the supply curve.

In part (b) of **Exhibit 3.10**, the rightward shift (an increase in supply) from S_1 to S_2 means that at all possible prices, sellers offer a greater quantity for sale. Suppose the number of sellers increases to from 10 to 15. At $3 per gelato, for instance, the market will increase its output from 2000 (Point A) to 3000 cones per week (Point B). Similarly, the other points on S_1 will also shift to the right. Hence, the whole market supply curve shifts to the right.

Alternatively, a fall in the number of sellers can cause a leftward shift of supply curve (a decrease in supply). **Exhibit 3.11** summarises the terminology for the effect of changes in price and non-price determinants on the supply curve.

Exhibit 3.11 **Terminology for changes in price and non-price determinants of supply**

Change	Effect	Terminology
Price increases	Upward movement along the supply curve	Increase in quantity supplied
Price decreases	Downward movement along the supply curve	Decrease in quantity supplied
Non-price determinant	Leftward or rightward shift in the supply curve	Decrease or increase in supply

In summary

- Under the **law of supply**, any change in price will cause a direct change in the quantity supplied by a firm, *ceteris paribus*.
- The **market supply curve** is the horizontal summation of individual supply curves.
- Under the **law of supply**, any change in price will cause a **change in quantity supplied**, represented by a movement along the supply curve.
- Changes in **non-price determinants** produce a shift in the supply curve and not a movement along the supply curve.
- An **increase in supply** causes a rightward shift of the supply curve. Vice versa.

4 Non-price determinants of supply

Now we turn to the way in which non-price factors affect supply.

Number of sellers

What happens when a severe storm destroys banana farms, or an outbreak of disease ruins an apple crop? The damaging effect of the disease means that apple growers will supply fewer apples at each possible price and supply decreases. Note that this fall in apple production is *not* due to a fall in apply price. A storm has a similar effect on the supply of banana. When the government eases restrictions on hunting kangaroos, the number of kangaroo hunters increases and the supply of kangaroo meat and skins increases.

Technology

Seldom have we experienced such an explosion of new production techniques as we see today. Throughout the world, new and more efficient technology is reducing the cost of production, making it possible to manufacture

more products at any possible selling price. For example, over the last two decades, the number of transistors in a computer chip doubles roughly every two years. The development of more powerful computer chips reduces production costs and increase the supply of all sorts of goods and services.

Input prices

Raw materials, labour, capital and entrepreneurship are all required to produce products, and the prices of these resources affect supply. For example, in the mid-1980s, the Hawke–Keating Labor government struck a deal with the Australian Council of Trade Unions to restrict pay rises in exchange for more social benefits (e.g., more family payments and introduction of Medicare), a deal known as the Accord. Before the Accord, wages rose at very fast paces. Since labour is an important input in most industries, the rise in labour cost reduced supply of most goods and services, pushing the supply curves of most goods and services to the left. Any reduction in production cost caused by a decline in the price of inputs will have an opposite effect and increase supply.

Taxes and subsidies

Certain taxes, such as expenditure taxes, have the same effect on supply as an increase in the prices of inputs. Suppose there is an increase in GST from the current 10 per cent to, say, 15 per cent. This imposes an additional cost on the production of, for example, frozen meals, and reduces their supply – the supply curve shifts leftward. On the other hand, payment of a subsidy by the government would have the same effect as lower input prices; that is, the supply curve shifts to the right. In most developed countries governments subsidise the arts, leading to an increase in supply of services such as theatre productions, symphony concerts and operatic performances.

Exhibit 3.12 **Summary of the impact of changes in non-price determinants of supply on the supply curve**

Non-price determinant of supply	Relationship with the supply curve	Examples
1 Number of sellers	Direct	Suppose Australia increases childcare rebates. Childcare is becoming a more profitable industry and more childcare providers enter the industry. The supply of childcare service rises, and the supply curve shifts to the right.
2 Technology	Direct	The National Broadband Network has greatly reduced the cost of business-to-business (B2B) and business-to-consumer (B2C) transactions undertaken by firms. This results in a general increase in supplies of goods and services in Australia.
3 Input prices	Inverse	A decline in the price of computer chips increases the supply of computers. An increase in the cost of farm equipment decreases the supply of tomatoes.
4 Taxes and subsidies	Inverse and direct	An increase in the tobacco tax reduces the supply of cigarettes. A government subsidy to university researchers increases the supply of research output.
5 Price expectations	Inverse	Oil companies anticipate a substantial rise in future oil prices, and this expectation causes these companies to decrease their current supply of oil.

Expectations of producers

Price expectations can affect current supply. For example, the weather bureaus predict a much colder winter season in the north hemisphere. The consumption of natural gas is expected to increase sharply. Anticipating the price of natural gas to rise Australian gas suppliers will stockpile their output and restrict current supply in order to sell it at higher prices during the coming northern hemisphere winter months to make more profits. The current supply of natural gas will be reduced, and the supply curve shifts to the left. An expectation of a warmer northern winter will have the opposite effects.

Exhibit 3.12 summarises the relationship between changes in the non-price determinants of supply and the supply curve, accompanied by examples for each type of non-price factor change.

In summary

- *Non-price determinants of supply* are the number of sellers, technology, input prices, taxes and subsidies, and price expectations.

5 Market supply and demand analysis

So far, we have discussed the buying intentions of consumers and selling intentions of producers, represented respectively by their demand and supply curves. For example, in the gelato market, Sophie's demand curve tells us how many cones she is willing to purchase per week at a particular price. Similarly, the supply curve shows the numbers of cones Gela will supply at various prices. The key question is: How is the price of gelato determined? As we will see, the price of a good is determined by the interaction of buyers and sellers of the good in the market. A **market** is any arrangement in which the interaction of buyers and sellers determines the price and quantity of goods and services exchanged.

Market
Any arrangement in which the buyers and sellers interact to determine the price and quantity of goods and services exchanged.

Exhibit 3.13 displays the market demand and supply schedules for gelato cones. Notice that, in column 1 of the table, price serves as a common variable for both supply and demand relationships. Columns 2 and 3 list the quantity demanded and the quantity supplied of gelato per week at different prices.

Equilibrium price and quantity

The important question for market supply and demand analysis is: Which selling price will prevail in the market? The economists call this price the **equilibrium** price. It is the condition where there is no pressure to cause the

Equilibrium
A market condition that occurs at any price for which the quantity demanded and the quantity supplied are equal.

Exhibit 3.13 Demand, supply and equilibrium for gelato cones

(1) Price per gelato	(2) Quantity demanded	(3) Quantity supplied	(4) Market condition	(5) Pressure on price
$6	1000	5000	Excess supply	Downward
5	2000	4000	Excess supply	Downward
4	3000	3000	Equilibrium	Stationary
3	4000	2000	Excess demand	Upward
2	5000	1000	Excess demand	Upward

price to alter. Equilibrium occurs at the price at which the quantity demanded and the quantity supplied in the market are equal. **Exhibit 3.13** shows that at the price of $4 per cone the consumers as a whole plan to buy 3000 cones per week, and the suppliers as a whole plan to produce 3000 cones per week. This means at that price the consumers can purchase the amount they plan to purchase, and the suppliers can sell the amount they plan to sell. The market is cleared. Economists also refer to the *equilibrium* price as *market-clearing* price.

To better understand equilibrium, we look at *disequilibrium*. Let's start by asking what will happen if the market price is above equilibrium, say, $6 per cone. At this relatively high price consumers are willing to purchase only 1000 cones per week and the suppliers are willing to produce $5000 cone per week. As a result, 4000 cones remain unsold. When quantity supplied is greater than quantity demanded in the market, **excess supply** occurs (**Exhibit 3.13** column 4).

Excess supply
A market condition existing at any price where the quantity supplied is greater than the quantity demanded.

How will sellers react to an excess supply? Excess supply creates disequilibrium in the market as the sellers cannot sell all their output. Competition between sellers forces sellers to bring down their selling price in order to attract more sales. If they cut the selling price to $5 per cone, there will still be an excess demand of 2000 cone per week, and pressure on sellers to cut their selling price will continue until the price falls to $4 per cones where the excess supply is completely eliminated.

Now imagine that, instead of being highly priced, gelato cones are sold for just $2 each. This price is very attractive to consumers, and **Exhibit 3.13** shows that at $2, the quantity demanded is 5000 cones per week. However, sellers are willing to supply only 1000 cones per week at this price. At $2 per cone, many consumers will be disappointed as the quantity available cannot meet all their demand. A shortage or excess demand occurs. **Excess demand** is a market condition where the quantity supplied is less than the quantity demanded.

Excess demand
A market condition existing at any price where the quantity supplied is less than the quantity demanded.

In the case of an excess demand, unsatisfied consumers compete to obtain the product by bidding to pay a higher price. Because sellers are seeking the higher profits that higher prices make possible, they gladly respond by setting a higher price of, say, $3 and increasing the quantity supplied to 2000 per week. At the price of $3 per cone, shortage will be reduced, but not eliminated. There is still unfulfilled demand exerting an upward pressure on price until the price rises to $4 where supply equals demand.

Applicable concepts: changes in supply and demand, market equilibrium price and output

Analyse the issue

Using market forces to weed out drug syndicates

iStock.com/FatCamera

In some countries in Europe and some states in the US, selling and growing of cannabis (or marijuana) is legal. Should cannabis trading also be allowed in Australia? Legalisation of cannabis has always been a hotly debated public issue. In a 2018 Gallup poll, 66 per cent of Americans supported legalisation of marijuana. The report cites the use of cannabis for medicinal benefits and freeing up police force to deal with other crimes as the main reasons for their support for legalisation.[1] The opponents, however, argued that besides more car accidents and the greater harms to mental health this stronger drug causes, legalisation will send the wrong message to people that taking marijuana is fine, encouraging young people to take the more addictive drug.

What does demand and supply analysis tell us about the impacts of legalisation on consumption of cannabis? Once legalised, there will be a rise in cannabis growers and traders. The increased competition between suppliers in the legal cannabis market can lead to a fall in market price and a rise in consumption.

However, Alice Mesnard, an economist based in the UK, argued that by heavily cracking down on trading of illegal drugs, overall cannabis consumption can be contained. She argues that heavier fines and sanctions on the illegal market will make it much more costly for drug syndicates to smuggle and sell illegal cannabis, reducing its supply.[2] On the other hand, heaver penalties will deter consumers from buying in the illegal market, resulting in a fall in demand for illegal cannabis.

In an economic simulation study published in 2019, Mesnard and her colleagues modelled the impacts of various policy alternatives on cannabis consumption.[3] Based on American data, they found that a right mix of heavy sanctions on selling and buying illegal drugs, and setting a sufficiently high price of legal cannabis can reduce drug-related criminal activities *and* at the same time prevent big increase in cannabis consumption.

What do you think?

Using the information provided, answer the following questions:

1 How will legalisation affect the number of legal cannabis growers and supply of legal cannabis? Illustrate with a diagram.
2 How are legal and illegal cannabis related? Do you think they are substitutes or complements? With the help of diagrams, show the effect of legalisation on the demand for (a) legal cannabis, and (b) illegal cannabis?
3 Explain how heavier fines and sanctions on illegal drugs will influence (a) the demand for, and (b) the supply of, illegal cannabis. Illustrate the changes in a diagram.
4 **(Extension activity)** Considering the legal cannabis market, discuss the effects of legalisation and heavy sanctions on illegal drugs on the equilibrium price and output of legal cannabis.
5 **(Extension activity)** Combining the effects of legalisation and heavy sanctions on illegal drugs, show that both the price and consumption of illegal cannabis *can* go down.

[1] J. McCarthy, 'Two in three Americans now support legalizing marijuana', *Gallup*, 2018, https://news.gallup.com/poll/243908/two-three-americans-support-legalizing-marijuana.aspx?g_source=link_newsv9&g_campaign=item_258149&g_medium=copy)

[2] A. Mesnard (2019) 'Economics of legalising cannabis – pricing and policing are crucial', *The Conversation*, 30 July 2019, https://theconversation.com/economics-of-legalising-cannabis-pricing-and-policing-are-crucial-119914

[3] E. Auriol, A. Mesnard, and T. Perrault, 'Defeating crime? An economic analysis of cannabis legalization policies', CEPR Discussion Papers 13814, C.E.P.R. Discussion Papers, 2019, https://ideas.repec.org/p/cpr/ceprdp/13814.html

Efficiency
A condition where the scarce resources are utilised to maximise the social surplus. Efficiency occurs at the competitive market equilibrium.

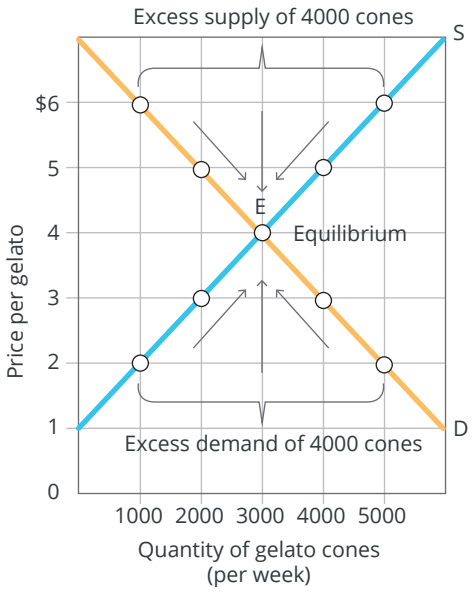
Social surplus
The total net gain by the society in an exchange. Social surplus is the sum of consumer surplus and producer surplus.

Exhibit 3.14 graphically shows how the market forces push the market price down when there is an excess supply and push it up when there is an excess demand. The market will come to an equilibrium when demand equals supply at $4 per cone in the gelato market.

Efficiency

It is important to note that when the forces of demand and supply are in balance at equilibrium, the market outcome is *also* efficient. Economists have a precise definition of efficiency. Recall economics studies how human beings deal with the problem of scarcity. **Efficiency** occurs when the market maximises the total net benefit it creates from the use of society's scarce resources. The net benefit of a unit of output is the difference between its utility to consumers and its cost of production. This is an important idea. Let's examine it in some detail. When Gela sells one cone to Sophie, the cost of production of this cone is $2 and the utility to Sophie is $6; a net benefit of $4 is therefore created. Using the same the argument, we can obtain the net benefits for the other 2999 cones sold in the market. In the gelato market at the equilibrium output of $3000, the total net benefit created – the total benefit of the 3000 cones to consumers minus the total cost of producing them – is represented by the green-shaded area as shown in **Exhibit 3.15**. The total *net* benefit created is known as **social surplus**.

Exhibit 3.14 **The supply and demand for gelato cones**



CAUSATION CHAINS

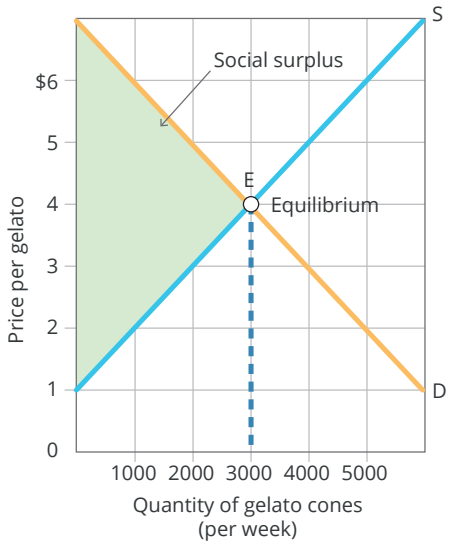
Quantity supplied exceeds quantity demanded →	Excess supply →	Price decreases to equilibrium price
Quantity demanded exceeds quantity supplied →	Excess demand →	Price increases to equilibrium price

The supply and demand curves represent the market for gelato. The intersection of the demand curve *D* and the supply curve *S* at point *E* indicates the equilibrium price of $4 per cone and the equilibrium quantity of 3000 cones bought and sold per week. At any price above $4, an excess supply occurs and pressure exists to push the price downward. At any price below $4, an excess demand (or shortage) condition provides pressure to push the price upward.

To see why social surplus is maximised at the equilibrium output of 3000, let us consider output levels *other than* 3000 cones per week. If output is reduced to 2000 cones per week, the social surplus will be reduced to the green-shaded area as shown in part (a) of **Exhibit 3.16**. Clearly this social surplus is less than when output is 3000 cones per week. Therefore, there is a loss to society. The social loss is represented by the blue triangle.

Alternatively, consider an output level beyond equilibrium, say, 4000 cones per week. Here, for any output beyond 3000, the marginal utility as measured by the demand curve is less than the marginal cost indicated by the supply curve. In other words, instead of creating value, the production and consumption of gelato beyond equilibrium *destroys* value. If 4000 cones were produced, the society as a whole would be worse off by an amount equal to the blue triangle as shown in part (b) of **Exhibit 3.16**. At this point it is important to note that consumers would not be willing to pay a price equal to the seller's marginal cost for any cone beyond 3000. Therefore, no mutually beneficial exchange can occur. In other words, in a free market, where exchange is voluntary, output beyond market equilibrium would not occur.

Exhibit 3.15 Social surplus at equilibrium output for gelato cones



In the gelato market, at the equilibrium market price of $4 per cone, 3000 cones are sold. At the equilibrium output, the social surplus – total benefit or utility to the consumers minus total cost of producing the 3000 cones, is maximised.

Exhibit 3.16 Social loss and inefficient output

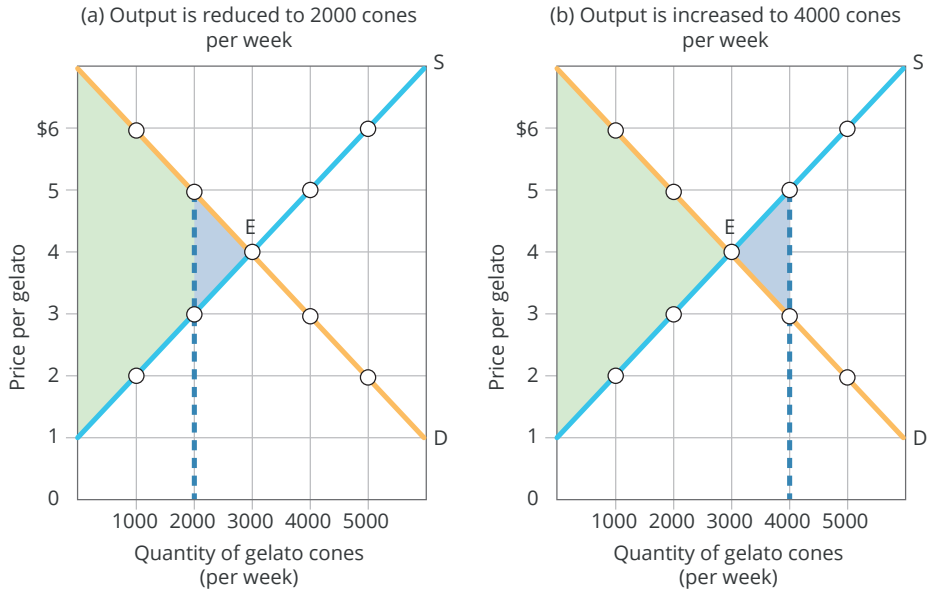


Diagram (a): If the production of gelato cones is reduced to 2000 per week below market equilibrium of 3000, social surplus is reduced to the green area between the demand (D) and supply (S) curves up to the output of 2000. A social loss, blue shaded triangle, occurs.

Diagram (b): If output were increased to 4000 cones per week, social loss (blue triangle) would occur. This is because MC > MU for output beyond equilibrium.

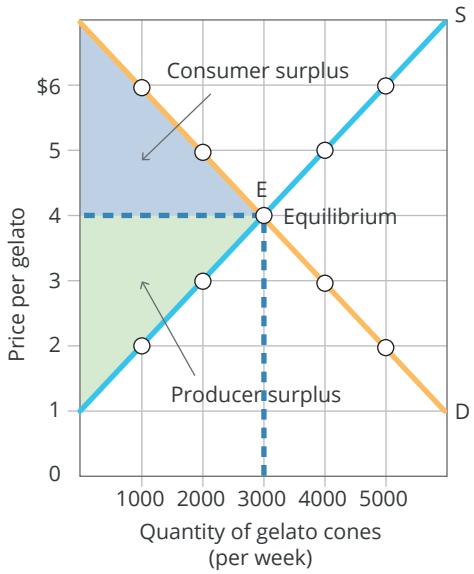
Our discussion so far shows that the social surplus is maximised at equilibrium in a market characterised by many sellers and buyers. It is silent as to how the social surplus created is shared between the consumers and producers. To complete this section, we will introduce two related concepts – *consumer surplus* and *producer surplus*.

Consumer surplus is defined as the gain by the consumer in an exchange, whereas **producer surplus** is the gain by the supplier. The consumer gains in an exchange if the additional satisfaction they get from consumption of a good or service is greater than the price they pay for it. For example, consuming the first gelato cone, Sophie gets a marginal utility of $6 but she pays only $4 for it. She gains a surplus of $2; we say her consumer surplus is $2 for this cone. Recall that demand is determined by marginal utility. **Exhibit 3.17** shows that at the market equilibrium (point E), the *consumer surplus* is represented by the blue triangle enclosed by the demand curve and the market price.

Consumer surplus
The gain by the consumer in an exchange. At the equilibrium market price, the consumer surplus is represented in the area under the demand curve and above the market equilibrium price.

Producer surplus
The gain by the producer in an exchange. At the equilibrium market price, the producer surplus is represented in the area above the supply curve and below the market equilibrium price.

Exhibit 3.17 **Social surplus, consumer surplus and producer surplus for gelato cones**



The consumer surplus is the area under the demand curve (D) and above the market price of $4. The producer surplus is the area above the supply curve (S) and below the market price of $4. The social surplus is the sum of consumer and producer surpluses. In the gelato market, the social surplus is equally split between the consumers and producers, but in general this is not the case.

On the other hand, the supplier gains if the price they receive is greater than the cost of producing the good. For example, Gela sells one gelato cone to Sophie for $4 and the cost of producing this cone is $2. Therefore, the producer surplus for this gelato cone is $2. And since supply is determined by the marginal cost of production, at the market equilibrium (point E) the *producer surplus* is the green triangle enclosed by the market price and the supply curve as shown in **Exhibit 3.17**.

In any exchange, it follows, the *social surplus* is the sum of consumer surplus and producer surplus. In the gelato example, note that the blue triangle and the green triangle have the same area; the social surplus is equally split between consumers and producers, but in general it is not the case. The split depends on the shapes of the demand and supply curves. (Do you know who will gain more if the demand curve, for example, is flatter?) We will learn more about the steepness of demand and supply curves and its applications in **Chapter 5**.

Economics and ethics

Are you for commercial surrogacy?

When a woman is unable to bear a child but can produce eggs that can be fertilised, there arises the possibility of having her fertilised egg implanted in the womb of a surrogate mother. In recent years this type of surrogacy – technically described as gestational surrogacy – has become increasingly popular, with demand for surrogate mothers increasing significantly. Many factors have contributed to this situation including improved reproduction technology, changing social attitudes to such a sensitive issue and increased publicity surrounding it. It seems, however, that governments have been slow to react to this increase in demand, with much confusion and many ad hoc changes to policy occurring as the medical, social and ethical issues that surround surrogacy play out.

Generally speaking, governments have tended to look favourably on altruistic surrogacy where the surrogate mother receives no payment for bearing the biological mother's child. Often these surrogate mothers are relatives or close friends of the biological mother. When it comes to a commercial relationship between the biological and surrogate mother, however, governments' views can be very different, with commercial surrogacy outlawed in many jurisdictions. Where such surrogacy is allowed in developed nations, it is commonly the case that the surrogacy fee is very high – around $100 000 in many cases. These high prices reflect supply and demand for surrogate mothers.

It is easy to see that with a given demand for surrogate mothers, the price of surrogacy will depend on the supply of these surrogate mothers. The high prices for surrogate mothers in developed countries suggest that, even though the number of women seeking surrogacy arrangements with surrogate mothers may be relatively small, the number of women willing, or legally able, to supply services as surrogates is also very small. These high prices, along with bans on commercial surrogacy, have encouraged women seeking surrogacy arrangements to look beyond their home country's borders for alternatives. In turn this has encouraged suppliers of medical services in low-wage countries to provide surrogacy services for these women. These trends resulted in the establishment of booming surrogacy businesses in developing countries such as India and Thailand. Sometimes described as 'baby factories for the rich', surrogacy clinics for foreigners have since been outlawed in these countries. Whereas a surrogacy arrangement might cost a biological mother $100 000 in the west, a similar service could be had in parts of Asia for $50 000. This significantly lower price reflects the much larger supply of surrogate mothers in these countries.

As you have no doubt learnt from the media, this development of cheap surrogacy services in developing countries has resulted in the airing of a wide range of contentious issues. These include the ethics of having poor women in these countries carry babies for generally rich westerners, problems of non-completion of contracts when one or other of the parties to the contract does not fulfil their side of the bargain, and issues surrounding the inequitable division of the surrogacy fee between the surrogate mother and the organisation that has set up the arrangement.[1] Taken together, these factors have caused many commentators to argue that for ethical reasons surrogacy should take place in the country of residence of the biological mother.

In a nutshell, high prices for surrogate mothers in developed countries brought about by high wages in those countries and by their legal restrictions on commercial surrogacy have discouraged surrogacy at home in favour of surrogacy abroad where low wages and limited legal restrictions have caused surrogacy to flourish. Just as the textile industry and manufacturing have migrated from developed countries to developing countries, so surrogacy has done the same thing. But it seems that from an ethical perspective we see the migration of surrogacy services in a different light from the migration of manufacturing.

Just as demand and supply analysis can help us understand the boom in provision of surrogacy services in developing countries, it can also help us understand, and deal with, issues surrounding the ethics of surrogacy. If, on ethical grounds, it is considered better to encourage surrogacy at home rather than abroad, an understanding of the market for surrogacy services can assist us in developing policies to bring this about.

There are many ways in which western governments could encourage surrogacy at home. They could increase supply at home by lifting restrictions on commercial surrogacy,[2] increase home demand by subsidising patients' costs through the national health system; or decrease foreign supply by restricting imports of surrogacy (by preventing

residents from purchasing these services abroad). They could also encourage developing countries to limit access by non-residents to their surrogacy services.

As you can see, an understanding of demand and supply can give us useful insights into the ethics of surrogacy.

[1] Perhaps the best-known example of a contractual breakdown in a surrogacy case is that of baby Gammy who, it was alleged, was not brought home from Thailand by his biological parents. His case can be investigated by typing <baby Gammy> into a search engine.

[2] The Australian parliament published a report into surrogacy in May 2016. The report recommended against introduction of commercial surrogacy in Australia. This report 'Surrogacy matters: Inquiry into the regulatory and legislative aspects of international and domestic surrogacy arrangements', can be found at https://www.aph.gov.au/Parliamentary_Business/Committees/House/Social_Policy_and_Legal_Affairs/Inquiry_into_surrogacy/Report, accessed 9 March 2021.

Let us summarise the key points from our discussion. The gelato market is an example of the competitive market, which will be fully presented in **Chapter 7**. The interactions between consumers and suppliers in a competitive market determines the equilibrium price and output. This market outcome is efficient in that the market utilises scarce resources to produce at the output level that maximises social happiness (social surplus).

Moreover, the price mechanism plays another important role. The market price serves the function of rationing output based on the consumers' willingness to pay. Considering the gelato market, at the equilibrium price of $4, the 3000 cones produced per week will go to the consumers who value gelato at or above $4. In our simplified scenario, it is assumed for simplicity all the 1000 consumers in the town have the same tastes, and hence the same demand curve. Each consumer will purchase three cones per week. In the real world, however, consumers typically have different tastes and incomes, and hence different willingness to pay. As a result, there will be no sale to consumers whose willingness to pay is less than $4 per cone. It is the same mechanism most markets use to ration their output to answer the *for whom* question in free market economies. For example, who drives a Mercedes? Who stays in a Hilton executive suite? This is in essence how the **price system** operates. In the next chapters, we will explore the allocation and rationing functions of different types of market in more detail.

Price system
A mechanism that uses the forces of supply and demand to determine the equilibrium price and output. The price system serves the function of resource allocation and rationing of output.

In summary

- **Equilibrium** is the stable price and quantity established at the intersection of the supply and the demand curves.
- At equilibrium, the competitive market is **efficient** as it creates the greatest social surplus, maximising the welfare of the society.
- The **price system** serves to ration output according to our willingness to pay.

Key concepts

Consumer sovereignty
Law of demand
Law of diminishing marginal utility
Demand curve
Change in quantity demanded
Change in demand
Normal good
Inferior good
Substitute good

Complementary good
Law of supply
Marginal cost
Change in quantity supplied
Change in supply
Market
Equilibrium
Excess supply
Excess demand
Efficiency
Social surplus
Consumer surplus
Producer surplus
Price system

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand how consumers make purchase decision – the law of demand

- The **law of demand** states that there is an inverse relationship between the price and the quantity demanded, *ceteris paribus*.
- The **law of diminishing marginal utility** explains why consumers will buy more of a good when its price decreases, *ceteris paribus*.
- A market **demand curve** is the horizontal summation of individual demand curves.
- Under the law of demand, any change in price, measured along the vertical axis will cause an inverse **change in quantity demanded**, measured along the horizontal axis – a movement along the demand curve.

2. Explore non-price determinants of demand and their effects on the position of the demand curve

- **Non-price determinants of demand** are the number of buyers, tastes and preferences, income (**normal goods** and **inferior goods**), expectations, and prices of related goods (**substitute goods** and **complementary goods**).
- Changes in non-price determinants produce a shift in the demand curve and not a movement along the demand curve.

3. Understand how suppliers make production decision – the law of supply

- The **law of supply** states that there is direct relationship between price and quantity supplied, *ceteris paribus*.
- An increase in price will cause an increase in quantity supplied. Only at a higher price will it be profitable for sellers to incur the higher opportunity cost associated with producing and supplying a larger quantity.
- A market supply curve is the horizontal summation of individual supply curves.
- Under the law of supply, any change in price along the vertical axis will cause a direct **change in quantity supplied**, measured along the horizontal axis, represented by a movement along the supply curve.

4. Explore non-price determinants of supply and the way in which they affect the position of the supply curve

- Non-price determinants of supply are the number of sellers, technology, input prices, taxes and subsidies, and expectations.
- Changes in non-price determinants produce a shift in the supply curve and not a movement along the supply curve.

5. Understand how interaction of consumers and suppliers determines market price – demand and supply analysis

- **Equilibrium** is the price and quantity established at the intersection of the supply and the demand curves. At the equilibrium price there is no tendency for the price to change, other things being unchanged.
- At equilibrium, the market is efficient since the **social surplus** (sum of **consumer surplus** and **producer surplus**) is maximised.
- The **price system** addresses the *for whom* question by rationing output according to consumers' willingness to pay.

Study questions and problems

1 Draw graphs to illustrate the difference between an increase in quantity supplied and an increase in supply of banana in Queensland. Give a possible reason for the change you have shown in each graph.

2 Suppose electricity prices fall significantly over a number of years as a result of the lowering of costs of generation using solar and wind power. What happens, and why, to the:
- **a** demand for home insulation
- **b** demand for battery-powered cars
- **c** demand for domestic gas
- **d** demand for air conditioners?

3 If a housing industry spokesperson announced during the coronavirus pandemic that there had been a change in demand conditions leading to a fall in house prices across Australia, would the spokesperson be talking about a decrease in demand or a decrease in quantity demanded? Draw a graph to illustrate your answer. What could be some of the factors leading to this change in the housing market?

4 Predict what will happen to either supply or demand in the following situations:
- **a** As a result of the improving economy, several large overseas companies decide to enter the supermarket business in Australia. What is the effect on the grocery market?
- **b** Consumers suddenly decide that tattoos are unfashionable. What is the effect on the market of tattoo service?
- **c** The Department of Employment issues findings of a study that shows that playing computer games increases young people's job prospects. What is the effect on the market of computer games?
- **d** The price of air conditioners increases significantly. What is the effect on the market for electricity?
- **e** Tobacco lobbyists successfully persuade the government to raise the excise tax paid by sellers on cigarettes. What is the effect on the cigarette market?

5 Using supply and demand analysis, explain and illustrate graphically the effect of the following situations.
- **a** A baby boom markedly increases the number of children aged five and under. What is the effect on the childcare industry?
- **b** There is a fall in the prices of resources used in the production of houses. What will happen to the housing market?
- **c** The government introduces a 'sugar tax' on soft drinks. How will this tax affect the consumption of soft drinks?
- **d** The income of consumers of personalised number plates decreases. How will the income increase affect the price of personalised number plates?
- **e** The local government constructs more cycling tracks in the city. How will the push bike market be affected?

6 In Adelaide, there is a shortage of Australian Rules football tickets for some games, whereas there is an over-supply for other games. Why do shortages and over-supply exist for different games?

7 List four non-price determinants of supply. Does each of these have a direct or indirect relationship with the supply curve?

8 Explain why the intersection of the supply curve and the demand curve is the only point consistent with a stable equilibrium. You may use a diagram to assist with your explanation.

Answer to 'You're the economist'

Can housing become an exception to the law of demand?

As the price of houses began to rise, the expectation of still higher prices caused investors and speculators to buy more at any given price; that is, demand increased. As shown in **Exhibit 3.18**, suppose the typical price for a capital city house was initially at P_1 and the quantity demanded was Q_1 on demand curve D_1 (point *A*). Then the boom caused the demand curve to shift rightward to D_2. Along the new demand curve D_2, investors and speculators increased their quantity demanded to Q_2 at the higher price of P_2 (point *B*).

Exhibit 3.18 **Increase in demand for houses**

The expectation of ongoing increases in house prices caused 'an increase in demand', rather than 'an increase in quantity demanded' in response to a higher price. If you said there are no exceptions to the law of demand, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 The law of demand states that the quantity demanded for a good (A) falls, other things being equal, when:
- **a** the prices of other goods rise.
- **b** consumer income falls.
- **c** the price of other good A rises.
- **d** the number of buyers falls.

2 Which of the following will cause a shift of the supply curve for electric kettles to the left?
- **a** A fall in the price of a close substitute.
- **b** A fall in the price of electric kettles.
- **c** An increase in demand for electric kettles.
- **d** An increase in the price of resources used to make electric kettles.

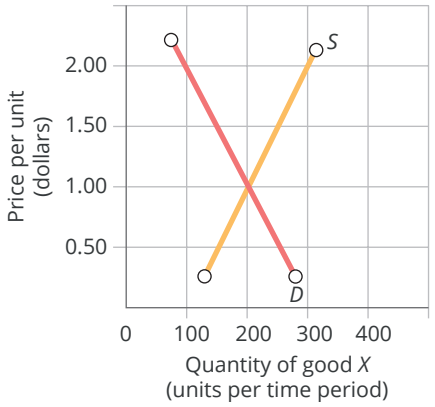
3 Assuming noodles and oyster sauce are complements, a decrease in the price of noodles will:
- **a** decrease the demand for noodles.
- **b** increase the demand for noodles.
- **c** increase the demand for oyster sauce.
- **d** decrease the demand for oyster sauce.

4 Assuming cosmetics are a normal good, an increase in consumer income, other things being equal, will:
- **a** cause a downward movement along the demand curve for cosmetics.
- **b** shift the demand curve for cosmetics to the left.
- **c** cause an upward movement along the demand curve for cosmetics.
- **d** shift the demand curve for cosmetics to the right.

5 Yesterday, Gela supplied 300 cones per week at $5 per unit. Today, Gela supplies the same quantity of cones at $4 per unit. Based on this information, we can say that there is:
- **a** a decrease in supply.
- **b** an increase in supply.
- **c** an increase in quantity supplied.
- **d** a decrease in quantity supplied.

6 An improvement in technology causes:
- **a** a leftward shift of the supply curve.
- **b** an upward movement along the supply curve.
- **c** a downward movement along the supply curve.
- **d** a firm to supply a larger quantity at any given price.

Exhibit 3.19 **Supply and demand curves**



7 In **Exhibit 3.19**, at a price of $2, the market for good *X* will experience:
- **a** an excess supply of 100 units.
- **b** an excess demand of 200 units.
- **c** an excess supply of 200 units.
- **d** no excess demand or excess supply.

8 In **Exhibit 3.19**, the equilibrium price and output of good *X* are, respectively:
- **a** $2 and 100 units.
- **b** $1 and 200 units.
- **c** $2 and 200 units.
- **d** $0.5 and 150 units.

9 In **Exhibit 3.19**, at the market equilibrium output:
- **a** the consumer surplus is maximised.
- **b** the producer surplus is maximised.
- **c** the social surplus is maximised.
- **d** there is a social loss.

10 At a price of $10 per kilogram, the quantity supplied of beef is 100 units per week, whereas the quantity demanded for beef is 120 units per week. Given the above information, which of the following about the beef market is true?
- **a** The market is in equilibrium.
- **b** The market clearing price is below $10 per kilogram.
- **c** At a price of $10 per kilogram, there is an excessive supply in the market.
- **d** There will be upward pressure on the market price of beef.

Markets in action

4

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand how changes in demand and supply lead to changes in market equilibrium.
2 Answer the question: can the laws of supply and demand be repealed if governments control prices by implementing price ceilings or price floors?
3 Understand why government intervention may be necessary to correct market failure.

Getty Images/Auscape

Cyclone Yasi hit the northern Queensland coastline in February 2011. While there was minimal harm to people, banana farms were devastated. In the weeks that followed, bananas became a rare addition to the weekly shopping basket. Early in the COVID-19 pandemic, petrol prices dropped below $1 a litre – a level not seen for more than a decade in Australia. In this chapter, using the demand and supply model established in **Chapter 3**, we will discuss the impact of Cyclone Yasi on the banana market, as well as why petrol prices plummeted by more than 40 per cent in early 2020. Then we will outline why the laws of supply and demand cannot be repealed. Government policies to control markets have predictable and, at times, unforeseeable consequences. For example, we will explore what happens when the government limits the maximum rent landlords can charge, and who benefits and who loses from minimum wage laws.

In this chapter, we will also take a look at situations in which the market mechanism may fail. If you visited London at the turn of the last century, you would lament the smog that blanketed the River Thames and the beautiful surroundings. The 'Great Smog' was the result of burning great amounts of low-grade coal by households and power stations. Why did this and other pollution problems occur? And lastly, in Australia, the government offers free immunisation against measles and other contagious diseases for young children. Why free immunisation, but not free coffee for parents? Let us now start the journey of markets in action.

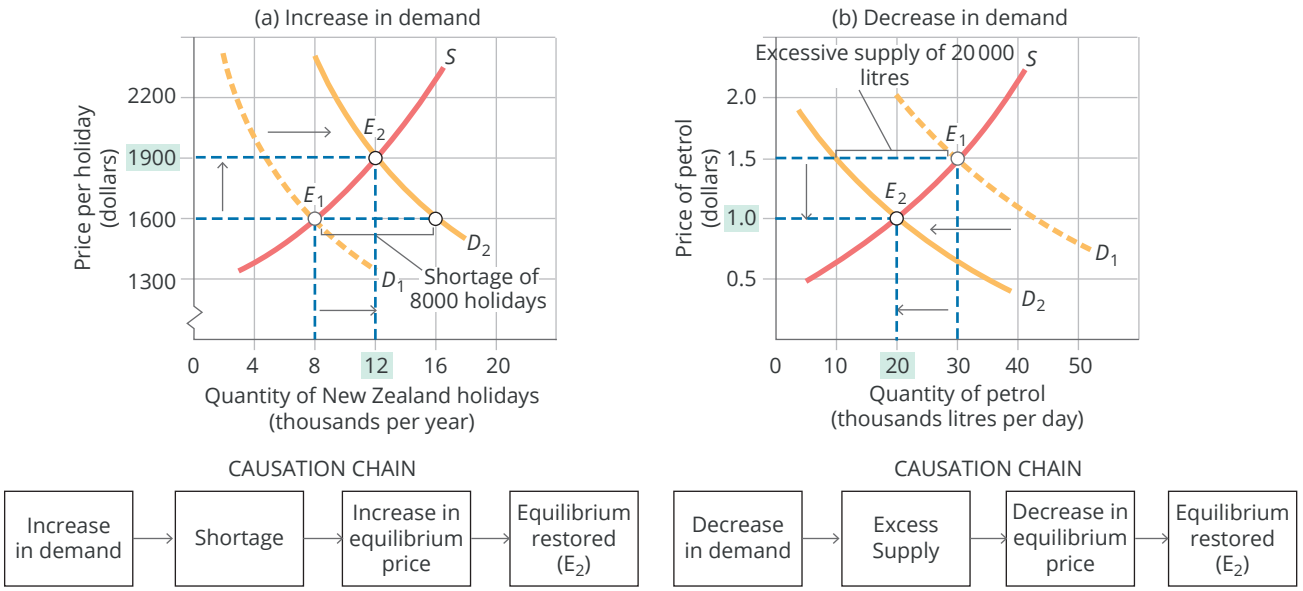
1 Changes in market equilibrium

Using market supply and demand analysis is like turning on the windscreen wipers in your car. Suddenly the fuzzy world in front of you becomes clear. The following examples will open your eyes and help you to see that economic theory has something important to say about so many things in the real world.

Changes in demand

The market for a particular type of New Zealand packaged holiday is shown in part (a) of **Exhibit 4.1**, where it is assumed that the position of the market supply curve, S, remains unchanged and that market demand increases from D_1 to D_2. Why has the demand curve shifted rightward in the figure? We will assume that the popularity of these holidays in New Zealand has suddenly risen sharply due to extensive advertising (changes in tastes and preferences). Given supply curve S and demand curve D_1, the initial equilibrium price is $1600 per holiday and the initial equilibrium quantity is 8000 holidays (holiday packages sold) per year, shown as point E_1. After the impact of advertising, the new equilibrium point, E_2, becomes 12 000 holiday packages per year at a price of $1900 each. Thus, the increase in demand causes both the equilibrium price and the equilibrium quantity to increase.

Exhibit 4.1 **The effects of shifts in demand on market equilibrium**



In part (a), there is an increase in the demand for New Zealand holiday packages because of extensive advertising and the demand curve shifts rightward from D_1 to D_2. This shift in demand causes a temporary shortage of 8000 holidays per year at the price associated with the initial equilibrium of E_1. This disequilibrium condition encourages firms in this segment of the holiday package industry to move upward along the supply curve to a new equilibrium at E_2.

Part (b) illustrates a decrease in the demand for petrol caused by the pandemic-induced recession. This leftward shift in demand from D_1 to D_2 results in a temporary excess supply of 20 000 litres of petrol per day at the price associated with the initial equilibrium of E_1. This disequilibrium condition forces petrol retailers to move down along its supply curve, reducing the petrol price to $1 a litre. Equilibrium is restored at E_2.

It is important to understand the force that caused the equilibrium to shift from E_1 to E_2. When demand initially increased from D_1 to D_2, there was a temporary shortage of 8000 holidays at $1600 per holiday. Firms in this segment of the packaged holiday business responded to excess demand (or shortage) by hiring more people, offering more

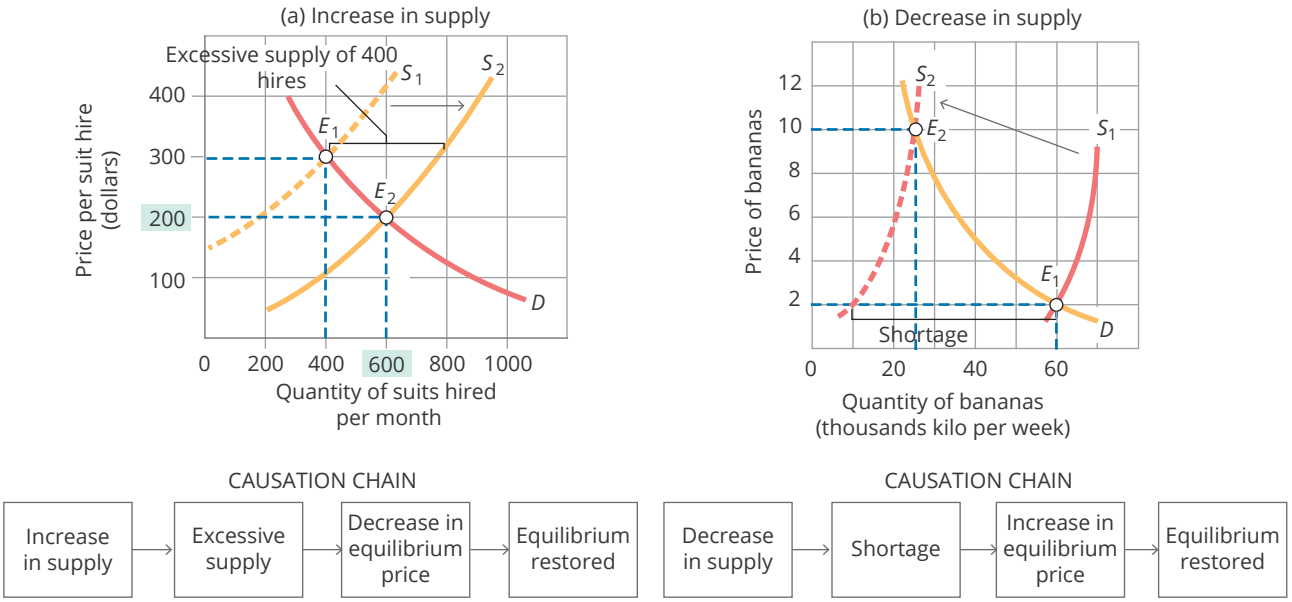
packaged holidays in New Zealand and raising the price. Tour companies will therefore move upward along the supply curve (increasing quantity supplied, but not changing supply). During some period of trial and error, sellers of packaged tours will increase their price and quantity supplied until a shortage no longer exists at point E_2. Therefore, the increase in demand causes both the equilibrium price and the equilibrium quantity to increase.

Part (b) of **Exhibit 4.1** displays what happened to the demand for petrol during the COVID-19 pandemic in Australia. Because a great part of the country was shut down due to lockdown measures, economic activities were greatly curtailed. As a result, factories and business activities came to a standstill, many people worked from home, and non-essential travel was banned, and there was a great fall in the use of both private and commercial vehicles. Demand for petrol fell, as shown by left shift of the demand curve from D_1 to D_2 in part (b) of **Exhibit 4.1**. At the initial equilibrium (E_1), with a price of $1.50 per litre, the quantity supplied now exceeds the quantity demanded by 20 000 litres per day. This unwanted inventory (excessive supply) forces petrol retailers to reduce the price and quantity supplied. As a result of this movement downward on the supply curve, market equilibrium changes from E_1 to E_2. The equilibrium price falls from $1.50 to $1.00 and the equilibrium quantity falls from 30 000 to 20 000 litres per day.

Changes in supply

Now assume demand remains constant, and allow some non-price determinant to shift the supply curve. In part (a) of **Exhibit 4.2**, begin at point E_1 in the formal suit-hire industry with an equilibrium price of $300 per hire and 400 hires per month. Suppose new firms enter the suit-hire market, market supply increases and the market

Exhibit 4.2 **The effects of shifts in supply on market equilibrium**



In part (a), begin at equilibrium E_1 in the formal suit-hire industry and assume that an increase in the number of suit-hire firms shifts the supply rightward from S_1 to S_2. This shift in supply causes a temporary excess supply of 400 hires per month. This disequilibrium condition causes a movement downward along the demand curve to a new equilibrium at E_2. At E_2, the equilibrium hire price declines and the equilibrium quantity rises.

In part (b), Cyclone Yasi causes the supply curve for banana to shift leftward from S_1 to S_2. This shift in supply results in a temporary shortage of 50 000 kilos per week. Competition between consumers for the available banana raises the price. As a result, the market moves upward along the demand curve to a new equilibrium at E_2 and the quantity demanded falls.

Exhibit 4.3 The effect of shifts in demand and supply on market equilibrium

Change	Effect on equilibrium price	Effect on equilibrium quantity
Demand increases	Increases	Increases
Demand decreases	Decreases	Decreases
Supply increases	Decreases	Increases
Supply decreases	Increases	Decreases

supply curve shifts from S_1 to S_2. This creates a temporary excess supply of 400 hires at the initial equilibrium price of $300. Formal suit-hire firms respond to having so many suits on their racks by reducing the hire price. As the price falls, hirers move down along their demand curve and hire more suits each month. When the hire price falls to $200, the market is in equilibrium again at point E_2 where consumers hire 600 suits per month.

Part (b) of **Exhibit 4.2** illustrates the market for bananas. Before Cyclone Yasi this market was at equilibrium E_1, the going price was $2 per kilogram, and 60 000 kilograms were bought and sold per week. In the summer of 2011 Cyclone Yasi destroyed two-thirds of banana farms in Queensland. This means the market supply curve shifted leftward from S_1 to S_2 and a temporary shortage of 50 000 kilograms of bananas occurred at the initial equilibrium price of $2 per kilogram. The market price subsequently rose from $2 to $10 per kilogram. The increase in price caused consumers to reduce their banana purchase. At the same time, in response to the higher price and more profit, suppliers of banana would divert bananas from other uses (e.g., producing banana chips) to the fresh banana market. A new equilibrium point was established at E_2, where 25 000 kilos were bought and sold.

Exhibit 4.3 gives a concise summary of the impact of changes in demand and supply on market equilibrium.

Analyse the issue — Applicable concepts: demand, supply and market equilibrium

Binge drinking and the market for alcohol

Binge drinking, which can be defined as irregular excessive drinking designed to get an alcohol fix, is different from alcoholism, which involves continual reliance on alcohol. It is also different from casual consumption of alcohol where there is not an excessive intake. Binge drinking – especially by young people – is an activity that governments around the world are increasingly seeing as problematic.

Getty Images/Zak Kaczmarek

Some attempts at a quantitative definition of binge drinking have also been made, with consumption of five standard drinks a night or more for men, and four standard drinks a night or more for women being common yardsticks.

Many of you may think that these quantitative definitions are wide of the mark because they treat 'having a few drinks' as a binge. Be that as it may, the reality is that drinking to excess to get an alcohol fix can be very harmful to drinkers themselves (liver damage; cancer of the mouth, oesophagus and bowel) and to others (excessive noise, motor vehicle accidents and family violence).

So, how have governments set about tackling the binge drinking problem?

A quick search on the internet reveals that the most popular approach involves education programs that target those young people who are, or are likely to become, binge drinkers. The most popular medium for these education programs has been television advertisements.

Another approach involves the idea of restricting access to alcohol. This may be done in a number of ways, including lockout laws whereby pubs and nightclubs have to close earlier, restrictions on the numbers of liquor outlets, raising the legal age at which people can purchase alcohol, and more stringent policing of existing restrictions.

A different and often more contentious approach is to raise the price of alcohol. This can be done, for example, by increasing manufacturers' and retailers' costs or by increasing the tax on alcohol. In early 2020, the Australian government raised the tax on beer by 1.2 percentage points, making it the fourth highest in the world at $2.26 per litre, and that is before the 10 per cent GST, according to the Brewers Association of Australia.[1] An earlier example of this policy approach can be found in the introduction in 2009 of a 70 per cent increase in the tax on premixed alcoholic drinks. These drinks, commonly known in Australia as alcopops, have spirits already mixed with soft drink in the can or bottle, and have been very popular with young drinkers.

What do you think?

1 Using demand and supply analysis, explain (with words and diagrams) the effect on alcohol consumption as a result of the following changes:
- **a** The government introducing an education campaign designed to discourage binge drinking.
- **b** The increase in beer tax in 2020.
- **c** The closure of pubs and nightclubs during COVID-19 restrictions (on the consumption of alcohol in these venues).
- **d** People spending more time at home during the COVID-19 crisis (on alcohol consumption at home). (To find out more about the impacts of lockdown restrictions during COVID-19 on mental health, read Moodie and Soller [2020].)[2]
- **e** Combining changes in (c) and (d), do you think overall the consumption of alcohol has gone up or down during this period?

2 The government introduces an additional tax on alcopops while keeping the tax on other alcoholic beverages at the current level:
- **a** Show the effect of the tax increase on price and quantity sold of alcopops.
- **b** Now show the effect of the alcopop tax increase on price and quantity sold of other alcoholic drinks for which the tax has remained unchanged.

[1] **Brewers Association, 'Excise duties on beer: Australia in international perspective' (Updated February 2020)', at https://www.brewers.org.au/read/420/excise-duties-on-beer-australia-in.html**

[2] **R. Moodie and T. Soller, 'Australia's COVID-19 relationship with booze', University of Melbourne, 2020, https://pursuit.unimelb.edu.au/articles/australia-s-covid-19-relationship-with-booze**

In summary

- When non-price factors change, demand and/or supply change, bringing about changes in the position of the supply and/or demand curves.
- Changes in market equilibrium brought about by changes in supply and/or demand enable us to see how new prices and quantities sold in the market are brought about.

2 Can the laws of supply and demand be repealed?

In some markets, the objective of governments is to prevent prices from rising to the equilibrium price. In other markets, the government's goal is to maintain a price higher than the equilibrium price. Although these government policies are often well-intentioned, when they intervene with the market, there are consequences that policy makers might not foresee. Market supply and demand analysis is a valuable tool for understanding what might happen when the government considers fixing prices. There are two types of price controls: *price ceilings* and *price floors*.

Price ceilings

Price ceiling
A legally established maximum price a seller can charge.

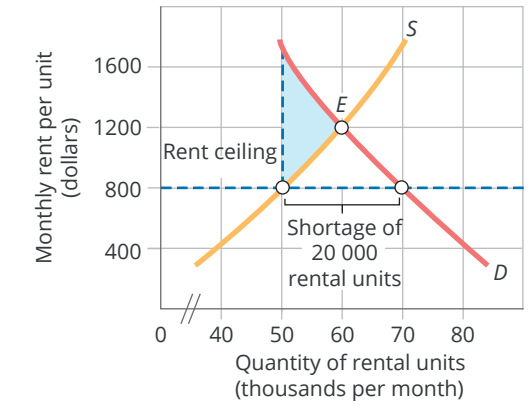
What happens if the government prevents the price system from setting a market price 'too high' by mandating a price ceiling? A **price ceiling** is a legally established maximum price a seller can charge. Rent controls are an example of the imposition of a price ceiling in the market for rental units. Many cities in the US have rent controls, which set a maximum price that private landlords can ask for the rental accommodation they own. In Australia and in many Asian countries, such controls over private landlords are rare. However, many of these countries do have rent control on government-provided housing, which is made available to tenants at subsidised rents well below the market equilibrium level. The rationale for rent control is to provide an 'essential service' that would otherwise be unaffordable to many people at the equilibrium rental price set in the market. Let's see why many economists believe that rent controls are counterproductive. In the following example we consider the private rental market. Many of the undesirable effects of rent control in this market can be shown to be equally applicable to rent controls applied in the government housing sector.

Exhibit 4.4 presents the supply and demand for rental units in a hypothetical city. Without rent controls the equilibrium is at point *E*, with a monthly rent of $1200 per month and 60 000 units occupied. Now, suppose the government imposes a rent control (price ceiling) designed to make housing more affordable. Assume the controlled rent is $800 per month. What does market supply and demand theory predict will happen? At the low rent ceiling of $800, the quantity demanded of rental units will be 70 000 while the quantity supplied will be only 50 000. Consequently, the price ceiling creates a persistent market shortage of 20 000 rental units. The shortage will persist because the rental price cannot rise to clear the market.

What might be the impact of rent controls on consumers and landlords? At the controlled rent, the quantity of rental units supplied and traded is reduced to 50 000 rental units, which is below the equilibrium output of 60 000 rental units. As discussed in **Chapter 3**, in a free market the equilibrium output is efficient since resources are utilised to create the greatest possible social surplus. The rent control reduces output to 50 000 rental units resulting in social loss represented by the blue-shaded area in **Exhibit 4.4**. Moreover, consumers must spend more time on waiting lists and searching for housing, as a substitute for paying higher prices. This means consumers incur an *opportunity cost* added to the $800 rent set by government. Considering both monetary and non-monetary costs (in terms of waiting and searching), is the government successful in reducing the cost of housing by imposing a price ceiling? Furthermore, an illegal market, or *black market*, can arise because of the shortage. Since some house seekers are willing to pay at a price above the controlled price, a profit motive exists that encourages tenants to risk breaking the law in order to sublet the unit to the highest bidder over $800 per month.

From the seller's perspective, rent control encourages two undesirable and unintended consequences. First, a controlled low rent may mean that landlords skimp on maintenance with the result that the stock of decent, well-maintained rental units will decline in the long run. So, *quality adjusted*, the price of housing can rise above the controlled level. Second, landlords may use discriminatory practices to replace the price system. Once owners realise that there is an excess demand for accommodation at the controlled price, they may resort to preferences based on family size, race and so on in order to determine how to allocate scarce rental space. Overall, rent control results in housing shortage without really bringing down the full cost of housing to needy households.

Exhibit 4.4 **Rent control results in a shortage of rental units**



CAUSATION CHAIN

Rent ceiling → Quantity demanded exceeds quantity supplied → Shortage

If no rent controls exist, the equilibrium rent for a hypothetical rental unit is $1200 per month at point *E*. However, the government imposes a rent ceiling of $800 per month, and a shortage of 20 000 rental units occurs. The quantity of rental units traded will be reduced to 50 000 units, creating a social loss (blue-shaded area). Because rent cannot rise by law, one outcome is that consumers must spend more time searching for an available unit. Other outcomes include a black market, bribes, discrimination and other illegal methods to deal with a shortage of rental units.

It should be noted that a rent ceiling at or above $1200 per month would have no effect as market forces will push it down to the equilibrium level of $1200 – a ceiling is only effective if it is below the market equilibrium price.

If governments accept economists' arguments against the use of rent controls, is there some alternative way to make housing available to targeted groups at a lower cost than the equilibrium rent? Many economists would argue for a system of welfare payments to low-income earners to enable them to pay market rents, leaving it up to the individual to choose how and when the welfare benefit might be spent. Other economists, who are concerned that welfare benefits may not be spent on housing, might propose a voucher system that would enable low-income earners to trade a permit issued by government for rental accommodation at market-determined rents. Such a scheme would ensure that recipients used their benefit for housing and also that the market for housing could achieve equilibrium.

Do you know what policy the Australian government use to provide affordable housing to poorer families? Go to the Centrelink website and do a key word search for *rent assistance* to find out.

Price controls on many goods and services, particularly those provided by government, were a common feature of many market economies throughout most of the 20th century. In recent years, however, governments, acting on the advice of economists, have pursued policies such as deregulation and privatisation of government-owned enterprises. These policies have increased efficiency by significantly reducing the extent of market distortions associated with government controls on prices. In Australia, these policies have been part of a set of wide-ranging adjustments known as microeconomic reform.

Nonetheless, in some areas where prices are controlled, there is still a long way to go. For example, urban water shortages brought about by drought conditions can be seen as the result of a de facto price ceiling. During drought, the supply curve of water shifts to the left, raising the equilibrium price of water in the process. If, when this happens, government bodies do not raise prices, there will be an excess demand for water that gives rise to rationing measures, such as bans on use of hoses, washing of cars and so on. Economists would much prefer that

higher prices be used to ration water rather than mandatory restrictions on the behaviour of householders. In this way, consumers who place great value on the availability of water (e.g., keen gardeners) can buy the water they want, while others can choose to purchase less.

You're the economist

How can better pricing policies help solve our infrastructure problems?

With strong growth in the Australian economy during the first two decades of this century, the demands on infrastructure provided by governments have increased enormously. Power outages, water shortages and heavily congested roads are just some of the manifestations of this tendency. Many solutions have been suggested for these problems, including calls for state governments, which are responsible for most of these services, to fund massive increases in public infrastructure. There might be political reasons for governments funding more infrastructure. However, a policy of facilitating large increases in infrastructure is fraught with dangers – not the least of which are the adverse environmental effects (known as negative externalities, which we will discuss in the next section) that accompany construction of capital assets, such as power plants and highways.

To reduce gridlocks in major capital cities, in 2019 Infrastructure Australia called on the government to invest more on roads and rails.[1] Rather than proposing increases in infrastructure as a solution to the problem, many economists suggest that an increase in the prices of these services is what is required. They say that in many cases the problems would be solved if prices were increased, but only during certain seasons, at certain times of the day or on certain days of the week. Thus, electricity prices could be higher when peak load occurs in the afternoon or evening, and charges for toll roads and tunnels could be raised during weekday peak periods. Such a policy would echo the private sector where price increases during high demand times are common. You probably know that Uber has surge pricing for busy periods, and that holiday rents are nearly always higher during peak periods, such as school holidays.

As far as road pricing goes, economists also point out that many of Australia's new urban roads, bridges and tunnels are taking less traffic and having a lesser effect on diversion of traffic from existing congested roads than is possible because these new urban infrastructures are subject to user-pay tolls that discourage demand.

In relation to tolls discouraging use of new road infrastructure designed to, among other things, reduce traffic congestion, is there a different pricing approach (apart from time of year, day or week pricing) that could help solve the problem?

[1] S. Wright and E. Bagshaw, 'Sydney, Melbourne coming to a standstill as infrastructure struggles', *Sydney Morning Herald*, 13 August 2019, https://www.smh.com.au/politics/federal/sydney-melbourne-coming-to-a-standstill-as-infrastructure-struggles-20190812-p52g9n.html

Price floors

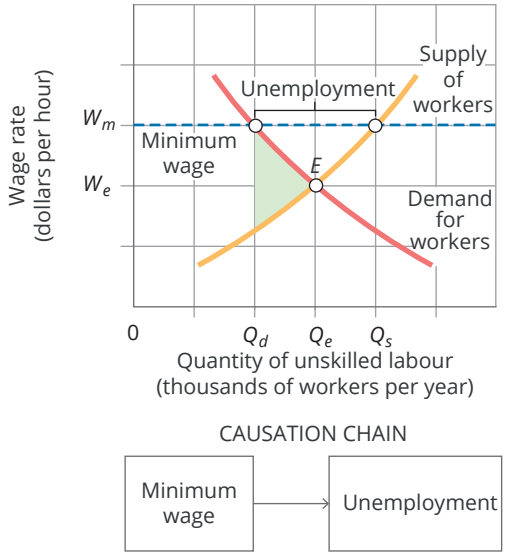
Price floor
A legally established minimum price a seller can be paid.

The other side of the price-control coin is that the government sets a **price floor** because it fears that the price system may establish a price viewed as 'too low'. A price floor is a legally established minimum price a seller can be paid. There is a long history of the use of price floors for agricultural products around the world. Even in the European Union where long-standing price floors for most agricultural products have been abandoned, major commodities, such as dairy products, beef, veal and sugar, still have guaranteed floor prices.

Globally, one of the most common price floors is the minimum wage. In recent years, in Australia, minimum wages have been in the news for all the wrong reasons. There have been many cases of low skilled workers – particularly those employed by franchisees in service jobs – who have been paid at rates well below the legislated minimum. Not only has this behaviour resulted in hefty fines being imposed on employers but it has also resulted in them having to pay compensation to workers who have been underpaid.

We begin our analysis of minimum wages by noting that the demand for unskilled labour is the downward-sloping curve shown in **Exhibit 4.5**. The wage rate on the vertical axis is the price of unskilled labour, and the amount of unskilled labour that employers are willing to hire varies inversely with the wage rate. At a higher wage rate, businesses will hire fewer workers. At a lower wage rate, they will employ more workers.

Exhibit 4.5 **A minimum wage results in a surplus of labour**



If the central wage-fixing authority sets a wage floor above the equilibrium wage, a surplus of unskilled labour develops. The supply curve shows the number of workers offering their services per year at each possible wage rate. The demand curve represents the number of workers employers are willing and able to hire at various wage rates. Equilibrium wage W_e will be the result if the price system is allowed to operate without government interference. At the minimum wage of W_m there is a surplus of unemployed workers ($Q_s - Q_d$). It is noted that if the minimum wage rate is set below the equilibrium wage rate, the forces of supply and demand for labour will establish the equilibrium wage, W_e regardless of the minimum wage legislation. Therefore, a minimum wage set at or below the equilibrium wage rate is ineffective.

On the supply side, the wage rate determines the number of unskilled workers willing and able to work per year. At higher wages, more workers are willing to give up leisure to go to work while lower wages will result in fewer workers available for hire. The upward-sloping curve in **Exhibit 4.5** represents the supply of labour.

Assuming the freedom to bargain, the price system will establish an equilibrium wage rate of W_e and an equilibrium quantity of labour employed of Q_e. But suppose a central wage-fixing authority sets a minimum wage, W_m, which is a price floor above the equilibrium wage, W_e. The intent of this policy is to make lower-paid workers better off, but consider the undesirable consequences. One result of an artificially high minimum wage is that the number of workers willing to offer their labour increases upward along the supply curve to Q_s while the number of workers that firms are willing to hire decreases to Q_d on the demand curve. The predicted outcome is an excess supply of unskilled workers, $Q_s - Q_d$. *Excess supply of labour* is just another term for unemployment. The minimum wage reduces the quantity of workers hired to Q_d, resulting in social loss (green-shaded area in **Exhibit 4.5**). The minimum wage is counterproductive because employers tend to lay off the lowest-skilled workers who are, ironically, the type of workers minimum wage legislation intends to help. Moreover, in the longer term, employers are encouraged to substitute machines for the unskilled labour previously employed at equilibrium wage W_e. If more income equality is the policy objective, it would be far better, according to many economists, to dispense with minimum wages and use the welfare system to top up the market-determined wages of low-income earners. This issue will be further discussed in **Chapter 13**.

Supporters of the minimum wage are quick to point out that those employed (Q_d) are better off. They also point out that even though the minimum wage may cause a reduction in employment, a more equal or fairer income distribution is worth the loss of some jobs. Moreover, the slope of the labour demand curve may be much steeper than is shown in **Exhibit 4.5**. If this is the case, the unemployment effect of a rise in the minimum wage would be small. It is also argued by advocates of minimum wage laws that the incentive they give employers to substitute capital for labour may have the long-term effect of raising the productivity of workers, and with it the capacity of employers to hire more workers and pay them higher wages in the long term.

The debate in relation to the minimum wage is far from settled. The above discussions illustrate the complexity of, and opposing views of economists on, these social and public issues, some of which will be further discussed in this book.

You're the economist — Sail price bargains

Although the age of sail is long gone, square-rigged sailing ships are still used for training purposes. In fact, many of the world's governments subsidise national sail training vessels that travel the world under their country's flag. In Australia's case, our training vessel is the *Young Endeavour*, which was a gift to the Australian people from the British government to celebrate the bicentenary of white settlement in 1988. This 44-metre-long vessel, which is operated by the Australian Navy, provides sail training for young Australians who can apply to join the vessel as crew to make ocean voyages.[1]

Australians aged between 16 and 23 years are invited to join the vessel to experience and endure the challenges at seas for days. In 2021 these voyages were not cheap, with the standard 11-day youth development voyage priced at $1950. Fee-paying crew members are also expected to fund their own travel to the points of departure and arrival of the vessel for each leg. Expecting high demand for places on each of these legs, the organisers advise applicants that a ballot will be held to determine which of the eligible applicants would be awarded a place on board.

Given that a ballot is held to determine which of the applicants would be awarded a place, are the prices charged akin to a price ceiling, a floor price or a price close to the equilibrium price?

Australian Government Department of Defence

[1] To read about the *Young Endeavour* and its voyages, go to the Young Endeavour Youth Scheme at http://www.youngendeavour.gov.au

In summary

- **Price ceilings** and **price floors** are maximum and minimum prices enacted by law to prevent the forces of supply and demand determining prices in the market.
- If a price ceiling is set below the equilibrium price, a shortage will persist. If a price floor is set above the equilibrium price, an excess supply will persist. In both cases, social loss occurs.

3 Market failure

In this and the previous chapter you have gained an understanding of how markets operate. The price system coordinates economic activities, bringing about efficient market outcomes. But markets do not always produce efficient outcomes from society's point of view. It is now time to step back with a critical eye and consider markets that produce socially unacceptable results, which economists call **market failure**. Market failure means that the price system fails to operate efficiently, creating a problem for society. In this section, you will study three types of market failure: externalities, public goods and lack of competition. Also, in this section is a discussion of an ethical aspect of the price system – income inequality, which is seen by some economists as another type of market failure.

Market failure
A situation in which the price system fails to operate efficiently, creating a problem for society.

Externalities

Some markets may fail because they suffer from the presence of side effects economists call **externalities**. An externality is a cost or benefit imposed on some *third parties* – people other than the consumers and producers of a good or service. Externalities are also called *spillover effects* or *neighbourhood effects.* Externalities may be either negative or positive; that is, they may be detrimental or beneficial to the third party.

Externality
A cost or benefit imposed on people other than the consumers and producers of a good or service.

Suppose you are trying to study and your neighbour is listening to Beyoncé at full blast. The action of your neighbour is imposing an unwanted *external cost* or *negative externality* on you and other third parties who are trying to study or sleep. Because it results from your neighbour's consumption of music, this is an example of a *consumption externality*. Externalities can also result in an *external benefit* or *positive externality* to non-participating parties. When householders proudly display their beautiful gardens and freshly painted homes, neighbours and visitors are third parties who did none of the work but enjoy the benefit of the pleasant scenery. Because it results from your neighbour's consumption of beautiful gardens and homes, this is another example of a consumption externality, but this time *positive externality.*

Let's now look at externalities arising from production. When some of the benefits of a firm's research and development (R&D) effort are captured by other firms, this is an example of positive *production externality*. As is discussed in **Chapter 12**, this externality is considered to be an important factor contributing to economic growth. Australia's National Broadband Network (NBN) was not expected to be profitable in financial terms. Nonetheless, the government proceeded with this project because it believed that the positive externalities that it would bring would more than compensate for any financial losses. These external benefits, which flow from the increased connectedness of the community, include increases in business productivities, improved educational outcomes and better delivery of health services.

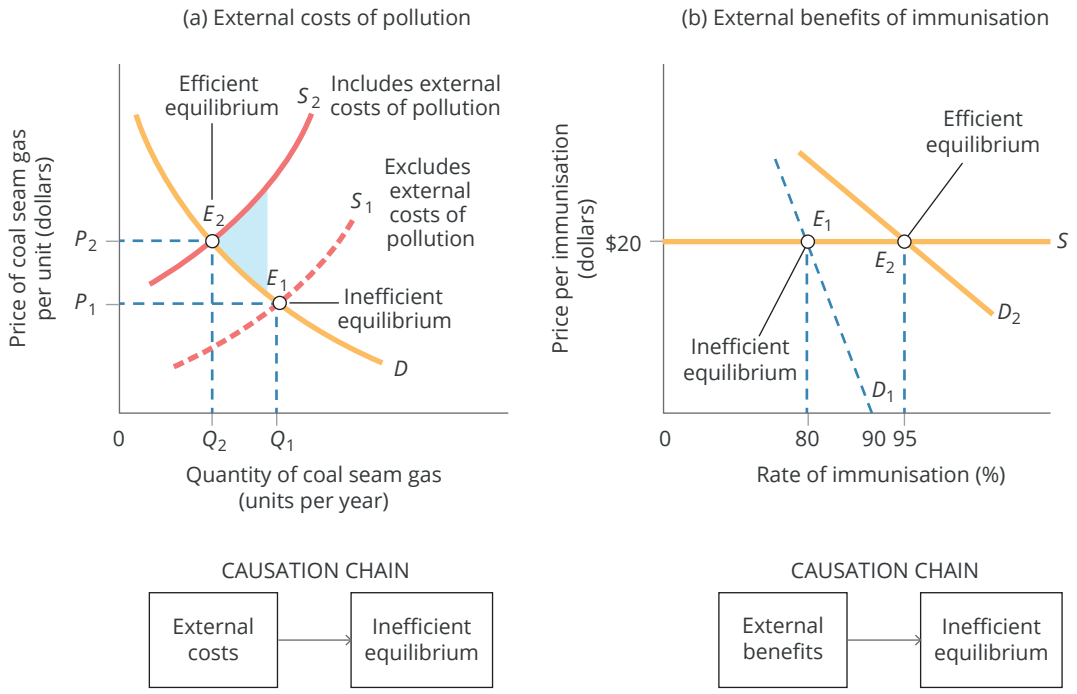
Why will externalities result in market failure? This will become clear in the following analysis of two examples of externalities – one negative, the other positive. First, drilling for coal seam gas, which has become a contentious issue in Australia and many other countries as a potential source of negative externalities, hits the headlines regularly. The adverse effects of this activity include encroachment on farmers' properties with consequent loss of productivity, contamination and wastage of groundwater, and blighting of the landscape. These are significant negative production externalities that have caused state governments in Australia to rethink their policy approaches to drilling for coal seam gas. The second externality is the massive reduction in disease and death that accompanies widespread immunisation. This is a positive consumption externality. Using demand and supply analysis, these two externalities are analysed graphically in **Exhibit 4.6**. We look first at drilling for coal seam gas.

A graphical analysis of coal seam gas drilling

Part (a) of **Exhibit 4.6** shows the market for gas produced from underground coal seams. Collectively, the drilling, fracking and transport of gas from these coal seams results in environmental damage. The demand curve D and supply curve S_1 establish the normal market equilibrium E_1 in the market for coal seam gas. This is an inefficient outcome because no account is taken of the negative externalities produced by this market. Recall the supply

curve tells us the cost of production of each unit of output (marginal cost). However, the supply curve S_1 only tells us the costs to mining companies – known as private costs in economics, which include costs of fuel, labour and machineries. Not included in S_1 are the *external costs* to the public in terms of loss of business by farmers, destruction of wildlife and clean-up costs. Therefore, to the society (which includes the affected third parties) the total cost of production of coal seam gas is the sum of private and external costs. This total cost is known as the *social cost*. The social cost of producing coal seam gas is represented by the supply curve S_2. Note that S_2 is above S_1. Since the supply curve, S_2 includes the full costs of coal seam gas drilling, the intersection of S_2 and D (E_2) will represent the efficient level of coal seam gas production (Q_2).

Exhibit 4.6 **Externalities in the coal seam gas and immunisation markets**



In part (a), resources are over-allocated at inefficient equilibrium E_1 because coal seam gas companies do not include the external cost per unit of gas produced. Supply curve S_2 includes the external costs of gas produced. If firms are required to purchase equipment to reduce groundwater contamination, pay higher fees to governments for the right to drill or guarantee full compensation to adversely affected third parties, the market can move towards the efficient equilibrium of E_2.

Part (b) demonstrates that external benefits cause an under-allocation of resources. D_1 represents parents' demand for immunisation without considering the external benefit. Adding the benefits of immunisation to the whole society (external benefits) D_2 represents demand for immunisation from the society's perspective. Note D_2 is flatter than D_1. This is because at very low rates of immunisation the external benefits of immunisation are low because not enough people are vaccinated to ensure herd immunity. At high rates of immunisation, where herd immunity is highly probable, the external benefits of immunisation are high. At equilibrium point E_2, the efficient output of immunisation services is obtained if the government uses subsidies and/or legislation to incentivise parents to have their children immunised.

In a free market, however, producers only consider the costs of production they must bear (private costs), but not costs to the third parties (external costs). Therefore, the free market equilibrium output occurs at Q_1. For output between Q_2 and Q_1, the social cost (represented by S_2) will be above the benefit to the consumers (represented by the demand curve D, as discussed in **Chapter 3**). In other words, the production of gas at between Q_2 and Q_1 represents destruction rather than creation of value. As a result, the society as a whole will be worse

off; the social loss is represented by the blue-shaded area. In short, the presence of negative externalities in the production of coal seam gas means that the firms produce more output than is socially desirable. Economists call this market failure.

If the free market fails to motivate producers to restrict their output to the efficient equilibrium level of Q_2, how can we tackle the environmental problem of negative externalities, such as coal seam gas extraction and burning of fossil fuels?

There are a number of ways in which the market failure associated with coal seam gas production can be addressed. These include increased taxes and charges levied on producers, regulation of production processes and a requirement for firms to pay full compensation to affected parties.

1 *Increased taxes and charges* – this approach involves the government levying taxes or charges that increase drillers' costs and inhibit production by imposing an additional production cost per unit of gas produced. This shifts the supply curve leftward from S_1 towards S_2, moving the market towards the efficient equilibrium point E_2.

2 *Regulation* – an alternative approach is to legislate standards which force firms to significantly reduce the probability, for example, of groundwater depletion and contamination. This means firms must use different procedures or install costly equipment to reduce the likelihood of such outcomes. When the extra costs of these procedures and equipment are added to the production cost per unit of output, the initial supply curve S_1 shifts leftward towards S_2 and towards the efficient equilibrium of E_2. In some instances, regulation may involve the banning of coal seam gas extraction altogether, as has occurred in New South Wales where drilling is prohibited adjacent to towns and city suburbs. The implication of this latter approach is that the potential external costs of coal seam gas production in these areas are so high that the efficient level of output is zero.

3 *Payment of full compensation* – the firms can be required to fully compensate adversely affected parties for the incurred damages. To the extent that these damages can be actually measured, this method would have the effect of imposing the costs of negative externalities directly on the company responsible for them. In the language of economics, the externality would be internalised. You can see that when firms build this necessity to pay compensation into their overall cost structure this will increase their costs and move the supply curve towards S_2 and towards the efficient equilibrium of E_2. Furthermore, this need to guarantee compensation would encourage producers to install additional equipment to help minimise adverse effects. Can you explain why this is sure to happen when compensation costs are likely to be high but the costs of minimisation of adverse effects are low?

In **Chapter 10** a more comprehensive analysis of issues surrounding the treatment of negative externalities is discussed in the context of the abatement of carbon emissions, which are strongly implicated in climate change. We look next at the issue of positive externalities.

A graphical analysis of immunisation

As explained above, the supply curve can understate the *external costs* of a good or service. Now you will see that the demand curve can understate the *external benefits* of a good or service. One of the marvels of modern medicine is that it is possible to vaccinate against many life-threatening diseases, including polio, measles and whooping cough. This is best done through the immunisation of children – usually in the first two years of life. To achieve community protection from diseases like measles (i.e., herd immunity), at least 95 per cent of the population must be vaccinated. Can we rely on the free market to achieve this rate of vaccination? Part (b) of **Exhibit 4.6**, illustrates the market for immunisation against measles, where the horizontal axis measures the percentage of children vaccinated. The supply curve S shows the marginal cost of vaccination, which is assumed to be constant, that is, the additional cost of a vaccine is the constant at $20. Demand curve D_1 reflects the price consumers (i.e., the parents of children) would pay for immunisations to receive the benefit of a reduced probability of infection. However, very few of these parents would consider the benefit of immunisation to the whole society – the benefit of herd immunity (i.e., immunity for all). Demand curve D_2 includes this external benefit of vaccination. Note

that D_2 is above D_1; the vertical distance between D_2 and D_1 represents the external benefits of immunisation. From the social perspective, the efficient level of immunisation is 95 per cent, intersection of the supply curve S and demand curve D_2.

In the free market, since parents typically consider only the benefit of vaccination to their children (private benefit), parents at the lower end of the demand curve D_1 will choose not to vaccinate their children, since to them the value from vaccination is less than the price of $20, and these parents tend to be poor or are those who do not believe in vaccination. As shown in part (b) of **Exhibit 4.6**, the free market equilibrium vaccination rate, determined by the supply curve S and demand curve D_1, is 80 per cent, and is lower than the efficient level of 95 per cent to achieve herd immunity. Market failure occurs. What can the government do to increase immunisation rate to 95 per cent?

To prevent market failure in the market for immunisation, there are two approaches:

- *Subsidies* – the Australian government has, for a long time, subsidised immunisation by making it available free of charge. Although this means that immunisation rates are far higher than they would be if parents were required to pay the full market price, declining immunisation rates in recent decades have led to the view that not enough children are being immunised. This is illustrated in part (b) of **Exhibit 4.6**. Focusing on D_1, at zero price the immunisation rate of 90 per cent still falls short of the 95 per cent required to achieve herd immunity. One possible solution would be for the government to make a payment to parents for each child immunised. Since the subsidy amount is payable at any price along the demand curve, the demand curve in part (b) of **Exhibit 4.6** shifts to the right. If the subsidy is sufficiently large, the demand will shift to a position somewhere between D_1 and D_2 to achieve the required immunisation rate of 95 per cent. (Can you add this demand curve to **Exhibit 4.6[b]**?)
- *Regulation* – the government could boost immunisation levels by passing laws requiring all parents to immunise their children. This authoritarian approach to the problem is, however, difficult to pursue in a thriving democracy. (Witness the protests against lockdowns during COVID-19 in many western democracies.) This explains why Australian governments have tried other ways to raise immunisation rates. For example, the Australian government restricts access to government benefits by families whose children have not been immunised. In 2016 the Australian government introduced the 'No Jab, No Pay' law, whereby families are not entitled to their full family benefits if their children are not immunised. These policies do not make immunisation compulsory, but a decision not to immunise carries with it the high cost of loss of government benefits. This high cost results in the required rightward shift of the demand curve that might otherwise be brought about by a decision to make immunisation compulsory or by the payment of a subsidy.

Global perspective

Applicable concepts: market failure, positive externalities, and consumer sovereignty

Should globetrotting anti-vaccination advocates be allowed to visit Australia?

After many decades of universal support for immunisation, in recent years there has been a significant worldwide increase in the number and visibility of its opponents. Sometimes described as 'anti-vaxxers', these opponents of immunisation form an international community that tries to marshal global support for their cause in the name of freedom of choice. Activists in Australia include some medical practitioners on the fringe of their profession as well as prominent political leader Pauline Hanson.

Alarmed by the growth in this movement, governments, public health experts, doctors and ordinary citizens are joining in a cacophony of voices aimed at educating the public at large about the benefits of immunisation while at the same time endeavouring to expose the misleading nature of information peddled about its risks. Perhaps the most notorious of these latter claims is the discredited research that purports to show that certain childhood vaccinations

can lead to autism. Knowing that immunisation only protects the wider population – giving so-called herd immunity – if a very high percentage of individuals are vaccinated, immunisation advocates worry that antipathy and apathy can lead to rates of immunisation below required. In economic terms they know that immunisation provides enormous positive externalities that may be lost if immunisation rates fall.

Alamy Stock Photo/Robert Wallace/ Wallace Media Network

An example of the costs of falling rates of immunisation played out in the US when a record number of measles cases was reported in 2014, just 14 years after the country was declared measles free in 2000. In the vast majority of cases these outbreaks occurred in communities with much lower than average immunisation rates.

So seriously does the Australian government take the risks of falling immunisation rates that in August 2017 it denied entry to prominent US anti-vaxxer Kent Heckenlively, who travels the world discouraging parents from having their children vaccinated.[1] It seems that for the common good, governments are happy to override consumer sovereignty to ensure that positive externalities are reaped and market failure does not occur.

Protest against public health policies had also been the focus of many heated debates during the coronavirus pandemic. In Australian the state governments imposed various levels of restrictions to prevent transmission of the virus. As a result a lot of our daily routines were affected. Those restrictions ranged from limitations on the number of patrons to a restaurant, the number of spectators in a football stadium, the number of guests allowed to attend a wedding ceremony or funeral, and reduced duration and space of outdoor exercising. To contain the outbreak of the second wave of infection in Melbourne in mid-November 2020, the Victorian government completely locked down the city for a short period. The protests against these restrictions in Australia were relatively peaceful compared to other parts of the world. In the US and Europe many such protests led to serious confrontations with police and even riots. Those against lockdowns argued that such restrictions were not necessary for the majority of the population. For example, they claimed that young children and adults faced more risk from flu than COVID-19, many suffered mental stresses due to social isolations and urgent medical procedures were delayed during prolonged lockdowns. A group of human right advocates and health experts in the UK organised an anti-lockdown petition, the 'Great Barrington Declaration', endorsed by prominent public health experts and research scientists urging policymakers to take a focused approach to protect the most vulnerable instead of lockdowns.[2]

What do you think?

1 Consider the demand for vaccinations and answer the following:
- **a** Draw a community demand curve for vaccinations where no external benefits are accounted for; label it D_1. Now draw a second demand curve that takes into account the external benefits of vaccination; label it D_2. How does the slope of this second curve differ from that of the first curve?
- **b** Imagine that a strong anti-vaccination movement develops and prospers with the help of anti-vaxxers such as Kent Heckenlively. Suppose that, as a result of this development, at each possible price, demand for vaccinations falls to half its previous level. On your diagram draw the new demand curve for vaccinations where no account is taken of external benefits; label it D_1^*. Where does it sit in relation to the original demand curve D_1?
- **c** Now consider the combined effects of the rise of anti-vaxxers and government's 'No Jab, No Pay' legislation and draw the demand curve for vaccinations, D_1^* assuming the overall effect is such that the immunisation rate is raised to the required 95 per cent to achieve herd immunity.

2 Consider the social cost of human activities during the COVID-19 pandemic. We all appreciate being able to freely run our daily life (work, shopping, jogging, socialising, going to church, sports event etc.), but our daily activities follow the law of diminishing utility. That is, the marginal benefit from our daily activities declines as we do more

of them. How about the costs of running our daily activities? During normal times, in terms of risks of spreading diseases, the social cost is almost zero, regardless of the number of hours we spend in public places. However, during the coronavirus pandemic, the health risk is considerable; moreover, the social cost of this risk follows the law of increasing marginal cost. That is, if people all spend little time in public places, the health risk is small. But the health risk will increase substantially with the rise of human interactions when we spend more time in public places during a pandemic. Now answer the following questions:

a Model the marginal cost and marginal benefit of human activities in public places. The horizontal axis represents the number of hours of daily activities in public places (for simplicity, zero hours means complete lockdown and 24 hours complete freedom). And the vertical axis represents the cost or benefit of such daily activities. Draw the social marginal cost and marginal utility curves for daily activities in public places (i) during normal times with no risk of infectious diseases, and (ii) during the pandemic.

b During the pandemic, what is the optimal duration of daily activities if you are the Victorian government in the second wave of the pandemic? What if you are one of those petitioners who are strongly against lockdowns? (*Note:* The shorter the duration, the stricter is the restriction.)

[1] S. Peatling, "'World's number 1 anti-vaxxer" Kent Heckenlively denied entry to Australia', *Sydney Morning Herald*, 31 August 2017, https://www.smh.com.au/politics/federal/worlds-number-1-antivaxxer-kent-heckenlively-denied-entry-to-australia-20170831-gy7u82.html

[2] Read the following article to learn more about the Great Barrington Declaration: 'Coronavirus: Health experts join global anti-lockdown movement', *BBC News*, 7 October 2020, https://www.bbc.com/news/health-54442386

Public goods

Public good
A good or service that is non-rival and non-excludable.

Most goods and services are traded in markets, but not for goods and services such as national defence, police and streetlights. Have you wondered why? National defence is a classic example of a **public good**. A public good has two special properties: (1) the consumption of the good by one user will not reduce its benefit to other users – a public good is non-rival, and (2) it is not possible to prevent people who do not pay (free riders) from consuming the good – a public good is non-excludable.

In most cases, if a good or service is a public good, the government, rather than the market, must provide if it is to be made available in sufficient quantity. To see why the marketplace fails in relation to provision of a public good, imagine that a company called C-Surveillance commences patrols of the Australian coast, and offers to sell their coast watch services to people who want private protection against entry of illegal immigrants into Australia. First, once the system is operational, Mr Jones benefits from a lower incidence of illegal immigration without reducing the benefits to Ms Smith. Second, the *non-exclusive* nature of coastal surveillance means that it is impossible or very costly for any purchaser of coast watch services to prevent non-purchasers, the free riders, from reaping the benefits.

Given the two properties of a public good, why would any private individual purchase a C-Surveillance coast watch service? Why not take a free ride and wait until someone else buys coastal surveillance? Each person therefore wants a surveillance system, but each person does not want to bear the cost of the system when everyone shares in the benefits. As a result, C-Surveillance will not get any business and the market fails to provide coast watch services, even though the social benefit of the services (i.e., total benefit of the services to the community) is greater than the provision cost. Governments can solve this public goods problem by providing coastal surveillance services and taxing the public to pay. Note that while a government typically funds public goods through taxation, there is no need for it to actually undertake their production. This task can be outsourced to the private sector. Unlike a private firm, the government can use force to collect payments and prevent the free-rider problem. Other examples of public goods include the judicial system, street lighting, maritime navigation markers and the quarantine service.

Lack of competition

So far, we have argued that a free market is efficient, in the absence of externalities. However, for it to work efficiently, the market must be competitive. That is, there must be competition among both producers and consumers. In 1776, Adam Smith's famous work on economics, *The Wealth of Nations*, was published. Although Smith praised the virtues of the market system, he was acutely aware of the damage that anti-competitive practices could inflict. Smith observed that 'people of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some diversion to raise prices'. This famous quote clearly underscores the fact that in the real world, businesses will seek ways to replace consumer sovereignty with 'big business sovereignty'. What happens when a few firms rig the market and become the market's boss? By colluding to reduce competition, these firms can restrict output to enjoy higher prices and profits, at the cost of the wellbeing of the community.

In most countries there are legislations to prevent anti-competitive behaviours. In Australia, the *Competition and Consumer Act 2010* (formerly the *Trade Practices Act*), which is administered by the Australian Competition and Consumer Commission (ACCC), outlaws collusive behaviour. One example involves a fire protection industry cartel in Queensland. A cartel is an explicit collusion to exploit consumers by rigging prices. The fire protection industry cartel was responsible for raising prices of alarm installations by 15 per cent and sprinkler installations by 10 per cent. The ACCC estimated that the effect of these higher prices was that members of the cartel could pocket an additional $5 million per annum at the expense of the building owners, their tenants and, ultimately, the consumer. Among the clients named as victims of the cartel were two large universities, an airport, an aged care home, a soft drink manufacturer and two shopping centres. As a result of successful action by the ACCC against them, firms involved in the cartel, and some of their employees, have paid fines totalling millions of dollars as a penalty for their illegal (and unethical) behaviour.

In 2007, giant packaging company Visy agreed to settle a landmark case brought against it by the ACCC that alleged price fixing with competitor Amcor. It was reported that Visy had agreed to pay record penalties of $40 million – an amount far short of the more than $300 million that this price fixing was estimated to have cost the customers of these packaging companies. Between them, Visy and Amcor were estimated to control more than 90 per cent of the corrugated fibreboard packaging market in which prices were fixed.

In November 2010, the European Commission announced that it would fine 11 well-known international airlines, including Qantas, a total sum of nearly 800 million euros for taking part in an air cargo cartel. The illegal activity was said to have occurred over a six-year period, between December 1999 and February 2006.

These examples all involve markets dominated by a few firms, which economists call oligopolies. The behaviour of firms in this and other types of markets will be examined at length in the next three chapters.

Free market and income inequality

Economics and ethics

In the cases of insufficient competition, externalities and public goods, the marketplace is inefficient because it allocates too few or too many resources to producing particular outputs. However, even when these market failures are absent and the market is operating efficiently, a free market may be instrumental in producing a very unequal distribution of income. Under the impersonal price system, top golfers, CEOs of large corporations and medical specialists can earn very high incomes, while unskilled workers and the disabled earn very low incomes – if they are able to earn an income at all. Not all economists agree that these differences in incomes constitute market failure. Some point to the strong incentive effects that income inequality provides. Nonetheless, ethical considerations mean that there is widespread support for a policy of reducing income inequality below the level that would prevail if governments did not intervene in the market. This support is evidenced by the widespread adoption in all developed countries of income redistribution mechanisms, such as progressive income taxes and payment of welfare benefits to those most in need. The 2016 election of Donald Trump in the US and the majority support for Brexit in the UK are also seen as containing

an element of protest on the part of ordinary citizens, whose incomes have failed to keep pace with those of the professional and business classes. But how equal should the distribution of income be, and how much government intervention is required to achieve this goal? Some people wish to remove most inequality of income. Others argue for the government to provide a 'safety net' minimum income level for all citizens while higher incomes are lightly taxed so that they continue to act as an incentive for people to develop skills, work hard and accumulate wealth.

Nobel Prize-winning US economist Joseph Stiglitz argues that increased inequality of income is associated with insecurity on the part of a nation's poorer citizens, and this insecurity reduces the willingness of these people to take the risks that are necessary for an economy to reach its potential. So, not only may a widening divide between rich and poor disadvantage the poor, but it may also disadvantage many others in society as the economy fails to reach its full potential.[1] Ethical issues surrounding the distribution of income are not only complex but they are also contentious.

[1] J. Stiglitz, 'Divide between US rich, poor widens', *Sydney Morning Herald*, 15 October 2014, p. 28.

In summary

- **Market failure** is a situation in which the market mechanism fails to achieve the efficient outcome to maximise social surplus. Sources of market failure include externalities, public goods and lack of competition. Although sometimes controversial, government intervention is a possible way to correct market failure.
- An **externality** is a cost or benefit of a good or service imposed on third parties who are not directly involved in the production or consumption of that good or service.
- A **public good** is a good or service that (1) is non-rival and (2) non-excludable. Due to the free rider problem, a public good may not be produced in a private market even though its benefit to the society is greater than its cost of production.
- Lack of competition occurs when firms in an industry can exploit consumers by charging higher prices than would prevail if the market was competitive.

Key concepts

Price ceiling | Price floor | Market failure | Externality | Public good

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand how changes in demand and supply lead to changes in market equilibrium

- Changes in market equilibrium, which are brought about by changes in the position of the supply curve and/or the demand curve, enable us to see how new prices and quantities sold in the market are brought about.

2. Answer the question: can the laws of supply and demand be repealed if governments control prices by implementing price ceilings or price floors?

- **Price ceilings** and **price floors** are maximum and minimum prices enacted by law to prevent the forces of supply and demand determining prices in the market. A price ceiling is a maximum price mandated by government, and a price floor is a minimum legal price.
- If a price ceiling is set below the equilibrium price, a shortage will persist. If a price floor is set above the equilibrium price, an excess supply will persist. Price controls often lead to undesirable and unintended consequences.

3. Understand why government intervention may be necessary to correct market failure

- **Market failure** is a situation in which the market mechanism operates inefficiently and, as a result, does not achieve the most desirable result of maximising social surplus. Sources of market failure include externalities, public goods and lack of competition.
- An **externality** is a cost or benefit of a good or service imposed on people who are not buyers or sellers of that good or service.
  - Pollution is an example of an external cost. Because this cost is not incurred by the polluting firm, it is using too many resources to produce too much of the product responsible for the pollution. Two basic approaches to solving this market failure are pollution taxes and regulation.
  - Immunisation benefits the whole society by preventing spread of some diseases (external benefits). However, because many buyers of immunisation services base their purchases on the benefit to themselves and ignore the benefits to the wider community, the amount of immunisation services purchased is too low. Two basic solutions to this type of market failure are special subsidies or laws making consumption compulsory.
- A **public good** is a good or service that (1) is consumed collectively (i.e., non-rival), and (2) cannot be consumed by one person without others being able to consume it, regardless of whether they pay (i.e., non-excludable). A public good may not be provided as there is no incentive for an individual to provide it.
- Lack of competition occurs when firms in an industry can exploit consumers by charging higher prices than would prevail if the market was competitive.
- Although sometimes controversial, government intervention (e.g., legislations, taxes and subsidies) is often employed as a possible way to correct market failure.

Study questions and problems

1 Market researchers have studied the market for rice. Their estimates for the supply of and demand for rice per month are as follows:

Price per kilo	Quantity demanded (millions of kilos)	Quantity supplied (millions of kilos)
$2.50	100	500
2.00	200	400
1.50	300	300
1.00	400	200
0.50	500	100

a Using the accompanying data, graph the demand for rice and the supply of rice. Identify the equilibrium point as *E* and use dotted lines to connect *E* to the equilibrium price on the price axis and the equilibrium quantity on the quantity axis.

b Suppose the government enacts a rice support price of $2 per kilogram. Indicate this action on your graph and explain the effect on the rice market. Why would the government establish such a support price?

c Now assume the government decides to set a price ceiling of $1 per kilogram. Show and explain how this legally imposed price affects the rice market. What objective could the government be trying to achieve by establishing such a price ceiling?

2 In Victoria under 'No Jab, No Play' legislation, children who have not been vaccinated are barred from attending childcare. Explain why doctors would support this policy, using appropriate economic concepts.

3 Using market supply and demand analysis, explain the effect on the market for alcohol of an increase in the legal drinking age from 18 to 21 years. What do you think might be the effect of such an increase on the market for illegal drugs?

4 An increasing number of world cities, including London, Singapore and Stockholm, are embracing the idea of a congestion charge that involves motorists paying a fee when they bring their vehicles into the busiest part of the city. Explain how such a charge is an example of the price system at work.

5 Thinking about your answer to question 4, how do you think the authorities would go about determining the level of the charge if they have a particular target for the total number of vehicles entering the congestion charge zone each day?

6 Net revenues from the London congestion charge are to be invested in the city's public transport system. Once this investment occurs, what are some of the likely effects on the level of the charge, *ceteris paribus*?

7 Using the characteristics of public goods – namely that (1) users collectively consume benefits, and (2) there is no way to prevent people who do not pay from consuming the good or service – explain why the Australian Quarantine Service is considered to be a public good.

8 Discuss the advantages and disadvantages of legislating for minimum wages in a modern democracy. Do you think that employers and employees might have different views about the subject?

Answers to 'You're the economist'

How can better pricing policies help solve our infrastructure problems?

As the population grows rapidly in Sydney and Melbourne, new infrastructures, such as roads, bridges and tunnels, are needed to cope with the increasingly congested cities. Recently, for efficiency reasons, the private sector has been more involved in these infrastructure projects. And to recoup their investment, these private construction firms are granted the right to charge their road, bridge and tunnel users a toll. The problem is that tolls discourage travel on these new roads that need to be highly patronised so that congestion on existing roads can be reduced. Some alternative ways for these constructors to raise revenue to recoup their investment would need to be devised if tolls were not imposed on new roads. An obvious solution would be to make travel free on new, congestion-busting roads while charging a toll on congested sections of existing roads. The tolls will increase patronage of those new roads, reduce that of existing congested roads, and at the same time the tolls can be collected to fund the construction of these congesting-busting roads. Of course, there would need to be many, sometimes complex, changes in government policy, legislation and regulation to allow this to happen. If you said that new private-sector-funded road infrastructure could involve contracts with government that gave construction firms the right to toll existing untolled roads but not the new infrastructure, YOU ARE THINKING LIKE AN ECONOMIST.

Sail price bargains

The fact that the organisers of the Young Endeavour program decided to use a ballot to determine who would get tickets suggests that the price they set for each leg is lower than the market clearing price. In other words, at the prices they set they expect quantity demanded to exceed quantity supplied. The price is thus akin to a ceiling price. This being the case, why would the organisers not set prices at a higher level? There could be many reasons but the most important is probably that the Young Endeavour Youth Scheme is a not-for-profit organisation that has objectives other than maximising profit. It may, for example, take the ethical view that market clearing prices would exclude all but the very rich from participation. On the other hand, if the organisation were a for-profit company seeking to maximise its profitability it is likely that higher, market clearing prices would be charged. If you said that the prices charged are akin to a price ceiling YOU ARE THINKING LIKE AN ECONOMIST.

As to whether ethical considerations lead to an equal number of places being awarded to men and women, this question can be examined from a number of perspectives. On the one hand, equality of opportunity might dictate that an equal number of tickets be given to both sexes. On the other hand, if demand from one sex is greater than from the other it may seem inequitable to award the same number of places to each. (Imagine if a movie theatre were to allocate equal numbers of seats to men and women to see *Rocky* or to see *Love Actually*. Would men and women need to be charged different prices to fill the theatre?)

Multiple-choice questions

1 If firms in an industry agree among themselves to raise prices, this will result in:
- **a** more output being sold.
- **b** a decrease in demand.
- **c** higher profit for each firm.
- **d** a rise in supply.

2 Which of the following statements is true of a market?
- **a** An increase in supply, with no change in demand, will decrease the equilibrium price and the equilibrium quantity.
- **b** A decrease in supply, with no change in demand, will decrease the equilibrium price and increase the equilibrium quantity.
- **c** An increase in demand, with no change in supply, will decrease the equilibrium price and quantity.
- **d** None of the above are true.

3 Minimum wages are an example of:
- **a** a price ceiling.
- **b** an equilibrium price.
- **c** a price floor.
- **d** a public good.

4 For a price floor to be effective, it must be:
- **a** set by market forces.
- **b** set below market equilibrium.
- **c** set at market equilibrium.
- **d** set above market equilibrium.

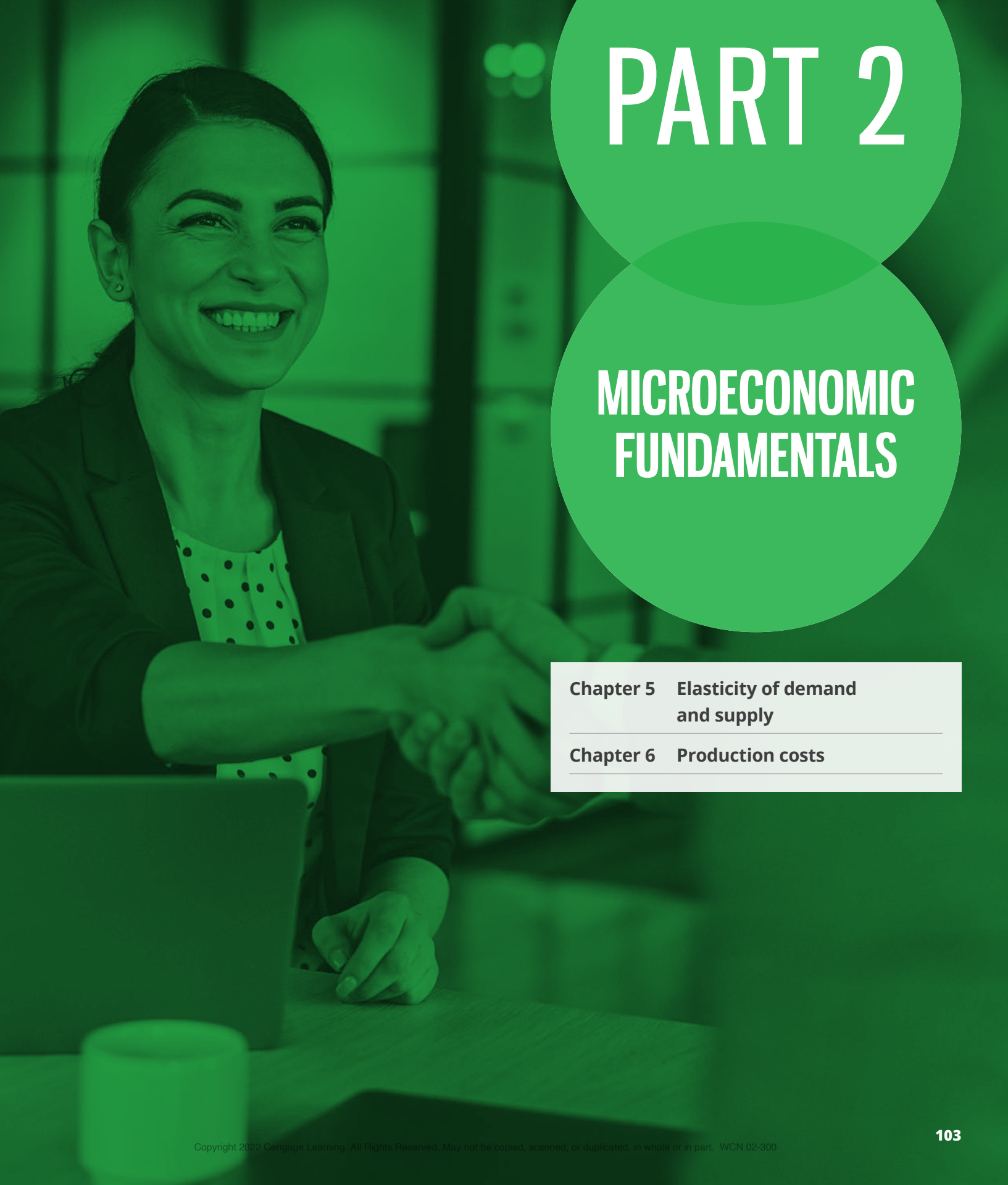
5 Blackouts and brownouts during peak periods of electricity use suggest that the electricity price at these times is:
- **a** just right.
- **b** too low.
- **c** too high.
- **d** irrelevant because building more solar and wind generation capacity is the way to solve the problem.

6 Consider the market for detached houses. An increase in the price of apartments will:
- **a** increase the demand for houses, creating a higher price and a greater number of houses purchased in the market.
- **b** increase the supply of houses, creating a higher price and a smaller number of houses purchased in the market.
- **c** decrease the demand for houses, creating a lower price and a smaller number of houses purchased in the market.
- **d** increase the supply of houses, creating a lower price and a greater number of houses purchased in the market.

7 Suppose the equilibrium price set by supply and demand is lower than the price ceiling set by the government. The result will be:
- **a** a shortage.
- **b** that quantity demanded is equal to quantity supplied.
- **c** an excess supply.
- **d** a black market.

8 Which of the following is/are the likely result from rent controls set below the equilibrium price for rental housing?
- **a** Non-price discrimination and black markets will result.
- **b** The existing rental housing will deteriorate.
- **c** People will demand more houses than are available.
- **d** All of the above.

9 Which of the following have strong public good elements: (i) streetlights, (ii) light houses, (iii) national defence?
- **a** (i) and (ii)
- **b** (i) and (iii)
- **c** (ii) and (iii)
- **d** (i), (ii) and (iii)



PART 2

MICROECONOMIC FUNDAMENTALS

5 Elasticity of demand and supply

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Explore the concept of price elasticity of demand and understand its importance in economics.
2 Understand how price elasticity of demand varies along a straight-line demand curve.
3 Investigate the determinants of price elasticity of demand.
4 Look at other measures of demand elasticity.
5 Explore the concept of price elasticity of supply.
6 Understand the relationship between price elasticity of demand and the incidence of taxation.

Getty Images/EyeEm Premium/John Dunn

Gyms were a big profitable type of business in Australia, but they were hard hit by the coronavirus pandemic due to serious restrictions on public gatherings. According to a survey reported in ABC News (28 May 2020)[1] many gym members looked forward to returning to gym exercise once restrictions were lifted. However, about 15 per cent of them were hesitant, citing potential health risk as their major concern. During the pandemic it was also observed that the sales of home gym equipment jumped. Many gym operators were concerned that people's exercise habit might have changed, preferring now to work out in the comfort of their home.

The 24-hour gym you manage sustained huge losses but managed to reopen after the pandemic. You are considering lowering the membership fees to attract more members to join your gym but wonder how customers will respond. You really want to know how many more memberships will be purchased if the price is decreased because the weekly membership proceeds may fall if there is only a small increase in the

[1] E. Manfield, 'Coronavirus restrictions leave gyms facing tough recovery as fitness industry looks for silver lining', *ABC News*, 28 May 2020, at https://www.abc.net.au/news/2020-05-28/gyms-hit-hard-by-coronavirus-crisis-as-fitness-industry-adapts/12293978

number of new memberships sold. Will it be better to keep the price unchanged or to charge a lower price and sell more?

This chapter teaches you to calculate price elasticity of demand, which is a measure of the responsiveness of quantity demanded of a good or service to changes in its price. Then we will see how this concept of elasticity relates to total revenue. This knowledge of the sensitivity of demand is vital for pricing and public policy decision-making. Next, you will see how quantity demanded responds to changes in consumer income, and how it responds to changes in the prices of related goods. You will also learn how supply responds to price changes. The chapter concludes by applying demand elasticity to examine the impact of taxation on a market.

1 Price elasticity of demand

The law of demand states that there is an inverse relationship between the price and the quantity demanded of a good or service. But it does not tell us by *how much* quantity demanded will change in response to its price change. The price elasticity of demand measures the responsiveness of consumption of a good to its price change. But before we present the formula to measure price elasticity of demand, let us look at one simple example involving two goods: A and B, to explore one important idea. Suppose the prices of both goods decrease by $1 per unit. As a result, the quantity demanded for good A increases by 10 units per month and for good B by 200 units per month. You are asked to decide the consumption of which good is more price sensitive. It is tempting to choose good B, but this conclusion may be wrong because we are missing one piece of important information – the original quantity demanded and price of each good. Consider quantity demanded only for illustration. Suppose before the price fall, 100 units and 10 000 units of good A and good B respectively are sold. It is easy to see that a 10 unit fall in consumption of good A represents 10 per cent fall in quantity demanded, whereas a 200 unit decrease in consumption of good B represents only a 2 per cent fall in quantity demanded. You can now see that it is good A that is more price sensitive, if the original prices of both goods are the same. This example shows that to measure the responsiveness of consumption to price changes we must consider the percentage changes in the price and in the quantity demanded for a good. The **price elasticity of demand** (E_d) is defined as the ratio of percentage change in quantity demanded to percentage change in its price. The formula is presented below:

Price elasticity of demand The ratio of the percentage change in the quantity demanded of a product to a percentage change in its price.

$$E_d = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in price}}$$

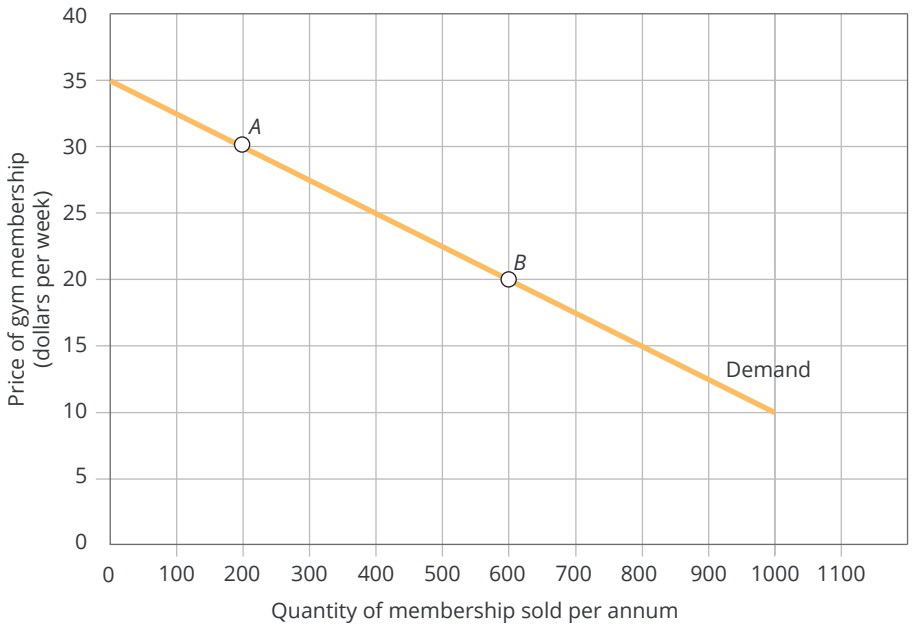
where *Ed* is the price elasticity of demand.

Let's now look at the 24-hour gym that you are managing. You are considering whether you should lower the price charged for silver (ordinary) gym membership. Based on recently collected survey data, you obtained the demand curve for silver membership presented in **Exhibit 5.1**. Suppose the gym lowers its current membership price of $30 to $20 per week and the number of memberships sold increases from 200 to 600 annually. This would involve a move from point *A* to point *B* on the demand curve in **Exhibit 5.1**. We can compute the elasticity coefficient for this change as:

$$E_d = \frac{\%\ change\ in\ Q}{\%\ change\ in\ P} = \frac{\frac{600-200}{200}}{\frac{20-30}{30}} = \frac{200\%}{-33.3\%} = -6$$

In this case, a 33.3 per cent drop in price will increase number of memberships sold by 200 per cent; the price elasticity of demand is -6. The number -6 means that the quantity demanded (membership) increases by 6 per cent for each 1 per cent decrease in price. Note price elasticity of demand is *always* minus because of the law of demand – quantity demanded and price are inversely related. Since it is always negative, economists drop the minus sign. From now on we will only consider the numerical value of elasticity coefficient.

Exhibit 5.1 The demand for gym membership



The price elasticity of demand coefficient calculated for a move from point *A* to point *B* is different from that calculated for a move from *B* to *A*. The midpoint formula enables an average coefficient to be calculated for the arc of the demand curve from *A* to *B*.

The price elasticity of demand – midpoint formula

Now consider the elasticity coefficient computed between these same points on the demand curve when the price is increased rather than decreased. You can see in **Exhibit 5.1** that starting at $20 per week, an increase of the membership price to $30 causes the number of memberships sold to fall from 600 to 200 per week. In the case of this move from point *B* to point *A*, you compute a very different elasticity coefficient, as follows:

$$E_d = \frac{\%\ change\ in\ Q}{\%\ change\ in\ P} = \frac{\frac{600-200}{600}}{\frac{30-20}{20}} = \frac{66.7\%}{50\%} = -1.33$$

The elasticity coefficient of 1.33 is smaller than the elasticity coefficient calculated previously for a fall in price. The disparity in the elasticity coefficients between the same two points on a demand curve (1.33 if price is raised, 6 if price is lowered) occurs because different initial points (P_1, Q_1 or P_2, Q_2) are used as the base to calculate percentage changes. Moving from point A to point B, the initial price is $30, an increase of $10 represents a 33.3% fall, whereas if we start at point B, at the initial price of $20 the same change of $10 now represents a 50 per cent increase. Similar analysis shows that the calculation of percentage changes in quantity demanded also depends on the initial point. Which elasticity coefficient should we use? Economists solve this problem of different base points by using the *midpoint* between the two possible initial values as the base. The *midpoint formula* for price elasticity of demand is:

$$E_d = \frac{\%\ change\ in\ Q}{\%\ change\ in\ P} = \frac{\dfrac{(Q_2 - Q_1)}{(Q_1 + Q_2)/2}}{\dfrac{(P_2 - P_1)}{(P_1 + P_2)/2}}$$

Expressed this way, we divide the change in quantity demanded by the *average* quantity demanded – $(Q_2 + Q_1) / 2$, to calculate the percentage change in quantity demanded. Similarly, to obtain the percentage change in price we divide the change in the price by the *average* price – $(P_1 + P_2) / 2$. The midpoint formula is also commonly called the *arc elasticity formula* because it refers to elasticity over an arc of the demand curve.

Using the midpoint formula the price elasticity of demand is 2.5 regardless of whether the gym raises the membership price from $20 to $30 or lowers it from $30 to $20, as shown in the following calculation.

$$Arc\ E_d = \frac{\%\ change\ in\ Q}{\%\ change\ in\ P} = \frac{\dfrac{(600 - 200)}{(600 + 200)/2}}{\dfrac{(30 - 20)}{(30 + 20)/2}} = 2.5$$

In economic analysis as discussed in the next section when we use revenue changes to test the demand elasticity, the midpoint formula is more accurate. Note also that if the changes in price and quantity demanded are very small, the discrepancy in the elasticity coefficients calculated with the two formulas is negligible.

The total revenue test of price elasticity of demand

As reflected in the elasticity formula, the *responsiveness* of the quantity demanded to a change in price determines the value of the elasticity coefficient. There are three possibilities in relation to the value of this coefficient: (1) the numerator (percentage change in quantity demanded) is greater than the denominator (percentage change in price) and *Ed* is greater than 1; (2) the numerator is less than the denominator and *Ed* is less than 1; and (3) the numerator equals the denominator and *Ed* is equal to 1. How does the price elasticity of demand relate to total revenue?

We will soon see price elasticity of demand is related to change in total revenue. **Total revenue** is the total number of dollars a firm receives from sales of a good or service; it is equal to its price multiplied by the quantity demanded. Depending on the value of the elasticity coefficient, a decrease in price may lead to an increase in total revenue, a fall in total revenue or no change in total revenue. **Exhibit 5.2** presents three cases that the gym may confront.

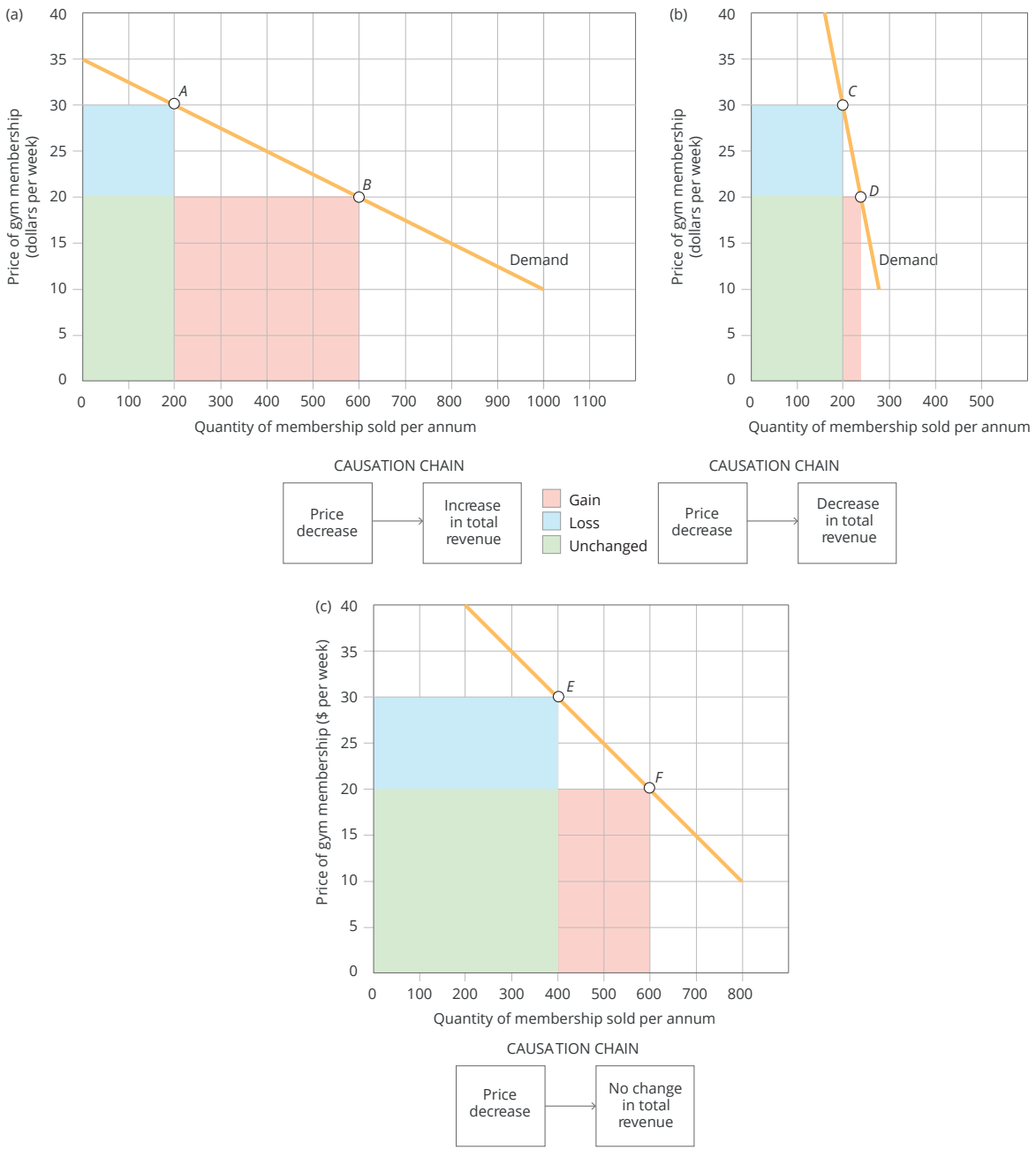
Total revenue
The total number of dollars a firm earns from sales of a good or service; equal to its price multiplied by the quantity demanded.

Elastic demand ($E_d > 1$)

Demand is elastic when the elasticity coefficient is greater than 1. **Elastic demand** is a condition in which the percentage change in quantity demanded is greater than the percentage change in price. Suppose the demand curve for gym membership is as depicted in part (a) of **Exhibit 5.2**. If the gym decreases its membership price from $30 to $20, the quantity demanded increases from 200 to 600. Using the midpoint formula, this means that a 40 per cent change in the membership price brings about a 100 per cent change in quantity demanded. Thus, $E_d = 2.5$. Since *Ed* is greater than 1, the percentage change in quantity is greater than the percentage change in price. Therefore, the drop in price is more than compensated by the rise in quantity and so *total revenue* rises. Part (a) of **Exhibit 5.2** further illustrates this increase in total revenue when the price is decreased from $30 to $20 per week. The total revenue at $30 is $6000 (point A). At $20, the total revenue increases to $12 000 (point B). Comparing the shaded rectangles under the demand curve, we can see that the green area is an amount of total revenue unaffected by the price change and at the lower price of $20 the pink area gained due to increase in memberships sold ($8000 = $20 × 400)

Elastic demand
A condition in which the percentage change in quantity demanded is greater than the percentage change in price.

Exhibit 5.2 The impact of a decrease in price on total revenue



These three different demand curve graphs show three different types of elasticity of demand. In part (a), the demand curve is elastic between points *A* and *B*. The percentage change in quantity demanded is greater than the percentage change in price, $E_d > 1$. As the membership price falls from $30 to $20, total revenue increases from $6000 to $12 000 per week.

Part (b) shows a case in which the demand curve is inelastic between points *C* and *D*. The percentage change in quantity demanded is less than the percentage change in price, $E_d < 1$. As the membership price decreases over the same range, total revenue falls from $6000 to $4800 per week.

Part (c) shows a unitary elasticity between point E and point F. The percentage change in quantity demanded equals the percentage change in price between *E* and *F*, and thus, $E_d = 1$. As the membership price decreases, total revenue remains unchanged at $12 000 per week.

is greater than the blue area lost due to reduction in price ($2000 = $10 × 200). This net gain of $6000 represents the increase in total revenue from $6000 to $12 000 when the gym lowers the price from $30 to $20. The important point to note is that, when total revenue increases (decreases) as a result of a fall (rise) in price, we know demand is elastic.

Inelastic demand (E_d <1)

Inelastic demand is a condition in which the percentage change in quantity demanded is smaller than the percentage change in price. Demand is *inelastic* when the elasticity coefficient is less than 1. Suppose now the demand for gym membership is depicted in part (b) of **Exhibit 5.2**. Here, a fall in price from $30 to $20 causes the quantity demanded to increase by just 40 memberships (from 200 to 240). Using the midpoint formula, a 40 per cent fall in price causes an 18.2 per cent rise in quantity demanded. This means that $E_d = 0.45$. Since E_d is smaller than 1, the fall in price is not fully compensated by the rise in quantity, and this causes total revenue to fall from $6000 to $4800. Note the changes in the shaded rectangles in **Exhibit 5.2**. In general, when total revenue decreases as a result of a fall in price, we can see demand is inelastic.

Inelastic demand
A condition in which the percentage change in quantity demanded is smaller than the percentage change in price.

Unitary elastic demand (E_d = 1)

An interesting case exists when demand is neither elastic nor inelastic. Part (c) of **Exhibit 5.2** shows a demand curve in which over the price range between $30 and $20 the percentage change in quantity demanded is exactly the same as the percentage change in price. That is, the demand is **unitary elastic** ($E_d = 1$) over this price range. Facing this demand curve, if the gym drops the price from $30 to $20, the quantity demanded rises from 400 to 600. Using the midpoint formula, the percentage fall in price is equal to the percentage change in quantity demanded, and its total revenue remains unchanged at $12 000. With unitary elastic demand, total revenue does not change as price changes.

Unitary elastic demand
A condition in which the percentage change in quantity demanded is equal to the percentage change in price.

Perfectly elastic demand (E_d = ∞)

Two extreme cases are shown in **Exhibit 5.3**. Suppose for the sake of argument that the demand curve for gym membership is horizontal, as shown in part (a) of **Exhibit 5.3**. At a price of $20, buyers are willing to buy as many memberships as the gym is willing to offer for sale. At slightly higher prices, buyers buy nothing. This means that when the price is increased by a very small amount (say to $20.01), the quantity demanded will be reduced to zero. If so, $E_d = \infty$ and demand is said to be *perfectly elastic*. **Perfectly elastic demand** is a condition in which a very small percentage change in price brings about a very large percentage change in quantity demanded, and the elasticity coefficient is infinitely large. There may not be any goods or services with perfectly elastic demand. Such a good would need to have a literally perfect substitute. A close example is the demand for a country's currency facing a seller in the forex market. Assuming no arbitrage costs, a slightly higher price charged by a seller will reduce its demand to zero – Why would I buy from this seller when I can buy at a lower price from many others? So, the demand for a country's currency facing this seller is perfectly elastic. You will learn more about the horizontal demand curve in **Chapter 7**.

Perfectly elastic demand
A condition in which a small percentage change in price brings about an infinite percentage change in quantity demanded.

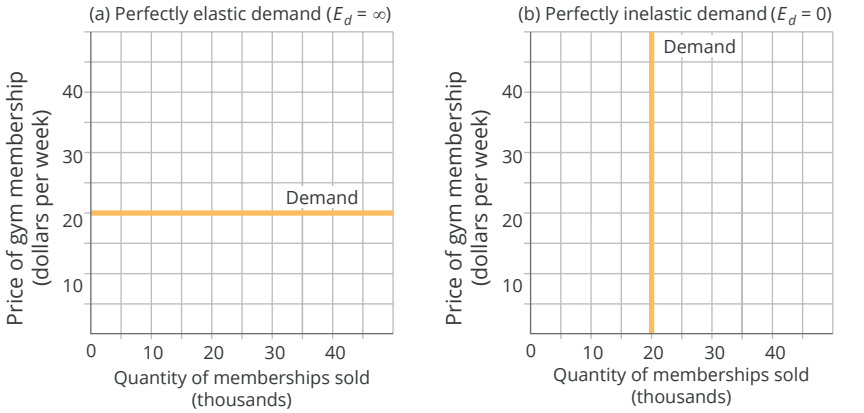
Perfectly inelastic demand (E_d = 0)

Part (b) of **Exhibit 5.3** shows the other extreme case, in which the demand curve for gym memberships is vertical. No matter how high or low the price is, the quantity demanded is unchanged at 20 000 memberships. If so, $E_d = 0$ and the demand is *perfectly inelastic*. **Perfectly inelastic demand** is a condition in which the quantity demanded does not change as the price changes. In real life there may be no product or service that has perfectly inelastic demand. However, the demand for a low-cost, life-saving drug with no available substitute is likely to have close to perfectly inelastic demand.

Exhibit 5.4 summarises the different types of price elasticity of demand. Note that elastic demand curve tends to be flat whereas inelastic demand curve tends to be steep. A special case is presented by the hyperbolic demand curve in part (c) of **Exhibit 5.4** illustrating unitary elasticity. We can show mathematically that the price elasticity of demand is always 1 at any part of this demand curve. In the next section, for a *linear* demand curve we will show that the price elasticity coefficient changes along the demand curve.

Perfectly inelastic demand
A condition in which the quantity demanded does not change as the price changes.

Exhibit 5.3 Perfectly elastic and perfectly inelastic demand



CAUSATION CHAIN

Very small price change → Very large change in quantity demanded

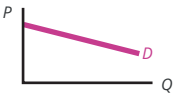
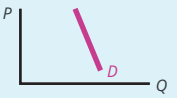
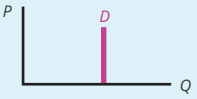
CAUSATION CHAIN

Price change → Zero change in quantity demanded

Here, two extreme demand curves are represented. Part (a) shows a demand curve that is a horizontal line. Such a demand curve is perfectly elastic. At $20 per week, the gym can sell as many memberships as it wishes. At any price slightly above $20, the quantity demanded falls to zero.

Part (b) shows a demand curve that is a vertical line. This demand curve is perfectly inelastic. No matter what the membership price, the quantity demanded remains unchanged at 20 000.

Exhibit 5.4 Price elasticity of demand terminology

Elasticity coefficient	Definition	Demand	
(a) $E_d > 1$	Percentage change in quantity demanded is greater than the percentage change in price	Elastic	
(b) $E_d < 1$	Percentage change in quantity demanded is less than the percentage change in price	Inelastic	
(c) $E_d = 1$	Percentage change in quantity demanded is equal to the percentage change in price	Unitary elastic	
(d) $E_d = \infty$	Percentage change in quantity demanded is infinite in relation to the percentage change in price	Perfectly elastic	
(e) $E_d = 0$	Quantity demanded does not change as the price changes	Perfectly inelastic	

In summary

- **Price elasticity of demand** is a measure of the responsiveness of the quantity demanded to a change in price.
- **Elastic demand** is a change of more than 1 per cent in quantity demanded in response to a 1 per cent change in price.
- **Inelastic demand** is a change of less than 1 per cent in quantity demanded in response to a 1 per cent change in price.
- **Unitary elastic demand** is a 1 per cent change in quantity demanded in response to a 1 per cent change in price.
- **Perfectly elastic demand** is a decline in quantity demanded to zero for even the slightest rise in price.
- **Perfectly inelastic demand** is no change in quantity demanded in response to price changes.

2 Price elasticity of demand variations along a straight-line demand curve

In **Exhibit 5.4**, we illustrate elastic demand with a flat demand curve, and inelastic demand with a steep demand curve. While a flat (steep) demand curve tends to be elastic (inelastic), this is not *always* true. The price elasticity of demand for a straight-line demand curve varies as we move along the curve. Look at **Exhibit 5.5**, which shows a linear demand curve in part (a) and the corresponding total revenue curve in part (b). Begin at $40 on the demand curve and move down to $35, to $30, to $25 and so on. The table in **Exhibit 5.5** shows variations in the total revenue and the elasticity coefficient (E_d) at different ticket prices. In upper part of the demand curve, the price elasticity of demand is greater than 1 and total revenue rises if price is lowered. For example, measured over the price range of $35 to $30, the price elasticity of demand is 4.33 and therefore this segment of demand is elastic ($E_d > 1$). When the price is reduced from $35 to $30, total revenue increases from $175 to $300. As we move down along the segment of the demand curve, price elasticity of demand falls. At $20, price elasticity is unitary elastic ($E_d = 1$) and total revenue is maximised at $400. As we move down the lower segment of the demand curve, price elasticity of demand falls below a value of 1 ($E_d < 1$). Over the price range of $15 to $10, for example, the price elasticity of demand is 0.45. When the price is reduced from $15 to $10, total revenue falls from $375 to $300.

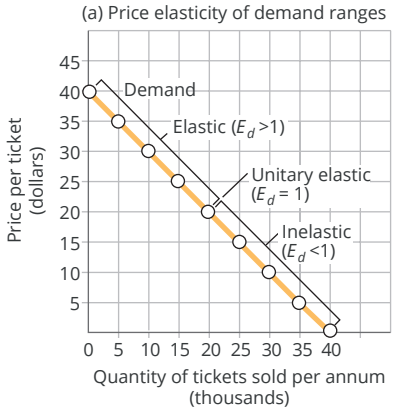
It is no coincidence that the demand curve in part (a) of **Exhibit 5.5** displays elastic, unitary elastic and inelastic demand at different points. In fact, *any straight-line demand curve displays all three of these types of price elasticity of demand*. As we move downward: first, there is an elastic range; second, a unitary elastic point; and third, an inelastic range. Why? Recall that price elasticity of demand is a ratio of percentage changes. At the upper end of the demand curve, quantities demanded are lower and prices are higher. A change of one unit in quantity demanded is a large percentage change. On the other hand, a $1 price change is a relatively small percentage change. Hence, elasticity is greater than 1. At the lower end of the curve, the situation reverses. A one-unit change in quantity demanded is a small percentage change. A $1 price change is a relatively larger percentage change. Hence, elasticity is smaller than 1.

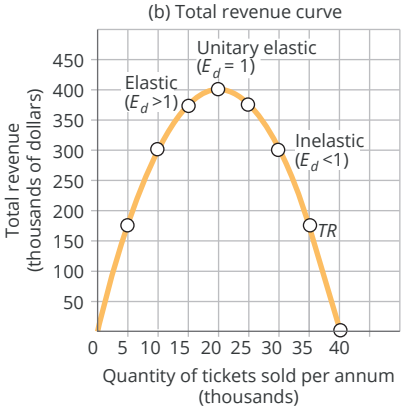
Exhibit 5.6 summarises the relationships among elasticity, price change and total revenue.

Exhibit 5.5 The variation in elasticity and total revenue along a hypothetical demand curve

Calculation of total revenue and elasticity along a hypothetical demand curve

Price	Quantity	Total revenue (thousands of dollars)	Elasticity coefficient (E_d)	Demand
$40	0	$0		
			15.00	Elastic
35	5	175		
			4.33	Elastic
30	10	300		
			2.20	Elastic
25	15	375		
			1.29	Elastic
20	20	400	1.00	Unitary elastic
			0.78	Inelastic
15	25	375		
			0.45	Inelastic
10	30	300		
			0.23	Inelastic
5	35	175		
			0.07	Inelastic
0	40	0		





Part (a) shows a straight-line demand curve and the three elasticity possibilities. In the $40 to $20 price range, demand is elastic. As price decreases in this range, total revenue increases. At $20, demand is unitary elastic and total revenue is at its maximum. In the $20 to $0 price range, demand is inelastic. As price decreases in this range, total revenue decreases. The total revenue (*TR*) curve is plotted in part (b) in order to trace its relationship to price elasticity.

Exhibit 5.6 Relationships among elasticity, price change and total revenue

Price elasticity of demand	Elasticity coefficient	Price	Total revenue
Elastic	$E_d > 1$	↑	↓
Elastic	$E_d > 1$	↓	↑
Unitary elastic	$E_d = 1$	↑ or ↓	No change
Inelastic	$E_d < 1$	↑	↑
Inelastic	$E_d < 1$	↓	↓

Can alcohol abuse be tackled with high alcohol taxes?

You're the economist

Whenever the subjects of alcohol abuse and alcohol-related behavioural problems are raised, there will always be some commentators who propose even higher taxes on alcoholic beverages as a way of discouraging consumption. Alcoholic drinks sold in Australia are already subject to a particular form of consumption tax called excise tax on top of the goods and services tax (GST). (Further explanation of excise taxes appears later in this chapter.) The application of this tax means that alcoholic drinks already bear a heavy tax burden in relation to most other goods. But would it be a good idea to further increase this tax in order to curb alcohol consumption and alcohol-related harm? A journal article by Jon Nelson sheds some light on this question.[1]

In his article, Nelson reviewed 19 papers in which elasticity of demand for alcohol had been calculated. In only two of these papers did their authors find a 'significant and substantial' response to price increases on the part of heavy-drinking adults (alcoholics) over the age of 26. In other words, price elasticity of demand for this group was so low that it was unlikely that increased taxes would have any worthwhile effect on these people's consumption of alcohol. On the other hand, moderate drinkers were found to be much more price sensitive. (It is this group that has been targeted by the alcopops tax discussed in **Chapter 4**.) Furthermore, some research has shown that drinkers who are well informed about the health effects of alcohol consumption have more elastic demand than do uninformed drinkers.

In a more recent UK study Pryce and his colleagues (2019) found that price elasticity of demand for alcohol is dependent also on the level of alcohol consumption. The heaviest drinkers are least price sensitive. This is consistent with Nelson's observation. And compared with lighter drinker, the heavy drinkers also substitute with low quality alcohol in response to price increase.[2]

On the basis of the information above, can you suggest ways in which policy might be framed to reduce alcohol consumption by heavy drinkers? What about moderate drinkers?

Alamy Stock Photo/Gordon Scammell

[1] J. P. Nelson, 'Does heavy drinking by adults respond to higher alcohol prices and taxes? A survey and assessment', *Economic Analysis and Policy*, Vol. 43, No. 3, December 2013, pp. 265–91.

[2] R. Pryce, B. Hollingsworth, and I. Walker. 'Alcohol quantity and quality price elasticities: Quantile regression estimates', *The European Journal of Health Economics*, Vol. 20, No. 3, 2019, pp. 439–454.

In summary

- On a straight-line demand curve, the price elasticity of demand is different at every point along the curve. As we move down along the curve there is first, an **elastic** range; second, a **unitary elastic** point; and third, an **inelastic** range.

3 Determinants of price elasticity of demand

Economists estimate price elasticity of demand for various goods and services. **Exhibit 5.7** presents some of these estimates. You can see that the elasticity coefficients vary a great deal. For example, the demand for cars and for chinaware is elastic. On the other hand, the demand for jewellery and watches and for theatre and opera tickets is inelastic. The demand for tyres is approximately unitary elastic. What makes the price elasticities of demand for these products different? The following factors cause these differences.

Exhibit 5.7 Estimated price elasticities of demand

Item	Elasticity coefficient	
	Short run	Long run
Motor cars	1.87	2.24
Chinaware	1.54	2.55
Heroin	0.80	–
Tyres	0.86	1.19
Commuter rail fares	0.62	1.59
Jewellery and watches	0.41	0.67
Medical care	0.31	0.92
Housing	0.30	1.88
Petrol	0.20	0.70
Theatre and opera tickets	0.18	0.31
Foreign travel	0.14	1.77
Air travel	0.10	2.40

Sources: T. A. Olmstead, S. M. Alessi, B. Kline, R Liccardo Pacula and N M. Petry, 'The price elasticity of demand for heroin: matched longitudinal and experimental evidence', *Journal of Health Economics*, No. 41, May 2015, pp. 59–71, available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4417427/, accessed 22 May 2017. A. Wijeweera and M. Charles, 'An empirical analysis of the determinants of passenger rail demand in Melbourne, Australia', *Economic Analysis and Policy*, Vol. 43, No. 3, December 2013, pp. 249–64; R. Archibald and R. Gillingham, 'An analysis of the short-run consumer demand for gasoline using household survey data', *Review of Economics and Statistics*, Vol. 62, November 1980, pp. 622–8; H. S. Houthakker and L. D. Taylor, *Consumer Demand in the United States: Analyses and Projections*, Harvard University Press, Cambridge, 1970, pp. 56–149; R. Voith, 'The long-run elasticity of demand for commuter rail transportation', *Journal of Urban Economics*, Vol. 30, November 1991, pp. 360–72. Also see R. E. Falvey and N. Gemmell, 'Are services income-elastic? Some new evidence', *Review of Income and Wealth*, Series 42, September 1996, pp. 257–69.

Availability of substitutes

By far the most important determinant of price elasticity of demand is the availability of substitutes. Demand is more elastic for a good or service with close substitutes. If the price of meals at Thai restaurants rises, consumers can switch to Chinese, Malaysian, Indian, Italian or Lebanese restaurants. There are plenty of good substitutes for Thai restaurant meals so the quantity demanded of Thai meals will tend to be very responsive to changes in their price. When consumers have limited alternatives, the demand for a good or service is more price inelastic. Take the example of milk. Because adult consumers find that milk has few substitutes, and because it is an essential component in the diet of young children, it has inelastic demand.

Price elasticity also depends on how broadly or how narrowly we define the good or service for which we wish to calculate elasticity. For example, the price elasticity of demand for Hyundai motor cars is greater than that for cars in general. Hyundai compete with cars sold by Ford, Toyota, Mitsubishi and other car makers. All of these other brands are substitutes for Hyundai cars. If we lump all brands together, however, we eliminate these specific brands of cars as competitors. In short, there are more close substitutes for a specific brand than there are for all cars, which explains why the demand for a Hyundai is more elastic than demand for cars in general.

In Australia, the demand for cars is less elastic than in other countries as there are few close substitutes for cars in Australia where public transport is not as convenient as other developed countries. We can see availability of substitutes is an important determinant of demand elasticity.

Share of budget spent on the product

Now consider a humble good – table salt. When the price of salt changes, consumers pay little attention. Why should they notice? The price of salt could double and this purchase would still remain a small percentage of

one's budget. If, however, international airfares, the prices of washing machines or housing prices double, people will attempt to economise in their expenditure on these goods and services. This is because these goods and services account for a large part of people's budgets.

Adjustment to a price change over time

The timeframe of adjustment also affects elasticities. **Exhibit 5.7** separates the elasticity coefficients into short-run and long-run categories. As time passes, buyers can respond fully to a change in the price of a product by finding more substitutes or changing lifestyles. Consider the demand for petrol. In the short run, people find it hard to cut back the amount they buy when the price rises sharply. They are accustomed to driving back and forth to work alone in their cars. The typical short-run response is to cut unnecessary travel and drive in a fuel-efficient manner. If high prices persist over time, car buyers will find additional ways to cut back. They can move closer to work, buy smaller cars with better fuel economy, form carpools and use public transport. This explains why the short-run elasticity coefficient of petrol in the exhibit is more inelastic at 0.2 than the long-run elasticity coefficient of 0.7.

Applicable concepts: price elasticity of demand and cross-elasticity of demand

Analyse the issue

Can elasticity of demand tell us something about the likely effect of 2020 higher education reforms?

Alamy Stock Photo/imageBROKER

In June 2020, the Education Minister announced a higher education (HE) reform package to better prepare Australia for the post-pandemic economic recovery. The focus of the reform was to channel more resources to train more job ready graduates in fields with more potential growth. The federal government increased its 2021 HE funding by 10 per cent. To incentivise students to choose areas expected to experience high growth, the government's contribution to courses in teaching, nursing, science, engineering and maths was increased. As a result, tuition fees in these courses would fall by 20 to 62 per cent. On the other hand, funding for humanity, management and commerce was reduced; therefore, tuition fees for management and commerce would increase by 28 per cent.[1] How will students respond to these fee changes? Would there be a massive increase in enrolment in nursing degrees where fees would fall by 46 per cent, and a big fall in management and commerce enrolment? The answers depend on the price elasticity of demand for university education.

Elasticity of demand for Australian university courses is low with a major factor contributing to this situation being that students are not required to pay the fee until they join the workforce and their salaries exceed the threshold income level. Researchers have estimated domestic students' price elasticity of demand for Australian university courses to be 0.026.[2] As of 2019, there were just over 1.6 million commencement university students, of which 15.5 per cent are enrolled in nursing courses

1 D. Tahan and M. Cash, 'Job-ready graduates to power economic recovery', 19 June 2020, https://ministers.dese.gov.au/tehan/job-ready-graduates-power-economic-recovery

2 See P. J. Dawkins and J. M. Dixon, 'Alternative approaches to fee flexibility: towards a third way in higher education reform in Australia', *Centre of Policy Studies Working Paper No. G-252*, March 2015, https://www.vu.edu.au/sites/default/files/business/pdfs/higher-education-reform-cops-working-paper- 16-march-2015.pdf, accessed 16 May 2017.

3 Australian Department of Education, Skills and Employment, 'Higher education statistics', https://www.education.gov.au/higher-education-statistics

and 25.1 per cent in management and commerce.[3] The following questions are based on a total enrolment of 1.6 million domestic fee paying students prior to the fee changes as contained in the HE reform package.

What do you think?

Using the information provided, and assuming the price elasticity of demand for university education is the same for all courses, answer the following questions:

1. If the tuition fees for nursing fall by 46 per cent, *ceteris paribus*, how many more students will choose to enrol in nursing courses?
2. Will the nursing departments' revenue from student tuition fees rise or fall as a result of the fee decrease?
3. In answering the above questions, the *ceteris paribus* assumption is made. One such assumption is the fees of other courses are unchanged. What if this *ceteris paribus* assumption is dropped? Would the fee-induced increase in quantity demanded likely to be more or less pronounced for nursing courses?

In summary

- *Determinants of price elasticity of demand* include: (a) the availability of substitutes; (b) the proportion of the budget spent on the product; and (c) the length of time allowed for adjustment.

4 Other measures of demand elasticity

So far, in our discussion of price elasticity of demand we examine the sensitivity of consumption of a good to changes in its own price. We know that consumption is influenced by other factors, such as income and prices of other goods. In the following sections we look at elasticity of demand in relation to changes in income and in relation to changes in the prices of related goods.

Income elasticity of demand

Income elasticity of demand The ratio of the percentage change in the quantity demanded of a good or service to a given percentage change in income.

Recall from **Chapter 3** that an increase in income can increase demand for a normal good and decrease demand for an inferior good, *ceteris paribus*. To measure exactly how consumption responds to changes in income, economists calculate the **income elasticity of demand**. Income elasticity of demand is the ratio of the percentage change in the quantity demanded of a good or service to the percentage change in income that brought about this change in quantity demanded. The formula is presented below:

$$E_Y = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in income}}$$

where E_Y is the income elasticity of demand.

As discussed earlier, since the price elasticity of demand is always negative, we ignore the negative sign of the elasticity coefficient. However, in calculating income elasticity of demand, the sign is important. For a *normal* good, demand and income move in the same direction. Thus, the income elasticity of demand is *positive*, $E_Y > 0$. For an *inferior* good the reverse is true, and the income elasticity of demand is *negative*, $E_Y < 0$.

Why is the income elasticity coefficient important? Returning to our gym membership example, the manager of the gym is interested to know the impact of a pandemic induced recession on membership sales. A Deloitte

Access Economics survey[2] showed that fitness and sport activities are positively linked to household income, as expected. Gym exercise is a normal good. During a downturn when consumers' incomes fall, demand for gym memberships falls. But by how much? The answer is pivotal to the gym manager in deciding on short- and medium-term business strategies. Should they cut back their staff? And by how much? The above-mentioned survey did not investigate the issue of income elasticity for gym memberships, so the below discussion is based on hypothetical figures. Consider the response of gym members who are in the lower income group. The gym manager reasons that gym membership for this income group is quite income sensitive. To illustrate, suppose the downturn caused an estimated 10 per cent fall in family incomes resulting in a 20 per cent decrease in the quantity of memberships sold. The income elasticity of demand for gym membership is 2. Since E_y is *greater than* 1, the demand for gym memberships is income elastic. Why does the gym manager believe that the demand for membership for this lower income group is sensitive to changes in income? It can be argued that to many lower income households, organised fitness and gym exercise is a *luxury* good, when their income falls, they will first cut back on expenditure on luxury items, rather than necessities. However, if you ask some fitness fanatics, they may disagree. To them gym exercise is a *necessity*, and its demand is not sensitive to income changes. Thus to this group of consumers, demand for gym membership is income inelastic, and E_y is positive but *less than* 1.

Let us now look at inferior goods. Potatoes in western countries and rice in Asia are examples of inferior good with negative income elasticity. There is ample empirical evidence that in Asia when a country experienced economic growth and households earned more income, their consumption of rice fell. For example, since the Second World War Japan enjoyed strong income growth and their per capita consumption of rice fell. Researchers found that consumption of rice in Japan had an income elasticity of −0.091 in 1964 and −0.708 in 1984.[3] The negative income elasticity shows that rice in Japan was an inferior good. Note also that the numerical value of income elasticity of rice during this period increased. What this means is that when Japanese household income increased by 1 per cent, their consumption of rice fell by 0.091 per cent in 1964, and this fall in rice consumption grew to 0.708 per cent in the 1980s. In other words, demand for rice became more income sensitive as income rises.

As you might expect, income elasticity for a particular good or service is likely to be different for different groups depending on the level of income they earn as illustrated in the above gym membership and rice examples. As another example, consider a Nigerian study of income elasticity of demand for beverages (both alcoholic and non-alcoholic). It showed that income elasticity for the lowest income quintile in that country was +0.92 whereas for the highest income quintile it was +2.1.[4] For both the highest and the lowest income groups, beverages are a normal good but the degree of responsiveness is much greater for high income earners. Can you think of some reasons why this might be the case?

So what is the practical relevance of income elasticity of demand? A good example of this practical relevance can be found in an examination of the differential effect of a recession on the profitability of business. When there is a downturn in the economy, it is often said that the only businesses that escape its effects are those of the receivers and liquidators who help to wind up companies that have failed. Nothing could be further from the truth. Although sellers of luxury cars and expensive jewellery may feel the pinch, many businesses selling inferior goods, such as fast foods or generic brand groceries, which experience increased demand as income falls, tend to prosper during a recession. In a 2017 article, journalist Catie Low quoted a Merrill Lynch report in which it was argued that because tough times encourage eating at home, major supermarkets benefit from difficult economic conditions.[5] During the pandemic induced recession in 2020, it was observed that supermarket sales in Australia continued to rise and the share prices of major grocery chains bucked the trend of falling share prices.

---

2 See the 2020 Australian Fitness Industry Workforce Report, https://fitness.org.au/articles/fitness-industry-workforce-report-2010-2020/14

3 K. Ogundari, 'A note on socio-economic characteristics and the demand for beverages in Nigeria: Does income matter?', *Economic Analysis and Policy*, Vol. 43, No. 3, December 2013, pp. 293–301.

4 W. S. Chern, K. Ishibashi, K. Taniguchi and Y Tokoyama, 'Analysis of food consumption behavior by Japanese households', April 2002, http://www.fao.org/3/ae025t/ae025t.pdf.

5 Catie Low, 'Supermarket sector to prosper from softening economy', *Sydney Morning Herald*, 31 March 2017, p. 21.

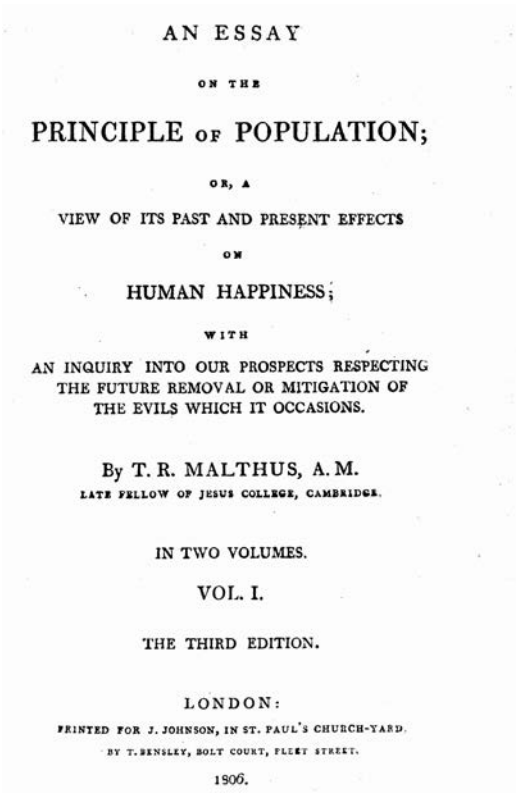
Lastly, the previous example of different income elasticities of demand for beverages in Nigeria shows that a knowledge of income elasticities would be very useful to a beverage manufacturer who is wondering how it will be affected by a changed tax regime that alters the disposable incomes of some groups in that country but not others.

Economics and ethics — Are children an inferior good?

The issue of population growth and its effect on social welfare has exercised the minds of commentators for centuries. In recent times there has been strong emphasis on the contribution of population growth to environmental degradation. Nowhere have concerns about population growth been emphasised more than in China, where sanctions on having children, including forced use of contraceptive devices and sterilisation, were introduced in 1980 as part of the notorious one-child policy. From the perspective of many in the west this was an authoritarian, illiberal policy that would be described as unethical if it were introduced in democratic nations. From the point of view of western commentators – especially economists – there was no need for such an authoritarian approach. Their view was that economic growth itself would solve the problem.

Although Thomas Robert Malthus, a Church of England minister and economist, had argued in his 1798 *Essay on the Principle of Population* that population would continually outstrip the food supply unless checks to population growth occurred, his prediction was wrong. In Malthus's day it was common for families to have 10 or more children. Today, the average family in developed countries has fewer than two children. Although many factors have contributed to this decline, including the improved effectiveness of contraception, some economists believe that one of the most important factors is the rise in family income that has occurred over the past 200 years. The association between rising family income and falling family size may suggest that children are an inferior good. Some reflection reveals this is *not* true. The real explanation has to do with the rising cost of raising children as the cost of living rises in these developed and rapidly developing countries. Think about the costs of piano and ballet lessons, coaching schools and private education to the many middle-class families. When the cost of raising children goes up, couples will choose to have fewer children. It's just the law of demand at work.

In the final analysis, critics of China's now-abandoned one-child policy (which was first replaced by a two-child policy, and recently a three-child policy) perhaps are right in arguing that income growth and the associated rising cost of raising children alone in China could have been sufficient to bring about the desired decline in the rate of population growth.[1] Thus China's strong economic growth in recent decades might have been relied on to solve the problem rather than government choosing the unethical, authoritarian solution of using sanctions to reduce fertility. The recent shift in birth policy in China lends strong support to this theory. In May 2021 China scrapped its two-child policy and allowed couples to have up to three children to cope with the problem of aging population due to its continuous falling birth rates.


AN ESSAY

ON THE

PRINCIPLE OF POPULATION;

OR, A

VIEW OF ITS PAST AND PRESENT EFFECTS

ON

HUMAN HAPPINESS;

WITH

AN INQUIRY INTO OUR PROSPECTS RESPECTING THE FUTURE REMOVAL OR MITIGATION OF THE EVILS WHICH IT OCCASIONS.

By T. R. MALTHUS, A. M.

LATE FELLOW OF JESUS COLLEGE, CAMBRIDGE.

IN TWO VOLUMES.

VOL. I.

THE THIRD EDITION.

LONDON:

PRINTED FOR J. JOHNSON, IN ST. PAUL'S CHURCH-YARD.

BY T. BENSLEY, BOLT COURT, FLEET STREET.

1806.

Alamy Stock Photo/World History Archive

[1] J. Zhang, 'The evolution of China's one-child policy and Its effects on family outcomes', *Journal of Economic Perspectives,* Vol. 3, No. 1, February 2017, pp. 141–60.

Cross-elasticity of demand

In **Chapter 3** you learned that goods are related – a change in the price of one good (say, Y) can cause the consumption of another good (say, X) to change (see 'prices of related goods' in **Exhibit 3.6** in **Chapter 3**). We observed that a rise in the price of petrol caused the number of high-fuel-consumption motor vehicles purchased to decline. This responsiveness of the quantity demanded to change in the price of some other good is estimated by the **cross-elasticity of demand**. Cross-elasticity of demand (E_c) is the ratio of the percentage change in the quantity demanded of a good to a given percentage change in the price of a related good.

Cross-elasticity of demand The ratio of the percentage change in the quantity demanded of a good or service to a given percentage change in the price of a related good or service.

$$E_c = \frac{\text{percentage change in quantity demanded of one good}}{\text{percentage change in price of another good}}$$

Like income elasticity, the sign of cross elasticity of demand is important as it reveals whether a good is a *substitute* or a *complement*. If two goods are *substitutes*, their E_c is positive because the numerator and denominator variables change in the same direction. Consider Coke and Pepsi. Suppose the price of Coke is increased by 10 per cent; consumers respond by buying less Coke and more Pepsi. If then the consumption of Pepsi increases by 5 per cent; the cross-elasticity of demand for Pepsi is +0.5 (+5 per cent/+10 per cent). Coke and Pepsi are *substitutes.* Note also that the larger the positive coefficient, the greater the substitutability between the two goods.

With complementary goods, the variables in the numerator and denominator change in the opposite direction. Now suppose that there is a 50 per cent *increase* in the price of nail polish and the quantity demanded of nail polish remover *decreases* by 10 per cent. The cross-elasticity of demand for nail polish remover with respect to the price of nail polish is –0.2 (–10 per cent/+50 per cent). Since $E_c < 0$, these two goods are complements. The larger the numerical value of the negative coefficient, the stronger the complementary relationship between the two goods.

Note that when we quote a figure for cross-elasticity we must include the sign – either positive or negative – to ensure that there can be no doubt as to whether we are dealing with substitutes or complements.

Exhibit 5.8 has some examples of own price elasticities and price cross-elasticities for various products. These estimates were part of a study into ways in which the diets of low-income Americans could be improved. They are thus relevant to this low-income group rather than to all Americans. Looking at the own price elasticity column, three of the four goods are demand inelastic while the other one is elastic. In relation to the cross-elasticities, it can be seen that some products are complements for vegetables while others are substitutes. Can you identify which products are the substitutes and which are the complements? Which product has the strongest complementarity with vegetables and which is the best substitute for vegetables? Are you able to explain the common sense of some of the relationships between the different products and vegetables?

Exhibit 5.8 Estimated own price and price cross-elasticities of demand

Product	Own price elasticity	Price cross-elasticity with respect to the price of vegetables
Milk	−0.790	−0.033
Meat	−0.758	−0.095
Egg	−0.471	+0.026
Juice	−1.165	+0.037

Source: B-h. Lin, S. T. Yen, D. Dong & D. N. Smallwood, 'Economic incentives for dietary improvement among food stamp recipients', *Contemporary Economic Policy*, Vol. 28, No. 4, October 2010, pp. 524–36.

It is not only in relation to demand that we can make use of the elasticity concept. As you will see in the next section, this concept is equally applicable to supply.

In summary

- **Income elasticity of demand** is a measure of the responsiveness of the quantity demanded to a change in income.
- **Cross-elasticity of demand** is a measure of the responsiveness of the quantity demanded to a change in the price of a related good or service.

5 Price elasticity of supply

Broadly defined, elasticity is a concept that indicates the response of one variable to changes in some other variable. In this section we see how we can use the concept in relation to supply.

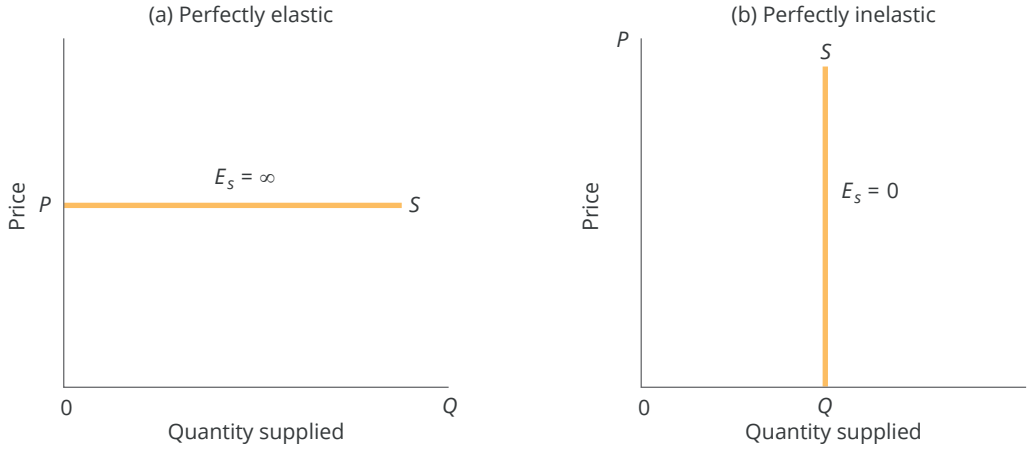
Price elasticity of supply The ratio of the percentage change in the quantity supplied of a product to the percentage change in its price.

The **price elasticity of supply** is the ratio of the percentage change in the quantity supplied of a product to the percentage change in its price. This elasticity coefficient is calculated using the following formula:

$$E_s = \frac{\text{percentage change in quantity supplied}}{\text{percentage change in price}}$$

where E_s is the price elasticity of supply. Since price and quantity supplied change in the same direction, the elasticity coefficient is always positive. Economists use terminology corresponding to that for the elasticity of demand to describe different types of price elasticity of supply. Supply is *elastic* when $E_s > 1$, *unit elastic* when

Exhibit 5.9 Perfectly elastic and perfectly inelastic supply



In part (a), supply is perfectly elastic – a small change in price changes the quantity supplied by an infinite amount: $E_s = \infty$. Part (b) shows that the quantity supplied is unaffected by a change in price: $E_s = 0$, and supply is perfectly inelastic.

$E_s = 1$, *inelastic* when $E_s < 1$, *perfectly elastic* when $E_s = \infty$ and *perfectly inelastic* when $E_s = 0$. **Exhibit 5.9** shows the last two cases.

Although there may be no goods or services in perfectly inelastic supply, there are some goods for which supply is close to being perfectly inelastic. For example, land is sometimes quoted as being in perfectly inelastic supply because no more of it can be made. Nonetheless, there is a slight degree of elasticity since land may be, and sometimes is, reclaimed from the ocean or wetlands (think of China's creation and/or reclamation of islands in the South China Sea, which led to much dispute over the ownerships of those territories around 2020). Sometimes, paintings by a famous deceased artist are said to be in perfectly inelastic supply when the sum total of this artist's work is known and all of these works have been found. Nevertheless, beware the forger!

Perfectly elastic supply implies that a very large quantity of the product can be produced at the same unit cost of production. With modern technology meaning that an existing recording of a piece of music, for example, can be reproduced ad infinitum at the same minuscule cost for each additional copy, it could be said there is a perfectly elastic supply of this recording. The same might be said of digitised books that can be reproduced electronically in increasing quantities at no additional cost per copy. We will look at costs of production and the seller's problem at length in the next few chapters.

From these examples you might have guessed that, in general, supply will be inelastic when costs rise greatly as output is increased. So, for example, the supply of performances by a famous opera singer may be highly inelastic because it is very costly, if not impossible, to coax extra performances from her. The supply curve of her performances tends to be inelastic and the supply curve steep. By contrast, supply will generally be elastic if costs rise only moderately when output is increased. A good example might be of an agricultural output, such as wheat, the supply of which farmers can, at times, easily increase by substituting it for alternative crops relatively easily. The supply of such crops tends to be elastic and the supply curve flat.

Exhibit 5.10 summarises the concepts of income elasticity and cross-elasticity of demand and price elasticity of supply.

Exhibit 5.10 **Summary of elasticity concepts**

Type	Definition	Elasticity coefficient possibilities	Terminology
Income elasticity of demand	$\frac{\text{percentage change in quantity demanded}}{\text{percentage change in income}}$	$E_Y > 0$ $E_Y < 0$	Normal good Inferior good
Cross-elasticity of demand	$\frac{\text{percentage change in quantity demanded of one good}}{\text{percentage change in price of another good}}$	$E_c < 0$ $E_c > 0$	Complements Substitutes
Price elasticity of supply	$\frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}}$	$E_s > 1$ $E_s < 1$ $E_s = \infty$ $E_s = 0$	Elastic Inelastic Perfectly elastic Perfectly inelastic

In summary

- **Price elasticity of supply** is a measure of the responsiveness of the quantity supplied to a change in price.

6 Price elasticity of demand and the impact of taxation

Taxes imposed on petrol, tobacco products and alcohol are an important source of revenue for governments in most countries. In Australia, these taxes are known as excise taxes. They are typically levied at a much higher rate than applies in the case of broad-based consumption taxes, such as Australia's GST.

But who pays the excise tax? One way to answer this question is to say that if the government places a fuel excise on petrol, the petrol companies pay the tax since they collect the tax when they sell petrol and remit the proceeds to the government. But this is not the whole story. Even when taxes are collected from sellers, buyers do not escape a share of the tax burden. **Tax incidence** is the share of a tax ultimately paid by consumers or by sellers. We will show that the tax incidence depends on the price elasticities of demand and supply. In this section, we will consider price elasticity of demand only. Let's look at two scenarios to understand how demand elasticity determines tax incidence.

Tax incidence
The share of a tax ultimately paid by consumers or by sellers.

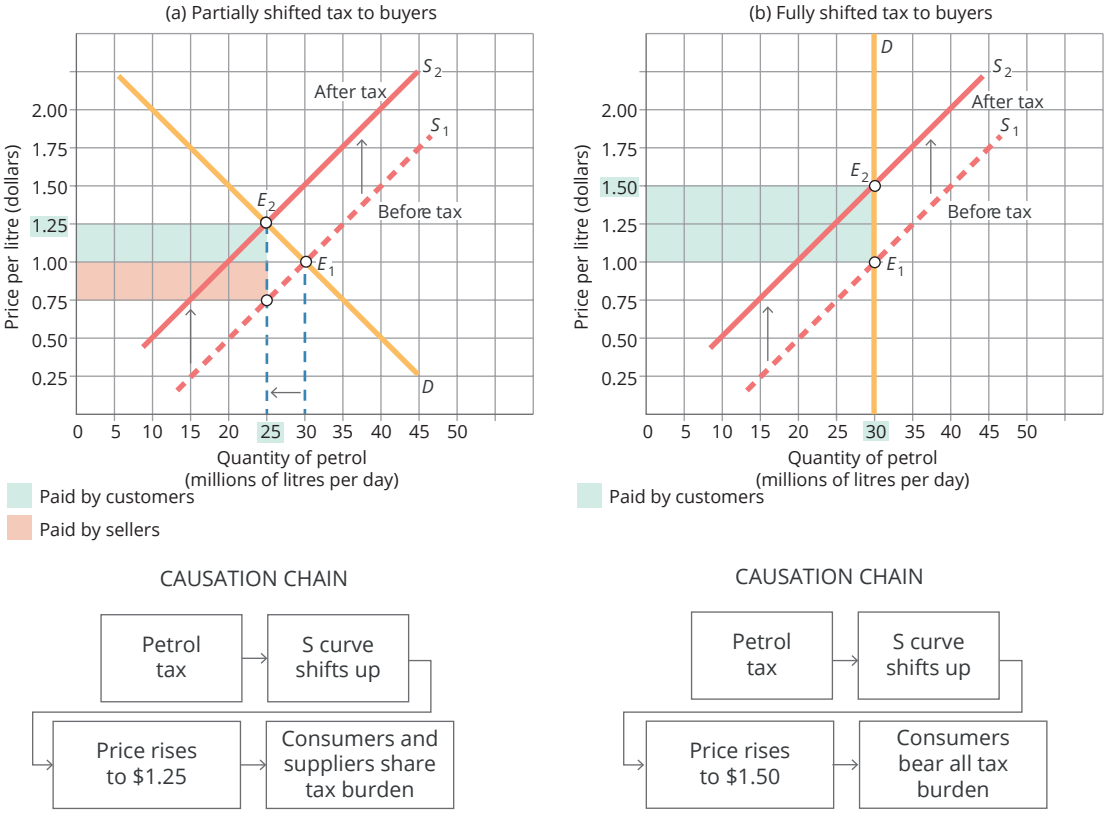
Suppose there is currently no tax on petrol. Now consider a decision by the federal government to impose a petrol excise of $0.50 per litre. **Exhibit 5.11** shows the impact of the tax in each of the two scenario. Consider Scenario (a) where the demand is typically downward sloping to the right. Before the tax, the equilibrium price is $1 per litre and the equilibrium quantity is 30 million litres per day (E_1). The effect of the tax is to shift the supply curve upward by the amount of petrol excise from S_1 to S_2. The vertical distance between the two supply curves (S_1 and S_2) represents the $0.50 tax per litre. From the sellers' viewpoint, the production cost of each litre of petrol increases $0.50. The effect is exactly the same as if the price of crude oil or any resource used to produce petrol increased.

Sellers would like consumers to bear the entire amount of the tax. This would occur if consumers were willing to pay $1.50 per litre for the same 30 million litres per day they purchased before the tax. But at the higher price of $1.50, the quantity demanded is reduced to 20 million litres, which is smaller than the quantity supplied of 30 million litres per day at this price, resulting in excess demand. Therefore, $1.50 cannot be the market equilibrium price. The shift upward of the supply curve following the fuel tax establishes a new market equilibrium at E_2. The new equilibrium price is $1.25 per litre and the equilibrium quantity of 25 million litres per day will be bought and sold. At E_2 the entire shaded area represents the tax revenue. The government collects $12.5 million per day, which equals the $0.50 per litre tax times the 25 million litres sold each day. Since consumers now pay $1.25, which is $0.25 more than before tax, they bear $0.25 or one-half of the tax. Now the sellers send $0.50 to the tax office and keep $0.75 compared to the $1 per litre they kept before the tax. Therefore, the sellers bear $0.25, the remaining half of the tax.

In this numerical example, the fuel tax is equally borne by the buyers and sellers. In general this is not the case. *Ceteris paribus*, the more inelastic the demand, the greater will be the share of the tax burden borne by consumers. This relationship is further illustrated by an extreme scenario in part (b). Part (b) of **Exhibit 5.11** is a special case in which demand for fuel is perfectly inelastic and the demand curve is vertical. In this case, the price rise due to the decrease in supply caused by the tax does not result in buyers responding by decreasing the quantity demanded. The quantity demanded is 30 million litres per day before and after the tax. The equilibrium price increases by exactly the amount of tax per unit from $1 to $1.50 per litre. After paying the tax, sellers receive a net price of $1 per litre, the same as before tax. Therefore, the burden of the tax is shifted entirely onto the buyers. The total tax revenue collected by the government is the shaded area. Each day $15 million is collected, which equals the $0.50 per litre tax multiplied by 30 million litres sold each day.

At the beginning of this section, we mentioned that to raise revenue the government imposes taxes on such goods as petrol, tobacco products and alcohol. You may wonder why these goods are chosen over other goods. The reason is that the demand for these consumer goods tend to be inelastic. When a tax is levied on an inelastic consumer good, its consumption will fall slightly, maximising government's tax revenue.

Exhibit 5.11 The incidence of a tax on petrol



In parts (a) and (b), S_1 is the supply curve before the imposition of a petrol tax of $0.50 per litre. The demand curve is not affected by this tax collected from the sellers. Before the tax, the price is $1 per litre and 30 million litres are bought and sold. The initial equilibrium is E_1.

In part (a), with the given downward sloping demand curve, the equilibrium price rises to $1.25 per litre at E_2 as a result of the tax. After the tax is paid, sellers are paid only $0.75 per litre instead of $1, as they were before the tax. Thus, buyers bear $0.25 of the tax burden and sellers bear the remaining $0.25. In this case the tax burden is equally split.

In part (b), in the unlikely situation of a vertical demand curve, the equilibrium point shifts to E_2 and the new equilibrium price is increased by the full petrol tax to $1.50 per litre. The consumers bear all tax burden.

As a final remark, you should be aware that although excise taxes collected by government may be put to good use, the way in which they distort markets leads to an inefficient outcome in the sense that they prevent the attainment of the free market equilibrium and the maximisation of social surplus. In a way similar to the effects of a price floor, an excise tax results in lower output and higher prices than would be if the tax had not been imposed. This assumes, of course, that the taxes are not designed to correct market failures like those discussed in **Chapter 4**.

Global perspective — Applicable concept: price elasticity of supply

How will scientific research output respond to a policy-induced increase in its demand?

Today's knowledge-based globalised economies rely heavily on scientific research output to ensure continuing high productivity growth. Seeking to promote growth in their domestic economies, governments around the world subsidise scientific research in many different ways. One way to increase scientific research output would be to implement policies designed to increase demand for scientific research; for example, by paying a subsidy to purchasers. Such an increase in demand would be expected to lead to an increase in quantity supplied of scientific research and its market price. But how successful would such a policy be? Would the price rise caused by an increase in demand be associated with a relatively large or a relatively small increase in scientific research output? In other words, is the elasticity of supply of scientific research high or low?

Some clues to the answer to this question can be found in the estimate, contained in a US study, that a 100 per cent increase in the number of research scientists would give little more than a 50 per cent increase in research output. Further increases in the number of scientists would increase output by an even smaller proportion. The fundamental argument is that the best research scientists are already in the industry – any increase in the number of scientists would involve employment of researchers with lower skills than those already employed.[1] This phenomenon can be described by the law of increasing opportunity cost presented in **Chapter 2**, which states that when output is increased, less efficient inputs are being used resulting in an increase in costs of production.

What do you think?

Using the above clues and assuming that the wage rate paid to all research scientists is the same as more researcher are employed, answer the following questions:

1. Draw a diagram showing the effect of increased demand for research output on the price and quantity of research output supplied.
2. If, as the increased demand for research output pushes the price of research output up, the companies and institutes undertaking this research respond to this price increase by employing more researchers, is the supply of research output likely to be elastic or inelastic?
3. As research output increases *further*, does the supply elasticity of research output become smaller or larger?
4. Is the elasticity of supply of research output likely to increase over a six-month period? Over a 10-year period?

[1] See S. P. Dresch, 'The economics of fundamental research', in J. W. Sommer, ed., *The Academy in Crisis: The Political Economy of Higher Education*, Transaction Publishers for the Independent Institute, San Francisco, New Brunswick, 1995. See also R. B. Freeman, 'The economics of science and technology policy' in *The Science of Science Policy: A Handbook*; K. Husbands Fealing, J. I. Lane, J. H. Marburger III & S. Shipp (eds), pp. 85–103, Stanford University Press, Palo Alto, CA, 2011. Freeman argues that elasticity of supply of scientists is also affected by the degree to which individuals can switch from an existing, or intended, profession to a career in science.

In summary

- **Tax incidence** looks at the share of a tax burden borne by buyers and sellers.
- The more inelastic demand is, the greater the tax burden on buyers. In the case where demand is perfectly inelastic, buyers bear the full burden of the tax.

Key concepts

Price elasticity of demand
Total revenue
Elastic demand
Inelastic demand
Unitary elastic demand
Perfectly elastic demand
Perfectly inelastic demand
Income elasticity of demand
Cross-elasticity of demand
Price elasticity of supply
Tax incidence

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Explore the concept of price elasticity of demand and understand its importance in economics

- **Price elasticity of demand** (E_d) is a measure of the responsiveness of the quantity demanded to a change in price. Specifically, E_d is measured as the ratio of the percentage change in quantity demanded to the percentage change in price. By convention it is always shown without the negative sign.
- A good has **elastic demand** if $E_d > 1$. If $E_d < 1$, the demand is **inelastic**.
- For **unitary elastic demand** $E_d = 1$. For a good with **perfectly elastic demand** its demand curve is horizontal and E_d, is infinity.
- A **perfectly inelastic demand** curve is vertical, and its E_d is zero.

2. Understand how price elasticity of demand varies along a straight-line demand curve

- On a straight-line demand curve, the price elasticity of demand is different at every point along the curve.

3. Investigate the determinants of price elasticity of demand

- *Determinants of price elasticity of demand* include: (a) the availability of substitutes, (b) the proportion of the budget spent on the product and (c) the length of time allowed for adjustment.

4. Look at other measures of demand elasticity

- **Income elasticity of demand** is a measure of the responsiveness of the quantity demanded to a change in income. For a normal good, income elasticity of demand is positive. For an inferior good, income elasticity of demand is negative.
- **Cross-elasticity of demand** is a measure of the responsiveness of the quantity demanded to a change in the price of a related good. When the cross-elasticity of demand is negative, the two products are complements; when it is positive they are substitutes.

5. Explore the concept of price elasticity of supply

- **Price elasticity of supply** is a measure of the responsiveness of the quantity supplied to a change in price. Specifically, price elasticity of supply is the ratio of the percentage change in quantity supplied to the percentage change in price.

6. Understand the relationship between price elasticity of demand and the incidence of taxation

- **Tax incidence** looks at the tax burden on buyers and sellers.
- The imposition of a tax on sellers of a product shifts the supply curve up and results in an increase in market price.
- For a good with elastic demand, the sellers will bear more of the tax, whereas for a good with inelastic demand, the tax burden will fall more heavily on the buyers. In the unlikely case where demand is perfectly inelastic, the price will rise by the full amount of a tax and buyers will bear the full burden of the tax.

Study questions and problems

1 It is observed that household incomes in Australia have continued to rise in recent years, online sales of clothing are continuing to rise while department stores and boutiques are experiencing declining demand. Does this situation suggest that clothes from department stores and boutiques are inferior goods?

2 In order to maximise revenue from an excise tax is the government more likely to choose a tax on salt or on peas?

3 A gym manager lowers the silver membership fee from $30 to $25 per week. As a result, the number of persons purchasing this level of membership rises from 300 to 500. Calculate the price elasticity of demand. Is demand for this level of membership elastic, unitary elastic or inelastic? Does total revenue rise, fall or stay the same when the fee is reduced?

4 A spokesperson from the government said in a media release that the increase in alcohol tax is very effective in raising tax revenue and at the same time reducing alcohol abuse. Do you see any contradiction in this statement?

5 Some commentators have argued that policing of the illegal drug trade, which emphasises detection and prosecution of dealers rather than users, does nothing more than raise the price of illegal drugs. Use a demand and supply diagram to explain why such a policy may have only one effect – the raising of prices of illegal drugs. What does this diagram tell you about the total value of drug sales following on from successful implementation of such a policy to detect and prosecute drug dealers?

6 Which of the following goods and services has the lower price elasticity of demand?
- **a** Cooking salt or Saxa cooking salt?
- **b** Electric scooters or bobby pins?
- **c** An emergency call-out of the plumber to fix a bursting pipe or a callout to install a dripping tap?

7 Is the price elasticity of supply of Sydney Harbour Bridge climbs likely to be high or low?

8 Suppose the income elasticity of demand for smartphones is 2 and the income elasticity of demand for house painting is 0.8. Compare the impact on smartphones and house painting of a recession that reduces consumer income by 5 per cent.

9 Assume that the price cross-elasticity of demand for cream with respect to the price of strawberries is –0.5. What does this tell you about the change in quantity demanded of cream when the price of strawberries rises by 10 per cent? Are cream and strawberries substitutes or complements?

Answer to 'You're the economist'

Can alcohol abuse be tackled with high alcohol taxes?

Tackling alcohol abuse is a very complex issue. Nonetheless, the findings in Nelson's study do strongly suggest that higher taxes on alcohol will have little effect on the behaviour of older problem drinkers (alcoholics). By contrast, the behaviour of moderate drinkers and younger drinkers is sensitive to tax increases and this sensitivity may be more marked among better-informed drinkers. A policy of increasing taxes combined with an education program explaining the ill-effects of excessive alcohol consumption should thus reduce consumption by moderate drinkers – particularly younger ones – and

help prevent some of them becoming heavy drinkers. Because taxes cannot easily be tailored to different consumption groups (in this case heavy drinkers as opposed to moderate drinkers) an increase in tax designed to reduce consumption of alcohol by moderate drinkers will not decrease consumption by heavy drinkers. However, heavy drinkers will, through these increased taxes, make even greater contributions to government coffers. If these contributions to government coffers were to be used for education programs and interventions targeted at heavy drinkers, there may be an eventual benefit for heavy drinkers from increased taxes. Nonetheless, if you concluded that Nelson's findings do call into question a policy approach involving increased taxes on alcohol YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 If an increase in train fares in Bangkok raises total revenue of the operators, this is evidence that demand for train rides in Bangkok is:
- **a** unitary elastic.
- **b** price elastic.
- **c** price inelastic.
- **d** perfectly elastic.

2 A straight-line downward-sloping demand curve will have:
- **a** a higher price elasticity of demand coefficient along the upper half of the demand curve.
- **b** the highest price elasticity coefficient along the middle part of the demand curve.
- **c** a constant price elasticity of demand coefficient throughout the length of the demand curve.
- **d** none of the above are true.

3 Suppose the quantity of fish purchased by the Li family is 22 kilograms per year when the price is $10.50 per kilograms and 18 kilograms per year when the price is $19.50 per kilograms. The price elasticity of demand coefficient for this family, using the mid-point method, is:
- **a** 2.
- **b** 1.
- **c** 0.5.
- **d** 0.33.
- **e** none of the above.

4 The Smith family has a small business providing forensic testing services. A spike in the crime rate has led to a 50 per cent increase in the price of forensic services. In response to this increase in price, the Smiths are willing to increase their supply of services by 25 per cent. The supply over this price range is:
- **a** elastic.
- **b** perfectly elastic.
- **c** unitary elastic.
- **d** inelastic.
- **e** perfectly inelastic.

5 The retailers of Port Underwood Gourmet Pies employ an economist to study the price elasticity of demand for this product. The economist estimates that the price elasticity of demand coefficient for a range of prices close to the selling price is less than 1. Given this estimate, what would be your recommendation if the business owner wishes to increase its revenue?
- **a** Lower the price of pies.
- **b** Raise the price of pies.
- **c** Keep the price unchanged.
- **d** Hire more workers.

6 The price elasticity of demand coefficient for a good will tend to be smaller:
- **a** if there are few or no substitutes available.
- **b** if a small portion of the budget will be spent on it.
- **c** in the short run than in the long run.
- **d** all of the above are true.

7 If the government wanted to raise tax revenue in such a way that the burden of the tax would be borne more by the seller than the buyer, it would impose a tax on a good with:
- **a** elastic demand.
- **b** inelastic demand.
- **c** perfectly inelastic demand.
- **d** none of the above.

8 To determine whether two goods are substitutes or complements, an economist would estimate the:
- **a** price elasticity of supply.
- **b** income elasticity of demand.
- **c** cross-elasticity of demand.
- **d** price elasticity of demand.

9 Over the past half century, the number of bathrooms in a typical new suburban house has more than doubled on average. This could reflect:
- **a** a positive income elasticity of demand for bathrooms.
- **b** a fall in the relative price of bathrooms.
- **c** a change in tastes.
- **d** the fact that bathrooms are a normal good.
- **e** all of the above.

Production costs

6

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Distinguish between economist's and accountant's definitions of costs and profit.
2 Understand how short-run production theory provides the foundation for short-run production costs.
3 Explore how short-run costs change as output changes.
4 Understand the marginal-average rule and the inverse relationship between marginal product and marginal cost.
5 Explore the long-run production costs.

iStock.com/Petko Ninov

Starting a business is not easy, and a very high proportion of business start-ups fail. Nonetheless, every year thousands of people have a go at starting their own business – and some of these people are successful beyond imagination. Suppose that you're one of those people starting a new business. You're very excited about the prospect of owning your own company instead of working for someone else. You are under no illusions; it is going to take hard work and sacrifice.

You are a very competent chef who has been working in a popular inner-city restaurant for some years; you even reached the quarter finals of *The Spatula-ette*. Somehow you've had enough of being someone else's employee and you want to start out on your own. With your skills you think you can create and sell packs of frozen food that provide a ready-to-eat meal that can be heated in a few minutes. So you quit your job and with your own savings and some help from a bank you start Fourmet (a fictional company with the motto 'Fourmet for Gourmet'). You lease commercial space, hire employees and purchase raw materials, and soon your company's frozen meals begin leaving your commercial kitchen. In this new business venture, production cost considerations influence each decision you make. The purpose of this chapter is to study the basic principles of production, how these principles impact on various types of costs and your business decision-making. Whether your company is new and small or an international giant, understanding costs is essential for its success. In this chapter and the next chapter, you will follow Fourmet and apply these important principles of production to understand firms' behaviour.

1 Costs and profit

A basic assumption in economics is that the motivation for business decisions is profit maximisation. Economists realise that managers of firms may sometimes pursue other goals, such as seeking a satisfactory profit rather than maximum profit, or building an empire to boost their ego. However, the profit maximisation goal has been shown to be a powerful way of explaining the behaviour of managers of firms who are responsible for making decisions about the appropriate level of output or price. To understand profit as a driving force for business firms, we must first understand how economists measure costs and contrast it with the way in which accountants measure costs.

After one year of operation, you want to know how well your frozen meal manufacturer Fourmet is doing. So you hire an accounting firm to prepare a financial report. Part (a) of **Exhibit 6.1** shows that Fourmet earned total revenue of $1 100 000 in its first year of operation. Costs for wages, ingredients and materials, interest and other payments totalled $1 030 000. Based on standard accounting procedures, this left a profit of $70 000. You show the financial report to your business partner who has formal economics training. Her conclusion gives you a big surprise – your business has made a loss! You wonder why.

Explicit and implicit costs

To understand why your business partner has come to a different conclusion, we need to understand how economists define costs. In economics, cost is *opportunity cost*. Economists define the total opportunity cost of a business as the sum of explicit costs and implicit costs. **Explicit costs** are payments to suppliers for provision of their resources. In our Fourmet example, explicit costs include the wages paid to labour, the cost of ingredients and other materials, the lease payments for the commercial kitchen, and other payments, such as the electricity bills, and the cost of workers compensation insurance. These resources are owned outside the firm and must be purchased with an actual payment to these 'outsiders'.

Explicit costs
Payments to non-owners of a firm for their resources.

Implicit costs
The opportunity costs of using resources owned by a firm.

Implicit costs are the opportunity costs of using resources owned by the firm. When you started Fourmet, you gave up the opportunity to earn a salary as a chef in someone else's restaurant. When you invested your nest egg in this company, you gave up the income it previously earned; this might have been interest on a bank deposit or the return on a share portfolio. You also use a large coldroom that you own to store finished Fourmet products ready for delivery. Although you made no actual payment to outsiders for using the coldroom, you gave up the opportunity to lease it out to someone else. The implicit costs of own labour, financial capital and equipment (coldroom) are presented in the third column in part (b) of **Exhibit 6.1**.

Economic and accounting profit

In calculating profit, accountants only consider explicit costs of running a business. Therefore, *accounting profit* is total revenue minus total explicit costs.

accounting profit = total revenue − total explicit costs

Because economic decisions must include implicit as well as explicit costs, economists use the concept of **economic profit** instead of accounting profit. Economic profit is total revenue minus all opportunity costs (including both explicit and implicit costs). Expressed as an equation:

Economic profit
Total revenue minus explicit and implicit costs.

economic profit = total revenue − total opportunity costs

or

economic profit = total revenue − (explicit costs + implicit costs)

Exhibit 6.1 **Fourmet's financial report**

(a) Fourmet's profit

Item	
Total revenue	$1 100 000
Less costs:	
Employees' wages and salaries	800 000
Ingredients and materials	105 000
Rent	100 000
Interest paid	0
Other payments	25 000
Total	1 030 000
Equals profit	**70 000**

(b) Fourmet's accounting versus economic profit

Item	Accounting profit	Economic profit
Total revenue	$1 100 000	$1 100 000
Less explicit costs:		
Employees' wages and salaries	800 000	800 000
Ingredients and materials	105 000	105 000
Rent	100 000	100 000
Interest paid	0	0
Other payments	25 000	25 000
Less implicit costs:		
Forgone salary	0	130 000
Forgone rent	0	25 000
Forgone interest	0	5 000
Equals profit	70 000	−90 000

Are they earning too much?

Economics and ethics

When it comes to small business, is it better to use economic profit rather than accounting profit as a guide to whether the customer is being ripped off?

Since time immemorial citizens have been troubled by individuals and organisations that are perceived to earn high profits. Indeed, much has been written about highly-profitable firms, which are often criticised for behaving unethically.

But how do we know if there is an ethical problem with an organisation that seems to be earning particularly high profits? A business earning high accounting profits may, after implicit costs are taken into consideration, earn only low or moderate economic profits.

Doctors at the local medical practice seem to be extremely well off. Is this business more profitable than it should be? If we believe we have sufficient information to make a call on this we should be careful that all relevant information

is considered – including implicit costs. So, for example, when calculating economic profit for the practice, are we including the income forgone when the doctors were studying full-time at university for up to six years? What about the implicit costs that doctors in this practice face when they are required to work in the evening or on weekends? Do they forgo time and income to fulfil continuing education requirements? Do they own the building in which their practice is located?

AAP Image/DAN HIMBRECHTS

Unless we know just what these implicit costs of carrying out their profession are, it is very difficult for us to make an ethical judgement about the profits earned by these doctors in their practice. The same is true of many other businesses. And to be sure, there will be businesses where implicit *benefits* that accrue to owners mean that economic profits could be *greater* than accounting profit. Think of the successful rock band that plays for the rich and famous, gets to travel to gigs all around the world, and earns profits because its members are enjoying their passion for making music. When we use economic profit as our guide, we can see that citizens worried about being swindled by unethical businesses might be best advised to look at rock bands rather than medical practices!

Part (b) of **Exhibit 6.1** illustrates the importance of the difference between *accounting profit* and *economic profit*. From the economic perspective, by including the implicit costs of using own resources, Fourmet made a loss instead of a profit. Let us look at these implicit costs more closely. Your $130 000-a-year salary as a chef was forgone in order to spend all your time running Fourmet. Also forgone were $25 000 in rental income and $5000 in interest that you would have earned during the year by renting your coldroom to someone else and putting your savings in the bank. The total cost of operation is $1 190 000. Subtracting both *explicit* and *implicit* costs from *total revenue*, Fourmet had an *economic loss* of $90 000. The firm is failing to cover the opportunity costs of using its resources in the food industry. If the analysis considers only accounting profit, Fourmet is profitable, because it ignores implicit costs. The distinction between explicit and implicit costs is important for us to fully understand the distinction between economic and normal profits, which we will discuss in **Chapter 7**.

You're the economist — A profitable newsagent?

There were over 5000 Australian newsagents in 2020, all of which sold a range of retail goods, including milk, lollies, birthday cards, bus tickets, Lotto and scratch cards, and more. Mr and Mrs Chen and their three teenage sons migrated to Australia about two years ago. Last year, feeling undervalued, Mr Chen quit his job as a restaurant manager, for which he earned $75 000 a year. With their savings (currently earning 5% of interest per annum) he decided to buy a newsagent in Sunnybank, a suburb in Brisbane close to where he lived, and became his own boss. The costs of buying the already well-established business in this fast-growing suburb include costs of goodwill, stock at wholesale value, government and legal fees, and other charges, totalling $120 000, which he paid with his family savings.


Alamy Stock Photo/Bill Bachman

Despite opening long hours each day, seven days a week, they hire only one full-time shop assistant to run this business. Mr Chen asked their sons to take turn and work at the newsagent after school and on weekends. Now after a year of operation and some number crunching, Mr Chen joyfully told his family at dinner their newly bought business has made a profit of well over $100 000 for the first 12 months. His wife asked how he arrived at this figure. Mr Chen showed her the following Trading, Profit and Loss Statement:

Sales	$765 400
Less Cost of goods sold	
Purchases	580 400
Less Change in stock (per stocktake)	19 600
Trading surplus	**165 400**
Add	
Commissions (Lotto and others)	101 800
Gross Profit	**266 500**
Less Expenses	
Rent and outgoings	84 000
Wages (1 staff)	68 000
Repairs and maintenances	2 300
Electricity and internet	7 500
Net profit	**104 700**

Based only on economic decision-making, do you agree with Mr Chen that he made a profit from running his own new business?

In summary

- **Economic profit** is equal to total revenue minus both **explicit** and **implicit costs**. Implicit costs are the opportunity costs of forgone returns to resources owned by the firm.
- Accounting profit equals total revenue minus explicit costs. Since business decision-making is based on **economic profit** rather than accounting profit, the word *profit* in this text always means economic profit.

2 Short-run production theory

Having presented the important distinction between explicit and implicit costs, the next step is to study production theory, which is the foundation for short-run cost theory.

Short run versus long run

The sales of your frozen gourmet meals have increased substantially during the coronavirus pandemic when people spent more time at home preparing their meals, and your current stock is running low. While you are very pleased with this result from a business point of view, you are faced with an important business decision. Should Fourmet expand its operation by running a bigger kitchen? It may be a wise decision to make if the increase in

sale is not temporary and will continue after the pandemic. However, to increase the scale of operation of your business, it will take at least three months. Before you can expand your kitchen, how are you going to meet the increase in demand for your products now? You can utilise your existing kitchen more intensively by hiring more employees and/or asking your employees to work longer hours.

Long run
A sufficient period of time to allow all inputs to be varied.

If the timeframe allows you to increase the scale of your operation (and if you assess it is profitable to do so), you are making a **long-run** decision. The long run is a period of time of sufficient duration to allow all inputs to be varied. In the long run, the firm can build new factories or purchase new machinery. Note that new firms can also enter the industry in the long run if they find it profitable to do so. In the long run, all factors of production (or inputs) – for example, the size of kitchen, number of fridges and freezers, and number of employees – are variable and there is no fixed input.

Short run
A period of time during which there is at least one fixed input.

However, if you need to increase output for meeting the increased demand within a short timeframe, you are making a **short-run** decision. The short run is a period of time during which there is at least one fixed input. For example, a firm can increase output by hiring more employees (variable input) only, while the size of the firm's plant (fixed input) remains unchanged. Within this short time period, you can only hire more employees and purchase more ingredients to raise output, while the kitchen size and its cooking equipment cannot be varied.

Fixed input
Any resource for which the quantity cannot change during the period of time under consideration.

Variable input
Any resource for which the quantity can change during the period of time under consideration.

In the short run there are two types of inputs – *fixed inputs* and *variable inputs*. A **fixed input** is any resource for which the quantity cannot change during the period of time under consideration. For example, the physical size of a firm's plant and the production capacity of heavy machines cannot change easily within a short period of time. They remain as fixed inputs while managers decide to vary output. A **variable input** is any resource for which the quantity can change during the period of time under consideration. For example, managers can hire fewer or more employees as needed.

You may ask: How long is the duration for us to call the time period the long run? Economists do not partition production decisions on the basis of any specific number of days, months or years. Instead, the distinction depends on the technology and method of production utilised in the industry. For example, the long run of a garment factory can be as short as 3 to 6 months, whereas for a car plant the long run can easily be 2 years or longer.

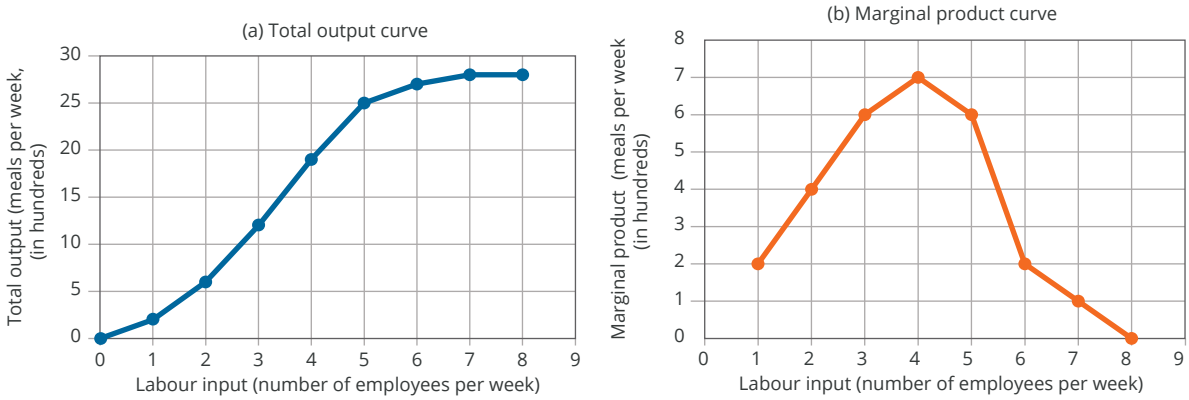
The production function

Production function
The relationship between the maximum amounts of output a firm can produce and various quantities of inputs.

Having defined fixed and variable inputs, we can now describe the transformation of these inputs into outputs, using a concept called a **production function**. A production function is the relationship between the maximum amounts of output a firm can produce and various quantities of inputs. An assumption of the production function model we are about to develop is that the level of technology is fixed. Technological advances would mean that more output is possible from a given quantity of inputs. We will relax this assumption in a later chapter to investigate economic growth.

Part (a) of **Exhibit 6.2** presents a short-run production function for Fourmet. For simplicity, we assume labour is the only variable input, and each employee is presumed to have equal job skills. The size of kitchen and appliances (ovens, freezers etc.) are fixed. Our production model is therefore operating in the short run. If no employees are employed, no meals will be produced. A single employee can produce 200 meals per week. But a lot of time is wasted when the only employee moves from one job to another; for example, purchases ingredients, takes orders, prepares and cooks foods, packs and delivers meals. Adding the second employee raises output to a total of 600 meals per week; output increases by 400 meals. The much bigger increase in output as a result of adding the second employee is *not* because the second employee is super-skilful, but because the employees divide the tasks and specialise. Division of labour makes the team more productive. The *increase in output* continues to increase when more employees are hired as division of labour can be carried to a greater extent. For the third employee, output increases by 600 meals, for the fourth employee by 700 meals.

Exhibit 6.2 **Fourmet's short-run production function and the law of diminishing returns**



Short-run production function of Fourmet

Labour input	Total output (number of meals per week, in hundreds)	Marginal product (number of meals per week, in hundreds)
0	0	–
1	2	2
2	6	4
3	12	6
4	19	7
5	25	6
6	27	2
7	28	1
8	28	0

Part (a) shows how the total output increases as the number of employees increases while the size of kitchen remains constant.

Part (b) illustrates the law of diminishing returns. The first employee adds 200 meals per week; his marginal product is 200 meals. Adding a second employee adds an additional 400 meals per week to total output; the marginal product of the second employee is 400 meals. The marginal product of labour continues to increase until it reaches its maximum of 700 meals per week when the fourth employee is hired. Thereafter, marginal product of labour diminishes, illustrating the law of diminishing returns.

Marginal product

The relationship between changes in total output and changes in labour is called the **marginal product**. Marginal product (MP) is the change in total output produced by adding one more unit of a variable input, with all other inputs used being held constant. When Fourmet increases its labour usage from zero to one employee, output rises from zero to 200 meals per week. The MP of the first employee is 200 meals. The MP is 400 meals for the second employee, 600 meals for the third, and 700 meals for the fourth. The observed increase in MP as discussed is due to division of labour and specialisation. Similar marginal product calculations generate the marginal product curve shown in part (b) of **Exhibit 6.2**. Economists also call rising marginal product due to specialisation the *short-run increasing returns*.

Marginal product
The change in total output produced by adding one unit of a variable input, with all other inputs used being held constant.

The law of diminishing returns

It can be observed that the MP of labour comes to a peak when the fourth employee is hired, after which MP starts to fall. But you might ask: Why won't the marginal product of labour continue to rise beyond the fourth employee? The reason is that, as more employees are added, they must share fixed inputs, such as ovens, with some employees standing around waiting for an oven to become available. As a result, marginal product declines. In the extreme case, marginal product could well be negative. At some point, as the number of employees increases, they start stepping on each other's toes and getting in each other's way when they have to work with a limited amount of bench space, mixing machines and other fixed inputs. A profit-seeking firm would never intentionally hire employees with zero or negative marginal product. In **Exhibit 6.2**, do you know when MP of labour becomes negative?

Law of diminishing returns
The principle that, beyond some point, the marginal product decreases as additional units of a variable factor are added to a fixed factor.

Although we are using a food industry example, this phenomenon of falling marginal product applies to production of all goods and services. Economists call this the **law of diminishing returns**, which states that beyond some point, the marginal product decreases as additional units of a variable factor are added to a fixed factor. For Fourmet, diminishing returns sets in after the fourth employee. Because the law of diminishing returns assumes that there is at least one fixed input, this principle is a short-run, rather than a long-run, concept.

In summary

- A firm operates in the **short run** whereby some inputs are **fixed**. The firm plans in the **long run** when all inputs are variable.
- A **production function** is the relationship between inputs and outputs. The short-run production function shows how total output changes as the amount of one input, such as labour, varies, keeping capital facilities unchanged.
- **Marginal product** is the change in total output caused by a one-unit change in a **variable input**, such as labour.
- The **law of diminishing returns** states that, after some level of output in the short run, each additional unit of the variable input yields smaller and smaller **marginal product**.

3 Short-run cost concepts

As an economic agent, a producer makes production decision by comparing the cost and benefit of output. Whether making production decisions in the short run or the long run, a business must therefore determine the costs associated with producing various levels of output. In this section using the Fourmet example, we will study the relationship between short-run costs and output of Fourmet's gourmet packs. Long-run costs will be discussed in the next section.

Total cost curves

Total fixed cost
Costs that do not vary as output varies and that must be paid even if output is zero. These are payments that the firm must make in the short run, regardless of the level of output.

As production expands in the short run, some inputs are fixed and some are variable. Therefore, the total costs of production are also divided into two basic categories: total fixed cost and total variable cost. We look first at total fixed cost.

Total fixed cost

Total fixed cost consists of costs that do not vary as output varies and that must be paid even if output is zero. These are payments for fixed inputs that the firm must make in the short run, regardless of the level of output. Fixed costs are therefore beyond management's control in the short run. The total fixed cost (*TFC*) for

Fourmet is $100[1], as shown in column two of **Exhibit 6.3**. Even if Fourmet produces nothing, it still must pay rent, interest on loans and general insurance.

Total variable cost

As the firm expands from zero output, **total variable cost** is added to total fixed cost. Total variable cost relates to the costs of variable inputs, which in Fourmet's case would include gas and electricity, ingredients and labour. In our simple example, in order to better understand the principles involved, we assume that the only input is labour. Put simply, the variable cost is the cost of the wages of the employees employed. As a firm uses more input to produce output, its variable costs will increase. Management can control variable costs in the short run by changing the level of output. **Exhibit 6.3** lists the total variable cost (*TVC*) for Fourmet in column three. Recall that, when explaining the production function, we assumed that all employees have the same level of job skills.

Total variable costs
Costs that vary as output varies.

Total cost

Given total fixed cost and total variable cost, the firm can calculate **total cost**. Total cost is the sum of total fixed cost and total variable cost at each level of output. As a formula:

$$TC = TFC + TVC$$

Total cost (*TC*) for Fourmet is shown in column four of **Exhibit 6.3**. Part (a) of **Exhibit 6.4** uses data from **Exhibit 6.3** to construct graphically the total cost, total fixed cost and total variable cost at each level of output.

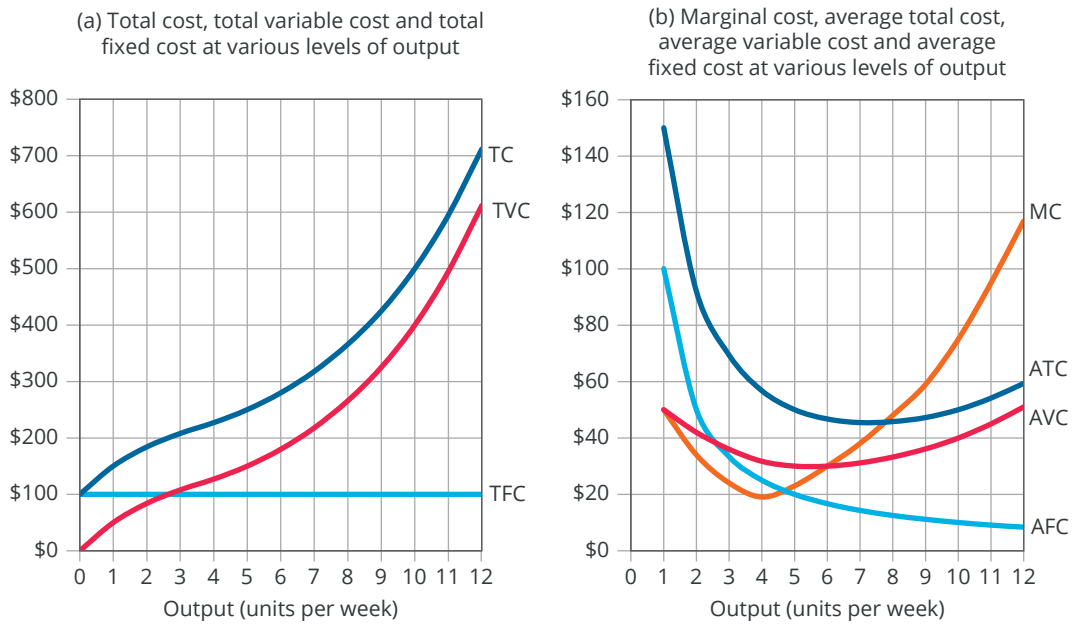
Total cost
The sum of total fixed cost and total variable cost at each level of output.

Exhibit 6.3 **Short-run cost schedule for Fourmet**

Total product (Q)	Total fixed cost (TFC)	Total variable cost (TVC)	Total cost (TC)	Marginal cost (MC)	Average fixed cost (AFC)	Average variable cost (AVC)	Average total cost (ATC)
0	$100	$ 0	$100	$ 0	–	–	–
1	100	50	150	50	$100	$50	$150
2	100	84	184	34	50	42	92
3	100	108	208	24	33	36	69
4	100	127	227	19	25	32	57
5	100	150	250	23	20	30	50
6	100	180	280	30	17	30	47
7	100	218	318	38	14	31	45
8	100	266	366	48	13	33	46
9	100	325	425	59	11	36	47
10	100	400	500	75	10	40	50
11	100	495	595	95	9	45	54
12	100	612	712	117	8	51	59

[1] The figures used in the Fourmet example are simplified for ease of highlighting the principles underpinning short-run cost relationships. Note also that the level of output is measured in some arbitrary unit and therefore one unit should not be taken as one frozen meal.

Exhibit 6.4 **Short-run cost curves**



The curves in this figure are derived by plotting data from Exhibit 6.3. Part (a) shows that the total cost (*TC*) at each level of output is the sum of total variable cost (*TVC*) and total fixed cost (*TFC*). Because the *TFC* curve does not vary with output, the shape of the *TC* curve is determined by the shape of the *TVC* curve. The vertical distance between the *TC* and *TVC* curves is *TFC*.

In part (b), the marginal cost (*MC*) curve at first decreases, then reaches a minimum, and then increases as output increases. The *MC* curve intersects both the average variable cost (*AVC*) curve and the average total cost (*ATC*) curve at the minimum point on each of these cost curves. MC, ATC and AVC curves are typically U-shaped. The average fixed cost (*AFC*) curve declines continuously as output expands. *AFC* is also the difference at any quantity of output between the *ATC* and the *AVC* curves.

Note that the *TVC* curve varies with the level of output and the *TFC* curve does not. The *TC* curve is simply the *TVC* curve plus the *TFC* curve. That is, the vertical distance between the *TC* and *TVC* curves represents *TFC*.

Average cost curves

In addition to total cost, *average cost* is of great interest to firms. Average cost, like product price, is stated on a per-unit basis. The last three columns in **Exhibit 6.3** present *average fixed cost* (*AFC*), *average variable cost* (*AVC*) and *average total cost* (*ATC*) calculated to the nearest whole dollar. These average, or per-unit, curves are shown in part (b) of **Exhibit 6.4**. These three concepts are defined and further discussed as follows.

Average fixed cost

Average fixed cost
Total fixed cost divided by the quantity of output produced.

As output increases, **average fixed cost** (*AFC*) falls continuously. The *AFC* is total fixed cost divided by the quantity of output (*Q*) produced. Written as a formula:

$$AFC = \frac{TFC}{Q}$$

As shown in part (b) of **Exhibit 6.4**, the *AFC* curve approaches the horizontal axis as output expands. This is because larger output numbers divide into *TFC* and cause *AFC* to become smaller and smaller.

Average variable cost

The **average variable cost** (*AVC*) is total variable cost divided by the quantity of output produced. Written as a formula:

Average variable cost
Total variable cost divided by the quantity of output produced.

$$AVC = \frac{TVC}{Q}$$

As shown part (b) of **Exhibit 6.4**, at first the *AVC* curve falls and then, after an output of six units per week, the *AVC* curve rises. Thus, the *AVC* curve is U-shaped. The explanation for the shape of the *AVC* curve is given in the next section.

Average total cost

The **average total cost** (*ATC*) is total cost divided by the quantity of output produced. Written as a formula:

Average total cost
Total cost divided by the quantity of output produced.

$$ATC = AFC + AVC = \frac{TC}{Q}$$

Like the *AVC* curve, the *ATC* curve is U-shaped, as shown in part (b) of **Exhibit 6.4**. At first the *ATC* curve falls because its component parts – *AVC* and *AFC* – are falling. As output continues to rise, the *AVC* curve begins to rise while the *AFC* curve falls continuously. Beyond the output of seven units per week, the rise in the *AVC* curve is greater than the fall in the *AFC* curve, which causes the *ATC* curve to rise in a U-shaped pattern.

Marginal cost

Marginal analysis asks how much it costs to produce an *additional* unit of output. Column five in **Exhibit 6.3** shows **marginal cost**. Marginal cost is the change in total cost when one additional unit of output is produced. Written as a formula:

Marginal cost
The change in total cost when one additional unit of output is produced.

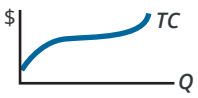
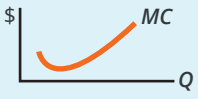
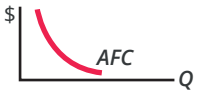
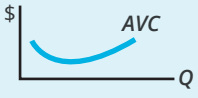
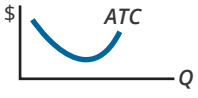
$$MC = \frac{\text{change in } TC}{\text{change in Q}} = \frac{\text{change in } TVC}{\text{change in Q}}$$

Note that marginal cost can also be calculated from changes in *TVC*. This is because the only difference between *TC* and *TVC* is total fixed cost. Since fixed cost does not vary with output, the changes in *TC* and *TVC* with each unit change in output are the same.

In our Fourmet example, since simplified data are used, the calculation of *MC* is straightforward. When output increases from zero to one unit per week, *TC* increases by $50 (from $100 to $150); hence, the *MC* of the first unit of output is $50. For the second unit, *TC* increases by $34 (from $150 to $184); the *MC* of the second unit is $34. Similar calculations produce the rest of the MC column. (You should be able to obtain the same *MC* values by examining the changes in *TVC* for each additional unit of output.) Part (b) of **Exhibit 6.4** shows this marginal cost schedule graphically. In the short run, a firm's marginal cost falls initially as output expands, eventually reaches a minimum (at four units of output for Fourmet), and then rises, forming a U-shaped curve.

Exhibit 6.5 summarises a firm's short-run cost relationships.

Exhibit 6.5 **Short-run cost formulas**

Cost concept	Formula	
Total cost (*TC*)	$TC = TFC + TVC$	
Marginal cost (*MC*)	$MC = \frac{\text{change in } TC}{\text{change in } Q} = \frac{\text{change in } TVC}{\text{change in } Q}$	
Average fixed cost (*AFC*)	$AFC = \frac{TFC}{Q}$	
Average variable cost (*AVC*)	$AVC = \frac{TVC}{Q}$	
Average total cost (*ATC*)	$ATC = \frac{TC}{Q}$	

In summary

- **Fixed costs** consist of costs of inputs that do not vary with the level of output (fixed inputs), such as rent for office space and general insurance **Total fixed cost** is the cost of all fixed inputs in the short run.
- Variable cost consists of costs of inputs that vary with the level of output (variable inputs), such as wages, raw materials. **Total variable cost** is the cost of all variable inputs used in production at each level of output.
- **Total cost** is the sum of total fixed cost and total variable cost at each level of output.
- **Marginal cost** is the change in total cost associated with one additional unit of output.
- **Average fixed cost** is the total fixed cost divided by total output at each level of output.
- **Average variable cost** is the total variable cost divided by total output at each level of output.
- **Average total cost** is the total cost divided by total output at each level of output which is also equal to the sum of average fixed cost and average variable cost.

4 Marginal cost relationships

Part (b) of **Exhibit 6.4** presents two important relationships that require explanation. First, we will explain the rule that links the marginal cost curve to the average cost curves. Second, we will return to the marginal product curve in part (b) of **Exhibit 6.2** and explain its connection to the marginal cost curve.

The marginal-average rule

Marginal-average rule When applied to cost relationships, the rule stating that when marginal cost is below average cost, average cost falls. When marginal cost is above average cost, average cost rises. When marginal cost equals average cost, average cost is at its minimum point.

Observe that the *MC* curve in part (b) of **Exhibit 6.4** intersects with both the *AVC* curve and the *ATC* curve at their minimum points.[2] This is not accidental. It is the result of a relationship called the **marginal-average rule**. The marginal-average rule states that when marginal value is below average value, average value falls; when marginal value is above average value, average value rises; and when marginal value equals average value, average value is at its minimum point. The marginal-average rule applies to grades, weights and many other figures.

Perhaps the best way to understand this rule is to apply it to a non-economic example. Suppose there are 20 students in your classroom and their average weight is (say) 60 kilograms. Now let another student who weighs 120 kilograms join the class. The new average weight of 21 students in the class rises to 62.9 kilograms. The average weight was pulled up because the weight of the additional student (*marginal weight)* was greater than *average*. The opposite is true if the additional student weighs less than 60 kilograms.

Applicable concepts: marginal cost, fixed cost and average fixed cost

Global perspective

Could marginal cost really be zero?

Recall that marginal cost is 'the change in total cost when one additional unit of output is produced'. What do you think the marginal cost of some of the goods and services you buy could be? How much does it cost a manufacturer to make an additional sports car that retails for $100 000? What about a $25 haircut or a $800 smartphone? Your answer is likely to be some fraction of the retail price. So, for example, you might guess that the marginal cost of the sports car is $60 000; perhaps the marginal cost of a haircut would be $4. Even if you were considering a single lolly at the sweet shop, you would likely guess that the marginal cost would amount to a cent or two. But would there really be goods or services that you use that would have a marginal cost of zero?

When did you last access some information or some entertainment on the web without having to pay for it? Have you noticed how many digitised copies of books from around the world are available free online? These are just some of the instances where marginal cost of providing an additional unit of a service to a consumer is zero, or very close to zero. In these examples it costs the 'seller' (or should we say provider) nothing when you click on their web address to access that additional electronic copy of the book, which is then made available to you. And just for the record, this is not a new idea; for example, the advent of public radio broadcasting in the 1920s meant that provision of radio service to an additional listener imposes no additional costs on the broadcaster. Similarly, the advent of the village clock meant that the time of day could be seen or heard by new residents and visitors at zero marginal cost.

However, you might ask, aren't there some fixed costs of providing these zero marginal cost services? The answer to this question is yes, there normally are non-zero fixed costs even when marginal cost is zero. Furthermore, these fixed costs are often reflected in the price you pay for services that have zero marginal cost. For example, when you download newspaper or magazine articles that have paywalls, you will pay a fee even though the cost to the supplier of your additional download is zero. However, it is not always practical or profitable to charge customers a fee for zero marginal cost services. In cases where there is no fee charged for these services, firms must use other ways of earning revenue to offset fixed costs. Advertising has been the main source of revenue for commercial television and radio broadcasters to cover costs. While advertising on internet sites is also a common way of doing this, in some

[2] This observation is only approximately illustrated in part (b) of **Exhibit 6.4** since discrete data are used to construct the various cost curves. We can mathematically prove that the *MC* curve cuts the *AVC* and *ATC* curves at their minimum points.

cases firms use their 'free' sites as a way of encouraging consumers to upgrade to premium sites where additional services are offered at prices that enable the costs of both free and premium services to be covered – where free services are cross-subsidised by premium services. Another model, sometimes used by non-profit organisations, is to ask for donations from customers. Let's say for example you are a classical music lover. While driving home from work in Brisbane, you tune to your favourite local classical music channel, 4MBS Classic. Have you wondered how the operation of this private radio station in Brisbane is funded? Visit their website (https://www.4mbs.com.au/support-us.html) to find out.

Finally, it should be made clear that although the advent of the internet has seen an explosion in the quantity supplied of zero marginal cost services, in the vast majority of cases marginal cost is positive.[1] It is this vast majority of cases that we are dealing with in this book.

What do you think?

1. Is the notion of zero marginal cost more likely to apply to tangible goods or intangible services? Why?
2. Using formulae developed in this chapter explain why, if marginal cost is zero, average variable cost must also be zero. Illustrate with a numerical example.
3. What are some examples of fixed cost that might be associated with online distribution of information or entertainment? (Note that when reference is made above to provision of additional copies of books or movies, it is distribution of the book or movie that we are talking about – not production of the book or movie itself.)
4. Give some examples of requests by non-profit organisations for donations from readers or audience to help cover the costs of providing free services.

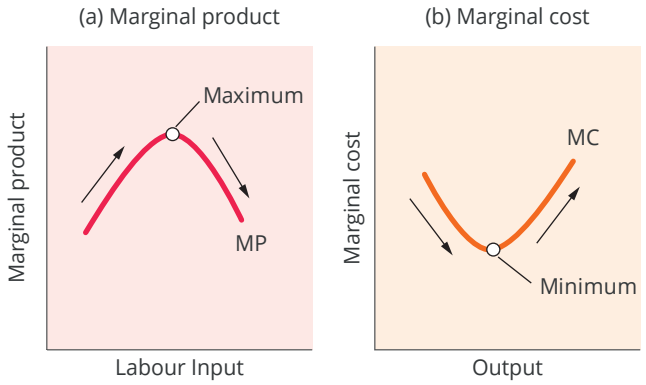
[1] Jeremy Rifkin has written a book about the global implications of the explosion of zero marginal cost services: J. Rifkin, *The Zero Marginal Cost Society: The Internet of Things, the Collaborative Commons, and the Eclipse of Capitalism*, 2014, Basingstoke, Palgrave MacMillan. For some summaries of his argument and some reviews of the book, do an internet search for <Rifkin Zero Marginal Cost Society>.

Now we can apply the marginal-average rule to explain the relationship between the MC and AVC curves. Consider the *MC* curve in part (b) of **Exhibit 6.4**. In the range of output from 0 to 6 units per week, the *MC* curve is below the *AVC* curve and *AVC* is falling. Beyond 6 units per week, the *MC* curve is above *AVC* and *AVC* is rising. Hence, the relationship between *AVC* and *MC* conforms to the marginal-average rule. It follows that the *MC* curve intersects the *AVC* at its lowest point. This analysis also applies to the relationship between the *MC* and *ATC* curves. Initially, the *MC* curve lies below the *ATC* curve, and *ATC* is falling. After seven units of output, the *MC* curve exceeds the *ATC* curve, causing the *ATC* curve to rise.

Marginal cost and marginal product inversely related

We stated earlier that the *MC, AVC* and *ATC* curves are typically U-shaped. What explains the U-shape of these curves? Let us consider the MC curve. Similar explanations apply to the AVC and ATC curves. **Exhibit 6.6** shows how the shape of the *MC* curve in part (a) is *inversely* related to the shape of the marginal product (*MP*) curve in part (b). Here is the explanation. In an earlier section, we have discussed the shape of Fourmet's MP curve. The *MP* of labour input rises as usage of labour rises due to specialisation as discussed in section 2. Since labour productivity and cost of output are inversely related, when *MP* rises in part (a), *MC* falls in part (b). Hence, specialisation explains both the rising part of the *MP* curve and the falling part of the *MC* curve. When labour's *MP* is at its maximum, *MC* of output is at its minimum. After the maximum point of the *MP* curve, diminishing returns set in, and *MP* of labour input declines when more labour is being used. This also causes *MC* of output to rise after its minimum point. Hence, the law of diminishing returns explains both the falling part of the MP curve and the rising part of the *MC* curve.

Exhibit 6.6 **The inverse relationship between marginal product and marginal cost**



Part (a) represents the marginal product of labour (*MP*) curve. At first, each additional employee hired adds more to output than does the previously hired employee and the *MP* curve rises until a maximum is reached. Then the law of diminishing returns sets in and each additional employee hired adds less output than previously hired employees and MP curve falls.

Part (b) shows the marginal cost (*MC*) curve as a U-shaped curve that is inversely related to the *MP* curve. Since labour productivity and cost of output are inversely related, as the *MP* curve rises, the *MC* curve falls. When the MP curve reaches a maximum, the *MC* curve is at a minimum. As diminishing returns set in and the *MP* curve falls, the *MC* curve rises.

In summary

- The **marginal-average rule** explains the relationship between marginal cost and average cost. When the marginal cost is less than the average cost, the average cost falls. When the marginal cost is greater than the average cost, the average cost rises. Following this rule, the marginal cost curve intersects the average variable cost curve and the average total cost curve at their minimum points.
- **Marginal cost** and **marginal product** are inversely related to each other. Specialisation causes marginal cost to fall and the law of diminishing returns causes marginal cost to rise. The two principles also explain the U-shaped marginal cost curve.

5 Long-run production costs

As explained earlier in this chapter, in the long run Fourmet can change the quantity of all inputs, including the size of its kitchen. A firm can, given sufficient time, build a larger or smaller factory or vary the capacity of its machinery. In this section we will discuss how a decision to change the size of a firm (e.g., its factory and the machinery in it) determines the relationship between output and costs in the *long run*.

Long-run average cost curves

Suppose Fourmet is making its production plans for the future. Taking a long-run view of production means that the firm is not locked into a small-, medium- or large-sized commercial kitchen. However, once the kitchen is built in any of these sizes, the firm operates in the short run because the kitchen has become a fixed input.

Exhibit 6.7 illustrates, for simplicity, a condition in which there are only three possible commercial kitchen sizes Fourmet might select. Short-run cost curves representing these three possible plant (commercial kitchen) sizes are labelled $SRATC_s$, $SRATC_m$ and $SRATC_l$. *SR* is the abbreviation for short run, and *ATC* stands for average total cost. The subscripts *s*, *m* and *l* represent small, medium and large plant size, respectively. In the previous sections, there was no need to use *SR* for short run because we were discussing only short-run cost curves and not long-run cost curves. The perceptive reader may notice that in this diagram the SRATC curve tends to be higher for small and large plants compared to the medium-size plant. We will come back to this point after discussing the relationship between the short-run and long-run cost curves.

The optimal plant size will be the one that minimises the average total cost at the chosen the level of output. Suppose Fourmet estimates that it will be producing an output level of six units per week for the foreseeable future. Which plant size should the company choose? It will build the plant size represented by $SRATC_s$ because this affords a lower cost of $30 per unit (point *A*) than the plant size represented by $SRATC_m$, which is $40 per unit (point *B*).

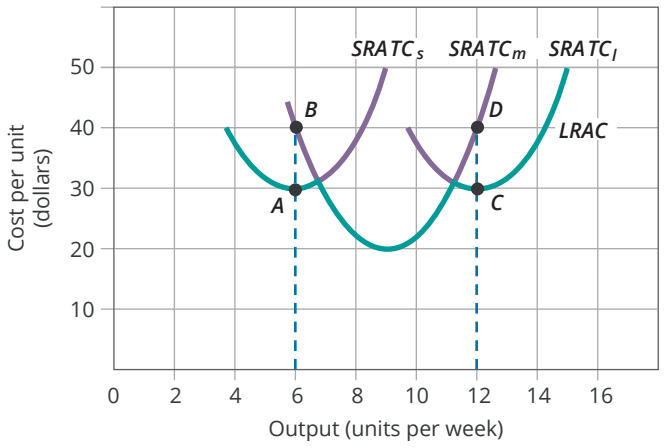
What if production is expected to be 12 units per week? In this case, the firm will choose the plant size represented by $SRATC_l$. This plant size gives a cost of $30 per unit (point *C*) rather than a cost of $40 per unit (point *D*). Do you know Fourmet's optimal plant size if the output level is 9 units per week?

Long-run average cost curve
The curve that traces the lowest cost per unit at which a firm can produce any level of output when the firm can build any desired plant size.

Using the three short-run average cost curves shown in **Exhibit 6.7**, we can construct the firm's **long-run average cost curve** (*LRAC* curve). The long-run average cost curve traces the lowest cost per unit at which a firm can achieve at any level of output when the firm is in a position to build any desired plant size, that is, in the long run. The *LRAC* curve is sometimes called the firm's planning curve. In **Exhibit 6.7**, the green curve represents the *LRAC* curve.

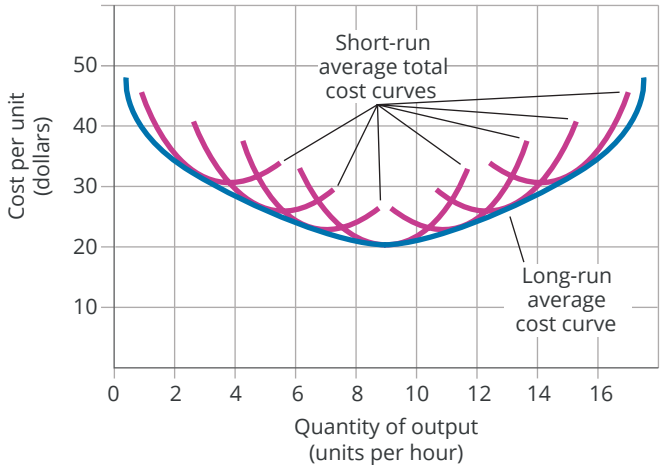
Exhibit 6.8 shows the case where there are an infinite number of possible plant sizes from which managers can choose in the long run. As the number of short-run average total cost curves corresponding to these

Exhibit 6.7 **The relationship between three factory sizes and the long-run average cost curve**



Each of the three short-run *ATC* curves in the figure corresponds to a different plant size. Assuming these are the only three plant sizes that can be chosen, a firm can choose any one of these plant sizes in the long run. For example, a young firm may operate a small plant represented by the U-shaped short-run average total cost curve $SRATC_s$. As a firm matures and demand for its product expands, it can decide to build a larger factory, corresponding to either $SRATC_m$ or $SRATC_l$. The long-run average cost (*LRAC*) curve is the green, scalloped curve joining the lowest parts of the short-run curves at different levels of output.

Exhibit 6.8 **The long-run average cost curve when the number of factory sizes is unlimited**



If there are an infinite number of possible short-run *ATC* curves that correspond to different plant sizes, the long-run average cost (*LRAC*) curve is the blue curve tangent to each of the possible red short-run *ATC* curves.

different-sized plants increases, the lumps in the *LRAC* curve shown in **Exhibit 6.8** disappear. With an infinite number of plant sizes, the corresponding short-run *ATC* curves trace a smooth *LRAC* curve. Notice that on the falling part of the *LRAC* curve, the points of tangency between the long-run and short-run curves are to the left of the minimum points on the short-run *ATC* curves. On rising part of the *LRAC* curve, the tangency points are to the right of the minimum points on the short-run *ATC* curves.

Increasing, constant and decreasing returns to scale

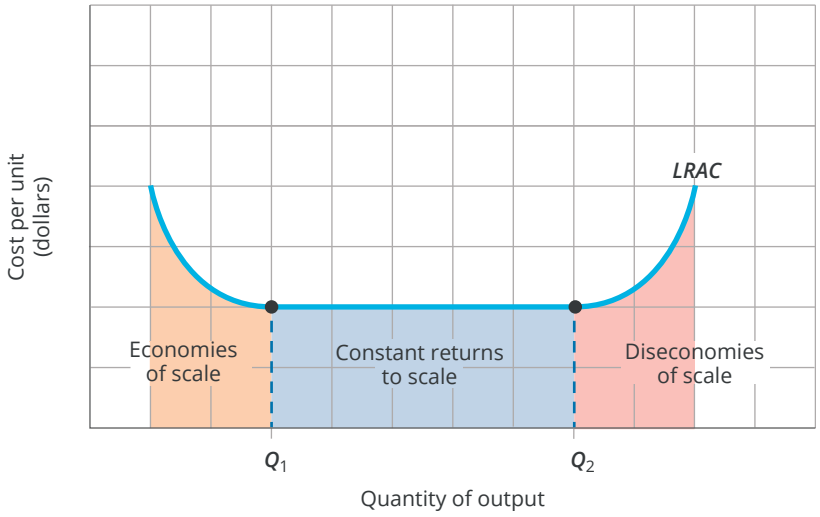
Now, we will discuss the reasons why *LRAC* first falls and then rises when output expands in the long run. For simplicity, **Exhibit 6.9** excludes the short-run *ATC* curves that touch points along the *LRAC* curve. As the scale of operation expands, the *LRAC* curve can exhibit three different patterns. Over the lowest range of output up to Q_1, the firm experiences *increasing returns to scale*, also known as **economies of scale**. Economies of scale exist when the long-run average cost curve declines as the firm increases output.

Economies of scale
A situation in which the long-run average cost curve declines as the firm increases output.

There are several reasons for economies of scale. First, a larger firm can increase the *division of labour* and *use of specialisation*. Adam Smith noted in *The Wealth of Nations*, published in 1776, that the output of a pin factory is greater when, rather than having each employee make a complete pin, one employee draws the wire, a second straightens it, a third cuts it, a fourth forms the point and a fifth makes the head. As a firm initially expands, higher output and having more employees allows managers to break a job into small tasks. Then each employee – including each manager – can specialise by mastering narrowly defined tasks rather than trying to be a jack-of-all-trades.[3] The classic example is Henry Ford's assembly line, which greatly reduced the cost of producing motor cars. Large scale of production and specialisation has been the driver behind mechanisation and technological advancement that explains much of the economic growth in the last two centuries.

[3] A. Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations*, 1776, reprint, Random House, New York, 1937, pp. 4–6.

Exhibit 6.9 **A long-run average cost curve including constant returns to scale**



The long-run average cost (*LRAC*) curve illustrates a firm that experiences economies of scale until output level Q_1 is reached. Between output levels Q_1 and Q_2, the *LRAC* curve is flat and there are constant returns to scale. Beyond output level Q_2, the firm experiences diseconomies of scale and the *LRAC* curve rises.

Second, economies of scale result from greater *efficiency of capital*. Suppose Fourmet can pack their gourmet meals with either machine A or B. Machine A is cheaper to purchase, has a smaller capacity and produces 1000 units per week at an average cost of $2 per unit. Machine B is technologically more advanced and has a capacity of 2000 units per week at an average cost of $0.80 per unit. However, if machine B is operated below its full capacity, say producing only 1000 units per week, the average cost will be much higher at $2.50 per unit. A low-output Fourmet will find it non-economical to purchase machine B, so it uses machine A and its average cost of packing is $2. If Fourmet operates at a higher output rate of 2000 units per week, it will be more cost effective to purchase the more advanced machine B and pack more efficiently at a per-unit cost of only $0.80.

The scale of operation is important for competitive reasons. Consider a young firm producing less than output Q_1 and competing against a more established firm that has reaped all economies of scale and is now producing in the range of output between Q_1 and Q_2. The *LRAC* curve shows that the older firm may have a cost advantage.

You're the economist — Just what are economies of scale?

The economist's concept of economies of scale is one of the most misunderstood and misused economic ideas of all. If you Google <economies of scale example> a significant proportion of the explanations of the term and the examples used to explain it are just plain wrong! For example, you are likely to find sites that tell you that economies of scale occur when a firm increases its production such that its fixed costs are spread over a larger amount of output leading to lower fixed cost per unit output. A specific example might be a bakery that, after undertaking a large advertising program, finds increased demand for its product causes its per-loaf costs to go down as sales and output increase. After you have done some further research, explain why examples like these are not describing economies of scale. What is the economic concept that these explanations are describing? Can you think of a definition of economies of scale that is acceptable to economists but also accurately explains the concept in a way which is easily understood by the non-economist?

The *LRAC* curve may not turn upward immediately and form the U-shaped cost curve in **Exhibit 6.8**. Between some levels of output, such as Q_1 and Q_2 in **Exhibit 6.9**, the firm increases its plant size but the *LRAC* curve remains flat. Economists call this **constant returns to scale**. Constant returns to scale exist when long-run average cost does not change as the firm increases its size. A firm with reserve capacity can exhibit constant returns to scale and maintain its LRAC at the same level when expanding its scale of operation.

Constant returns to scale A situation in which the long-run average cost curve is horizontal as the firm increases its scale of operation.

Applicable concept: economies of scale – What is CoBuy?

Analyse the issue

This is the question which CoBuy asks on its 'About' page on its website. CoBuy is a Seattle-based online platform designed to help home buyers share in the ownership of a home. In an era of increasing unaffordability of housing, CoBuy assists buyers to purchase a share in a house rather than being the sole owner. They assist by providing advice on how to deal with real estate agents, lawyers, mortgage providers, local authorities and so on.

Whereas, in times gone by, it was relatively easy for individuals and couples to purchase their own home, today, in many world cities, the situation is rather different with rising house prices in major capital cities across the world. Individuals and couples may find that it is only when they share ownership with one or a number of additional people that they can afford to get into the housing market. There are advantages of co-buying a property (which we will discuss in the next paragraphs) but at times there are many financial and legal potholes to avoid or deal with. CoBuy provides a service to help co-buyers navigate these barriers.

As a way of explaining the benefits of their services, CoBuy invokes the notion of economies of scale.

The benefits of economies of scale are explained by first pointing out that if a reader were to buy a one-bedroom home in Seattle on their own it would typically cost around $370 000. The blogger then asks readers to think bigger and consider joining with two friends to buy a three-bedroom home. With a three-bedroom home in Seattle averaging around $640 000, the cost per person is just under $215 000, or roughly 42 per cent less than the cost of a one-bedroom home. As the blogger declares: 'Woah! What a difference!'.[1]

In economic terms, as housing services are supplied to more occupiers by increasing the size of the home, the cost per resident goes down. Furthermore, assuming that three occupiers in a three-bedroom home share appliances, such as a stove, refrigerator, dish washer and a hot water system, the per person costs of these appliances is also likely to be lower. This will usually be true even if larger models of these appliances are installed.

Note that the concept of economies that is being written about in the blog is not short-run increasing returns that occur when output from a plant of a given size is increased; that is, they're not talking about what happens when more people occupy a house of a *given* size. Rather, they are describing long-run increasing returns or economies of scale that occur when the scale of a plant – in this case the size of the home, increases. They are talking about economies associated with a larger number of outputs (of services to occupants) occurring in a larger scale of plant (a larger home).

And what is more, these findings about economies of scale in housing are not just relevant to the owners. If more people live in larger houses with more occupants, there will likely be environmental benefits as fewer resources – including building materials and energy – are used to provide housing services to each individual. Do you remember what economists call these benefits external to the occupants?

Shutterstock.com/Richard Cavalleri

What do you think?

1 Explain why the long-run average cost curve for services to householders falls (i.e., economies of scale occur) as the size of houses increases. Draw a diagram that shows this effect using both short-run and long-run average total

cost curves. When drawing this diagram, assume that there are only two scales of plant available – large houses and small houses.

2 On the diagram you've drawn in the previous question, illustrate whether a small number of occupants in a large house results in lower average costs than a large number of occupants in a small house.

3 When the scale of operation (i.e., size of the house) continues increasing from three to four or more bedrooms, do you think diseconomies of scale will occur? What are they? Illustrate diseconomies of scale in this situation with additional short-run average cost curves.

4 When the authors of this text were students, many of their friends got together with four or five others to share large, old houses in inner-city locations. Since then a large amount of purpose-built student accommodation has been constructed. Typically, this accommodation takes the form of one-bedroom apartments occupied by only one or two people. *Ceteris paribus*, do you think that costs of providing services to students will have decreased or increased as a result of this change? Are there other factors unrelated to costs of services that might have been responsible for this change?

[1] **You can find the CoBuy blog at http://blog.gocobuy.com/economies-scale-housing.**

Diseconomies of scale
A situation in which the long-run average cost curve rises as the firm increases output.

As a firm becomes large and expands output beyond some level, such as Q_2 in **Exhibit 6.9**, it may encounter *decreasing returns to scale* also known as **diseconomies of scale**. Diseconomies of scale exist when the long-run average cost curve rises as the firm increases its scale of operation. The main reason is that a large-scale firm becomes harder to manage. As the firm grows, the chain of command lengthens and communication becomes more complex. Firms become too bureaucratic and operations bog down in red tape. Layer upon layer of managers are paid to shuffle paper and read emails that may have little or nothing to do with producing output. Consequently, it is no surprise that a firm can become too big and these management problems can cause productivity to fall and the average cost of production to rise. The owner of a firm in this situation may decide to break up the firm so that it consists of a number of smaller, more manageable units.

To summarise this section, we quote what the late Steve Jobs, founder of Apple Computer Company, has to say on bureaucracy:

> When you are growing [too big], you start adding middle management like crazy … People in the middle have no understanding of the business, and because of that, they screw up communications. To them, it's just a job. The corporation ends up with mediocre people that form a layer of concrete.[4]

In summary

- The **long-run average cost** curve is a curve drawn at a tangent to all possible short-run average total cost curves.
- When the long-run average cost curve decreases as output increases, the firm enjoys **economies of scale**.
- If the long-run average cost curve remains unchanged as output increases, the firm experiences **constant returns to scale**.
- If the long-run average cost curve increases as output increases, the firm experiences **diseconomies of scale.**
- The plant size selected by a firm in the long run will be the one that minimises cost for the expected level of production.

[4] D. Wise and C. Harris, 'Apple's new crusade', *Business Week*, 26 November 1984, p. 156.

Key concepts

Explicit costs
Implicit costs
Economic profit
Long run
Short run
Fixed input
Variable input
Production function
Marginal product
Law of diminishing returns
Total fixed cost
Total variable cost
Total cost
Average fixed cost
Average variable cost
Average total cost
Marginal cost
Marginal-average rule
Long-run average cost curve
Economies of scale
Constant returns to scale
Diseconomies of scale

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Distinguish between economist's and accountant's definitions of costs and profit

- **Economic profit** is equal to total revenue minus total cost (including both **explicit** and **implicit** costs).
- Accounting profit equals total revenue minus *only* explicit costs.
- Business decision-making must be based on economic profit rather than accounting profit.

2. Understand how short-run production theory provides the foundation for short-run production costs

- A **production function** is the relationship between inputs and outputs.
- The short run is the timeframe within which a firm has at least one input that cannot be varied (**fixed input**).
- The short-run production function shows how total output changes as the amount of **variable input**, such as labour, varies, holding all other factors of production constant.
- **Marginal product** is the change in total output caused by a one-unit change in a variable input, such as labour.
- The **law of diminishing returns** states that in the short run, after some level of output in the short run, each additional unit of the variable input yields smaller and smaller **marginal product**.

3. Explore how short-run costs change as output changes

- **Fixed cost** consists of costs that do not vary with the level of output, such as rent for office space and general insurance.
- **Variable cost** consists of costs that vary with the level of output, such as wages and cost of raw materials.
- **Total cost** is the sum of **total fixed cost** and **total variable cost** at each level of output.
- **Marginal cost** is the change in total cost associated with one additional unit of output.
- **Average fixed cost** is the total fixed cost divided by total output at each level of output.
- **Average variable cost** is the total variable cost divided by total output at each level of output.
- **Average total cost**, or the sum of average fixed cost and average variable cost, is the total cost divided by total output at each level of output.

4. Understand the marginal-average rule and the inverse relationship between marginal product and marginal cost

- The **marginal-average rule** explains the relationship between marginal cost and average cost.
- When the marginal cost is less than the average cost, the average cost falls. When the marginal cost is greater than the average cost, the average cost rises. Following this rule, the marginal cost curve intersects the average variable cost curve and the average total cost curve at their minimum points.
- Marginal cost and marginal product are inversely related to each other.
- Specialisation and the law of diminishing returns explains the inverted U-shape of the *MP* curve and the U-shaped MC curve.

5. Understand the long-run production costs

- In making a long-run production decision the firm selects the plant size that minimises average cost for the planned level of production. The **long-run average cost** curve is a curve drawn at a tangent to all possible short-run average total cost curves. **Economies of scale**, **constant returns to scale**, and **diseconomies of scale explains** the U-shape of the long-run average cost curve.

Study questions and problems

1 Explain why the marginal cost curve cuts the average variable cost curve at its lowest point.

2 What are some of the fixed inputs and variable inputs you would use in operating a small gift shop?

3 Explain conditions under which some type of labour might be treated as a fixed cost and conditions under which some would be treated as a variable cost.

4 Indicate whether each of the following is an explicit cost or an implicit cost:

a Payments to Officeworks for stationery.
b Interest forgone when an owner uses her own saving to run her own business.
c Wages forgone by children helping parents run their family business.
d Income forgone while caring for young children.
e Workers compensation policy payments a company makes for its employees.
f A manager's salary.

5 Using the following table complete tasks (a) and (b). All costs are in dollars.

a Complete the cost schedules for a firm operating in the short run. Do you know its fixed cost?
b Graph the average variable cost, average total cost and marginal cost curves.

Q	*TC*	*TVC*	*ATC*	*AVC*	*MC*
0	60				
1	80	20	80	20	
2	90	30		15	10
3	102	42	34		12
4	120		30	15	
5	145	85	29		25
6	180	120		20	35
7	228	168	32.6	24	
8	300		37.5		72

6 The operation of ride-sharing company Uber has been legalised in all states in Australia since 2017. Before then it continued to make very large losses. Can you think of a reason why their managers allow this to occur?

7 Explain why when marginal product is increasing marginal cost will be decreasing, and vice versa.

8 Refer to the below production function table.
   a Construct the marginal product schedule for the production function.
   b Graph the total output and marginal product curves and identify increasing and diminishing marginal returns.

Labour	Total output	Marginal product
0	0	
1	8	
2	18	
3	30	
4	43	
5	55	
6	65	
7	73	
8	79	
9	82	
10	80	

9 When we look at road vehicles used for transporting goods we find that in each market there is often a similarity in the size of vehicle used. Thus in the interstate trucking market, large semi-trailers and B-doubles are commonly employed while parcel transport in suburban areas sees the use of small trucks and vans. Can cost theory help explain why this is so?

Answers to 'You're the economist'

A profitable newsagent?

In Mr Chen's newsagent, the accounting profit is $104 700. An accountant would calculate profit as the total revenue (in this case, annual sales revenue of $765 400, plus commission from Lotto and other suppliers of $101 000) less the sum of the explicit costs of $761 800 per year. The explicit costs include costs of goods sold, commission, rent, wages and other expenses. However, the accountant would neglect implicit costs. The implicit costs of Mr Chen's business would include his forgone salary of $75 000, the forgone wages of his wife and children (not provided), and $6000 in forgone annual interest on the $120 000 he took out of his savings to purchase the business. Mr Chen's wife worked full-time and his children worked part-time in the newsagent. If their forgone wages added up to more than $23 700 ($104 700 − $75 000 − $6000) per annum, then Mr Chen made an economic loss. You visited the ABS website and found that the average full-time earning in 2018 was nearly $85 000.[5] If you assessed that their total forgone wages would likely exceed $23 700 and concluded that Mr Chen made a loss by including implicit costs, then YOU ARE THINKING LIKE AN ECONOMIST.

Just what are economies of scale?

The problem with many explanations of economies of scale is that their authors confuse short-run increasing returns with economies of scale. If the average costs decrease in the short run as output from a given plant increases then short-run

[5] Australian Bureau of Statistics, 'Employee earnings and hours, Australia', May 2018, http://www.abs.gov.au/ausstats/abs@.nsf/mf/6306.0

increasing returns are being experienced. This is what the bakery example describes. Only when average costs decline as the size of the plant being used increases (the size of the bakery increases) are economies of scale occurring. By definition, increasing the size of their plant means the firm is operating in the long run. A definition that is acceptable to economists yet is easily understood by non-economists is one which insists that economies of scale occur only when output increases *and* when the scale of the plant used by the firm also increases. If your explanation was similar to this, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 The law of diminishing returns applies
- **a** only in the long run.
- **b** only in the short run.
- **c** in both the short run and the long run.
- **d** when there is no fixed input.

2 The following table shows the total output produced by different numbers of workers in a shoe factory.

Number of workers	Output (pairs per day)
0	0
1	40
2	90
3	145
4	205
5	250
6	280

- **a** Refer to the table above. What is the marginal product of the third worker?
  - **i** 145 pairs of shoes
  - **ii** 90 pairs of shoes
  - **iii** 65 pairs of shoes
  - **iv** 55 pairs of shoes
- **b** The firm experiences diminishing returns after how many workers?
  - **i** 2 workers
  - **ii** 3 workers
  - **iii** 4 workers
  - **iv** 5 workers

3 Suppose a small clothing manufacturer has 10 sewing machines and 15 employees and is able to make 60 garments per day. When it adds an eleventh sewing machine, but no more employees, it is able to make 61 garments per day. The marginal product of the eleventh sewing machine is:
- **a** 61 garments per day.
- **b** 6.1 garments per day.
- **c** 1 garment per day.
- **d** 4.066 garments per day.

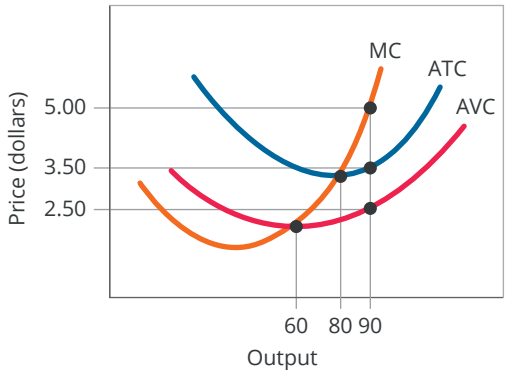
4 If the marginal cost curve is above the short-run average total cost curve, then as output increases the marginal cost curve must be:
- **a** horizontal.
- **b** rising.
- **c** falling.
- **d** none of the above.

5 If both the marginal cost and the average variable cost curves are U-shaped, at the point of minimum average variable cost the marginal cost must be
- **a** at its minimum.
- **b** rising.
- **c** equal to the average variable cost.
- **d** both rising and equal to the average variable cost.

6 Which of the following is true at the point where diminishing returns set in?
- **a** Marginal product is at a maximum and marginal cost at a minimum.
- **b** Marginal product is at a minimum and marginal cost at a maximum.
- **c** Both marginal product and marginal cost are at a maximum.
- **d** Both marginal product and marginal cost are at a minimum.

7 As shown in **Exhibit 6.10**, average total cost of producing 90 units of output is
- **a** below $2.50
- **b** $2.50
- **c** $3.50
- **d** $5.00

Exhibit 6.10 **Cost curves of a firm**



8 As shown in **Exhibit 6.10**, total fixed cost for the firm is:
- **a** zero.
- **b** $60.
- **c** $80.
- **d** $90.
- **e** $135.

9 The long-run constant returns to scale exist when:
- **a** the short-run average total cost curve is constant.
- **b** the costs of some inputs are constant.
- **c** the long-run fixed cost is constant.
- **d** the long-run average cost curve is flat.

PART 3

MARKET STRUCTURES

7 Perfect competition

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Distinguish the four types of market structures.
2 Describe the characteristics of perfect competition.
3 Analyse how a competitive firm determines its short-run profit maximisation output.
4 Determine a competitive firm's short-run loss minimisation output.
5 Derive the firm's supply and market supply under perfect competition in the short run.
6 Explain the difference between short-run and long-run equilibrium for a competitive firm.
7 Distinguish the three types of long-run market supply curve under perfect competition.

Towards the end of last century there was a boom in investment in timber plantations in countries with suitable soil and climate. Nowhere was this more apparent than in Australia, where the area of land under plantation grew eightfold during the 1990s. A number of factors led to this spectacular growth. They included favourable tax treatment that lowered growers' costs, falling prices for alternative agricultural outputs, recognition of the role that forests play in reducing greenhouse gases, and increasing dissemination of information about the highly profitable plantation investments.

This boom in plantations saw continuing investment in traditional pine plantations as well as innovative ventures involving species such as teak, sandalwood, blue gum and blackwood. The combination of tax benefits and the belief that demand would outstrip supply from existing sources led to expectations of handsome profits down the track when harvesting occurred.


Shutterstock.com/Ecopix

Suddenly, however, at the turn of the century, some commentators were warning that the high profits expected might never materialise. What factors were causing the alarm bells to ring? The problem appeared to be that too many growers were able to enter the market and supply timber of equal quality to that already being supplied. As the critics pointed out, thousands of farmers and

graziers in Australia alone had suitable land and could easily enter the plantation timber industry. Furthermore, information on how to develop, maintain and harvest plantations was readily available and, so long as plantation management standards were maintained, each grower would be in a position to produce high-quality timber in significant quantities and eventually drive the price of timber down.

Just a few years later, in 2004, it was observed in the business pages of *The West Australian* that the outlook for the plantation timber industry was again positive, with an expansion of sales and increased profitability encouraging further investment.[1] A few years later, however, the outlook has again become bleak with the collapse of a number of the country's largest plantation companies as a result of over-investment, legislative changes and the onset of the global financial and economic crisis (GFEC) in 2008. In 2014 it was reported that over 46 000 hectares of land on which plantation timber was growing was to be sold, with much of it expected to be returned to its former role as beef grazing land.[2] By 2017, however, poor management of logging in native forests and the ensuing closure of a number of native forest saw mills had led to renewed optimism about the future of plantation timbers.[3]

The ups and downs of timber plantations reflect changing consumer preferences and government tax policy and, most relevant to the focus of this chapter, they illustrate firms' decisions to channel resources in and out of an industry in response to profit opportunities. Timber plantation is a highly competitive industry. In the 18th century, foundational economist Adam Smith concluded that a competitive market acts like an 'invisible hand' that leads people who pursue their own interests to also advance the interests of society. In this chapter we will fully examine the role of the invisible hand in the perfectly competitive market, and you will see why all developed countries have legislations to foster a competitive economic environment.

We will first present how economists classify market types for understanding the behaviour of firms in different market conditions. Combining the concepts of demand and supply, cost of production and marginal analysis from previous chapters we will examine the working of the *perfectly competitive market* and why competition is so good. Other types of markets (*monopoly, oligopoly and monopolistic competition*) will be discussed in subsequent chapters.

1 Market structures

Firms sell goods and services under different market conditions. Economists categorised four sets of market conditions, which are called **market structures**. Examination of the business sector of our economy reveals firms facing differing degrees of competition because they operate in very different market structures. The defining characteristics of a market structure include (1) the number of firms in that market, (2) the degree to which the products they sell are similar, and (3) the ease of entry into and exit from the market. These characteristics influence the degree of competition faced by the firm and their behaviour in their particular markets.

Market structure
A classification system for the key characteristics of a market, including the number of firms, the similarity of the products they sell and the ease of entry into and exit from the market.

In **Chapter 3**, where the demand and supply model was presented, we have already discussed one market structure – perfect competition, although neither the concept of market structure nor the term perfect competition was used explicitly. Perfect competition is characterised by a large number of sellers producing identical product, each facing keen competition from its competitors. As shown in **Exhibit 7.1**, at the other extreme of the market structure spectrum is the type of market dominated by a single seller, which is called a monopoly. Between these two extremes we have monopolistic competition and oligopoly. Monopolistic competition is similar to perfect competition except that the products sold by the firms are close substitutes but

1 C. Bolt, 'Agribusiness agriboom', *The West Australian*, 20 November 2004, p. 86.
2 S. Goodwin, 'Up for grabs: timber country hits market', *The Land*, 14 March 2014, https://www.theland.com.au/news/agriculture/general/news/up-for-grabs-timber-country-hits-market/2691285.aspx, accessed 17 September 2014.
3 See, for example, E. Kirk, 'Nannup sawmill operation to shut down', *Busselton-Dunsborough Mail*, 12 September 2017, http://www.busseltonmail.com.au/story/4875440/nannup-sawmill-operation-to-shutdown/, accessed 12 September 2017.

not identical. Because of the large number of firms selling similar products, like perfect competition there is also a high degree of competition in monopolistic competition. On the other hand, oligopoly is similar to monopoly, but this time the market is dominated not by one but by a few firms. Each of these market structures has certain key characteristics that enable economists to analyse the way in which firms behave in that market. In this and subsequent chapters we will study each of the four market structures in detail and in the process we will see how different market structures bring out differences in the behaviour of firms. The first market structure we examine is *perfect competition*, to which this entire chapter is devoted.

Exhibit 7.1 **Comparison of market structures**

Market structure	Number of sellers	Type of product	Entry conditions	Examples
Perfect competition	Large	Homogeneous	Very easy	Small crops, international commodity markets
Monopolistic competition	Large	Differentiated	Easy	Boutiques, restaurants, motels
Oligopoly	Few	Usually differentiated but sometimes homogeneous	Difficult	Car making, tobacco products, oil
Monopoly	One	Unique	Extremely difficult	Public utilities

In summary

- The defining characteristics of a **market structure** include the number of firms in that market, the degree to which the products they sell are similar and the ease of entry into and exit from the market.
- The market structure will influence how firms make pricing and output decisions.

2 Perfect competition

Perfect competition A market structure characterised by: (1) a large number of small firms, (2) a homogeneous product and (3) very easy entry into or exit from the market.

Perfect competition is the most competitive market structure. It is characterised by: (1) a large number of small firms, (2) a homogeneous product and (3) ease of entry into or exit from the market. We now look at each of these defining characteristics in more detail.

Large number of small firms

How many sellers constitutes a large number? Certainly two, three or four firms in a market would not be a large number. In fact, the exact number cannot be stated. This condition is fulfilled when each firm in a market has no significant share of total output and therefore any change to the firm's output will have negligible impact on the market output. Hence, an individual firm in perfect competition has no ability to affect the product's market price. It follows that each acts independently without having to consider how its competitors react to this output or pricing decision. This condition is obviously not true if the industry only has a small number of firms, as in oligopoly, which we will examine in **Chapter 9**.

Homogeneous product

In a perfectly competitive market, all firms produce a standardised or homogeneous product. This means that the goods or services of all the firms are identical. Farmer Chek's stone fruit is identical to farmer Maureen's. Similarly, buyers may believe that the services of one taxicab are about the same as another's. This assumption rules out successful pursuit of competitive advantage through advertising, quality differences or locational advantage.

Combining the above two characteristics, we can see that a firm in a perfectly competitive market (in short, a competitive firm) will have no influence on the market price as in stone fruit market in Australia where there are hundreds of independent stone fruit growers producing (almost) identical stone fruits. A single stone fruit grower has to *take* the price set in the market; that is, it will not be able to raise or lower the market price by altering the amount of stone fruit it supplies to the market. This point will be further discussed in the next section.

Ease of entry and exit

Easy entry into a market means that a new firm faces no barriers to its entry into that market. Barriers can be financial, technical or government-imposed. These barriers include some classes of licences and permits, as well as patents. Anyone who wants to try his or her hand at a commercial ironing service needs only an iron, a phone and somewhere to do the ironing. Ease of exit from a market means that if a firm decides against continuing to produce, it can easily close down. In other words, there are no contractual or legal reasons why it must continue in a declining industry. We shall see that this ease of exit and entry of firms is important because it means that resources will be channelled into or out of the market in response to price and profit signals generated by changes in demand and supply conditions.

In addition to these assumptions, it is also presumed that economic agents – both businesses and consumers – are well informed about key aspects of the operation of the market. Thus businesses and consumers will both be familiar with the product being sold and its price, while business alone will have a good knowledge of the methods and costs of production. As we shall see later in this chapter, efficient operation of competitive markets does not require that consumers be familiar with production costs.

No real-world market exactly fits the three assumptions of perfect competition. The perfectly competitive market structure is a theoretical or ideal model, but some actual markets do approximate the model fairly closely. Examples include farm produce markets, the interstate road transport market and markets for home services, such as lawn-mowing or cleaning.

If no real-world market exactly fits the assumptions of perfect competition, you may be wondering why so much space is devoted to the model here. The answer is that the model provides a benchmark by which we may judge the structure and performance of markets that we observe in the real world. As you saw in **Chapter 4**, where we discussed market failure, a lack of competition in markets leads to a less efficient outcome than is potentially available if competition prevails. The notion that competition drives market efficiency is so important in contemporary market economies that every such economy in the world has a **competition policy** designed to promote efficiency through the encouragement of competition and the outlawing of anti-competitive practices.

Competition policy Government policy that has the objective of increasing competition in the economy or of encouraging firms that are not competitive to behave as if they were.

In Australia, competition policy is overseen by the Australian Competition and Consumer Commission (ACCC), which is a Commonwealth government statutory authority responsible for ensuring compliance with relevant parts of the *Competition and Consumer Act 2010* (formerly known as the *Trade Practices Act*). This Act has the objective of enhancing the welfare of Australians through the promotion of competition and fair trading, and providing for consumer protection. In performing its functions, the ACCC cooperates with, and complements, relevant state government bodies.

The ACCC's objectives include the following:

- improve competition and efficiency in markets
- foster adherence to fair trading practices in well-informed markets

- promote competitive pricing wherever possible and restrain price rises in markets where competition is less than effective
- inform the community at large about the *Competition and Consumer Act 2010* and its specific implications for business and consumers
- use resources efficiently and effectively.[4]

The New Zealand Commerce Commission (NZCC) has similar objectives, which are encapsulated in the following statement from its chairman:

> Ultimately the Commission exists to ensure New Zealanders and our country as a whole, experience the benefits of competition. Throughout our work we touch the lives of people each and every day. While much of our work is focused on education and compliance to ensure fair and competitive markets and prevent harm, we also operate at the other end of the spectrum, seeking compensation and redress for wrongdoing.[5]

You can see that even if perfect competition is evident only in some industries, the aim of government competition policies is to encourage all industries to become more efficient by behaving *as if they were* competitive.

It is also worth noting that many of the regulatory activities of bodies like the ACCC or the NZCC are designed to curb unethical behaviour in the business sector – behaviour that Adam Smith might have described as displaying unenlightened self-interest. Some of the undesirable business activities that governments do their best to eliminate include colluding to raise prices and profitability, abusing market power at the expense of suppliers and engaging in misleading advertising, some of which will be discussed in **Chapters 8 and 9**.

Having established the importance of the presence of competition in contemporary market economies, we now turn to further development of the model of perfect competition.

In summary

- **Perfect competition** a market structure with a large number of firms, a homogeneous product and ease of entry and exit.

3 Short-run profit maximisation for a competitive firm

Price-taker
A seller that has to accept or take the prevailing market price of its product.

As the first step in building our competitive model, we look at why the competitive firm cannot influence the price at which it sells its output and must accept or take the prevailing market price. If a firm has to take the market price, the firm is a **price-taker**.

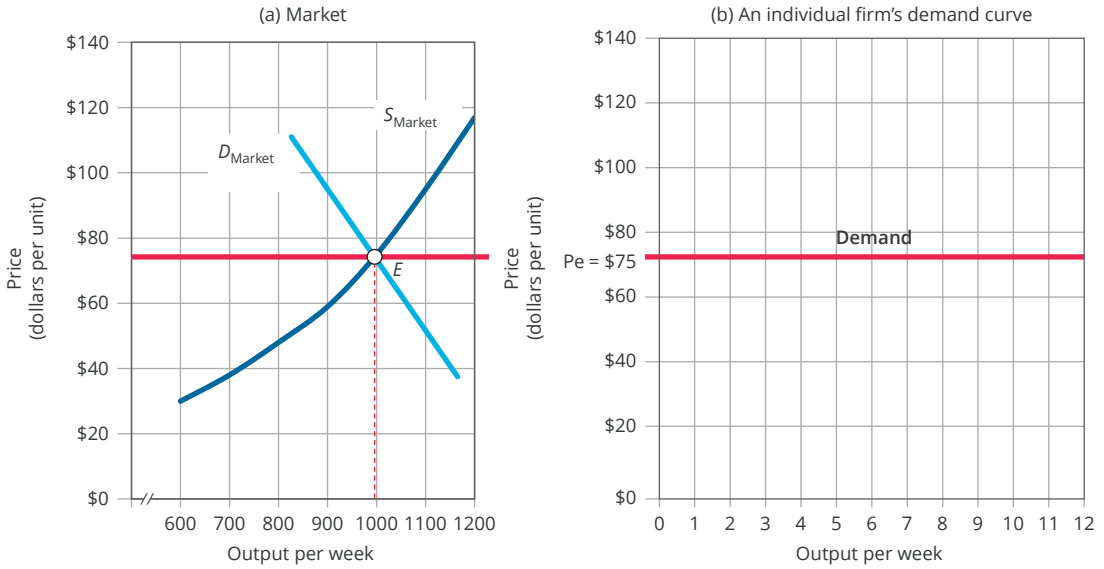
The competitive firm as a price-taker

Consider the frozen meals market in an economy. Part (a) of **Exhibit 7.2** presents the hypothetical market supply and demand for frozen meals. Here we will assume for the sake of discussion that the frozen meals industry is perfectly competitive. As presented in **Chapter 4**, given the market demand and supply as shown in part (a), the equilibrium price is $75 per unit and the equilibrium quantity is 1000 units per week.

4 ACCC, 'About us', https://www.accc.gov.au/about-us/australian-competition-consumer-commission/about-the-accc

5 Commerce Commission New Zealand, *Statement of Intent 2014–2018*, https://comcom.govt.nz/about-us/strategic-planning-and-accountability-reporting/statement-of-intent, accessed 1 November 2017.

Exhibit 7.2 **Market demand and supply, and demand facing the competitive firm**



In part (a), the market equilibrium price is $75 per kg. The competitive firm in part (b) is a price-taker because it is so small relative to the market. At $75, the individual firm faces a horizontal demand curve, *D*. This means that the firm's demand curve is perfectly elastic. If the firm raises its price even one cent, it will sell zero output.

Now we look at the demand curve *faced* by the competitive firm. The competitive firm's demand curve in part (b) of **Exhibit 7.2** is *perfectly elastic* (horizontal) at the prevailing market price of $75 per unit. A horizontal demand curve shows that the firm can sell any level of output at this price. (Note the difference between the firm's units per week and the industry's thousands of units per week; this firm has but a small part of total market output.) Why is the demand curve facing the competitive firm horizontal? Recall from **Chapter 5** that when a firm facing a perfectly elastic demand curve tries to raise its price one cent higher than the going price, no buyer will purchase its product and its quantity demanded falls to zero (see part (a) of **Exhibit 5.3** in **Chapter 5**). The reason is that there are many other firms selling the identical product at that price. Hence, the competitive firm will not set the price above the prevailing market price and risk selling zero output. Nor will the firm set the price below the market price because the firm can sell all it wants to at the going price and, therefore, a lower price would reduce the firm's revenue.

If the firm has no control over price, what does the firm control? The competitive firm makes only one decision – what quantity of output should be produced if profit is to be maximised. In this chapter, profit means economic profit as it was defined in **Chapter 6**. In this section, we develop two methods to determine the profit maximisation output for a competitive firm: the *total revenue–total cost* method and the *marginal analysis* method. The timeframe for our analysis is the short run with some fixed input, such as a farm or factory of a given size.

The total revenue–total cost method

Exhibit 7.3 provides hypothetical data on total revenue and total cost at various output levels for our frozen meals producer, Fourmet. The total cost figures in column three are taken from **Exhibit 6.3** in **Chapter 6**. By looking at total cost at zero output you can see that total fixed cost is $100.

At the prevailing market price of $75 per unit, because Fourmet is a price-taker, the total revenue from selling 1 unit is $75, $150 from selling 2 units, and so on. Total revenue is reported in column 2 of part (a) of **Exhibit 7.3** and is computed as the product price times the quantity. Subtracting total cost in column 3 from total revenue in

column 2 gives the total profit or loss (column 4) that Fourmet earns at each level of output. From the output rate of zero to 2 units per week, the firm earns losses; then a *break-even point* (zero economic profit) occurs at about 3 units per week. If the firm produces 10 units per week, it earns the maximum profit of $250 per week. **Exhibit 7.4** illustrates graphically that maximum profit occurs where the vertical distance between the total revenue and total cost curves is at a maximum. The perceptive reader would notice that Fourmet's profit is also maximised if it produced 9 units of output per week. Does it mean there are two profit maximisation levels of output? This situation of two profit maximisation output levels occurs because we use small discrete output values (1 unit, 2 units etc.). We will return to, and explain, this situation after introducing the marginal analysis method.

Exhibit 7.3 **Short-run profit maximisation for Fourmet**

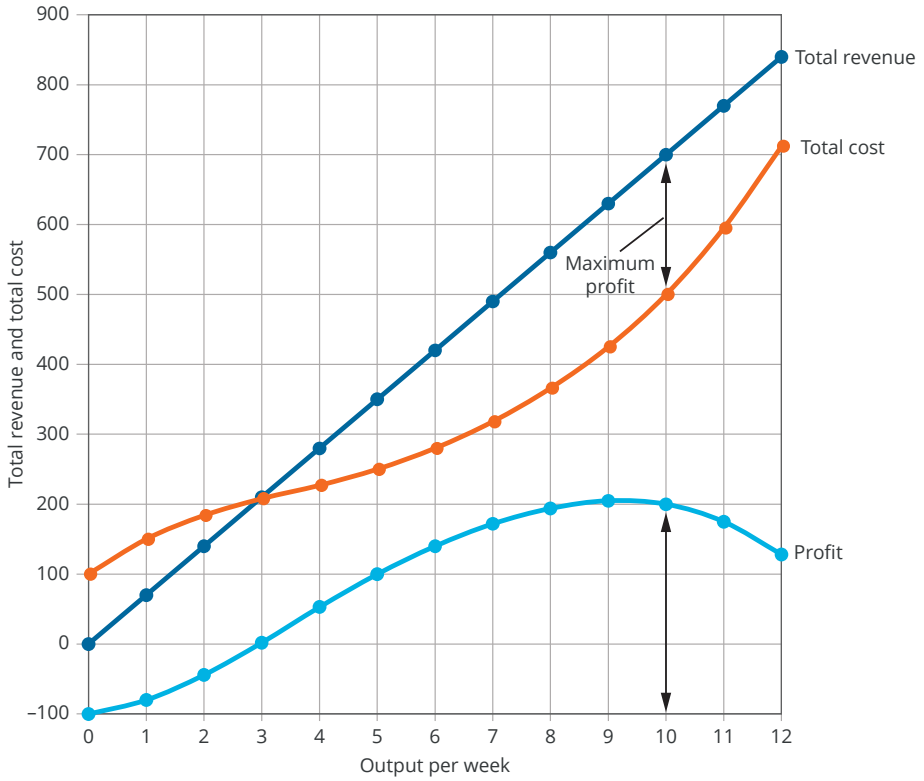
(a) Total revenue–total cost method

Output (units per week)	Total revenue ($)	Total cost ($)	Profit ($)
0	$ 0	$100	−$100
1	75	150	−75
2	150	184	−34
3	225	208	17
4	300	227	73
5	375	250	125
6	450	280	170
7	525	318	207
8	600	366	234
9	675	425	250
10	750	500	250
11	825	595	230
12	900	712	188

(b) Marginal analysis

Output (units per hour)	Marginal revenue ($)	Marginal cost ($)	Average variable cost ($)	Average total cost ($)
1	$75	$50	$50	$150
2	75	34	42	92
3	75	24	36	69
4	75	19	32	57
5	75	23	30	50
6	75	30	30	47
7	75	38	31	45
8	75	48	33	46
9	75	59	36	47
10	75	75	40	50
11	75	95	45	54
12	75	117	51	59

Exhibit 7.4 **Short-run profit maximisation using the total revenue–total cost method**



This exhibit shows the profit-maximising level of output chosen by a competitive firm, Fourmet. Given a market price of $75 per unit, the maximum profit is earned by producing 10 units per week. At this level of output, the vertical distance between the total revenue and total cost curves is the greatest.

The marginal analysis method

A second approach uses *marginal analysis* to determine the profit-maximising level of output by comparing marginal revenue and marginal cost. Part (b) of **Exhibit 7.3** presents the marginal cost data calculated in **Exhibit 6.3** in **Chapter 6**. Recall that marginal cost is the increase in total cost when an additional unit of output is produced.

Now we introduce **marginal revenue** (*MR*), a concept similar to marginal cost. Marginal revenue is the change in total revenue from the sale of one additional unit of output. Mathematically:

Marginal revenue The change in total revenue from the sale of one additional unit of output.

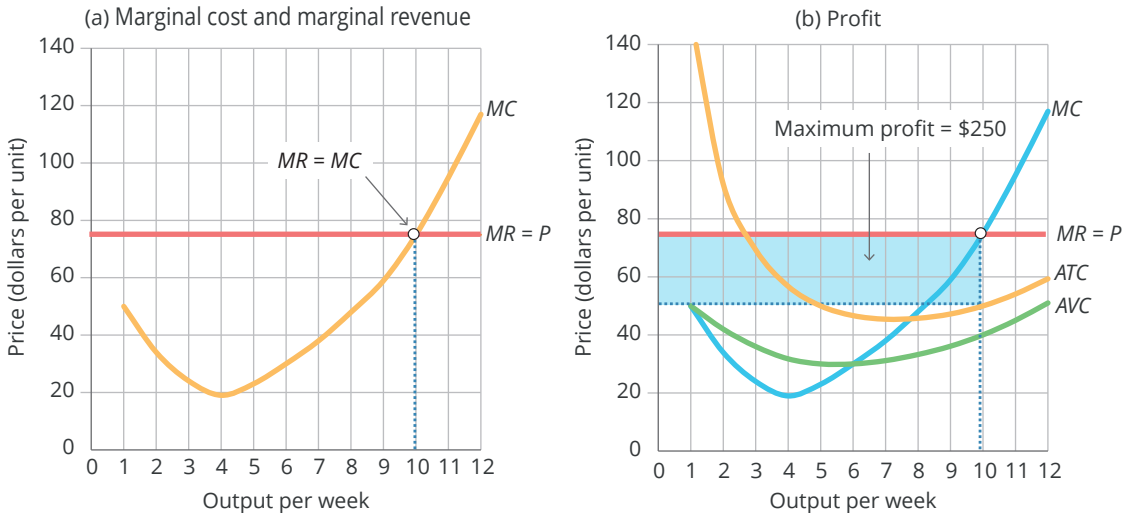
$$MR = \frac{\text{change in total revenue}}{\text{change in quantity}}$$

In our simplified example, since Fourmet is a price-taker facing a horizontal demand curve, it can sell any number of units at the going market price. At the market price of $75 per unit, one additional unit of output always adds $75 to its total revenue. Therefore, Fourmet's marginal revenue is $75, the same as the market price.

Now compare marginal cost and marginal revenue in part (b) of **Exhibit 7.3**. At first, marginal cost is below marginal revenue, and this means that producing each additional unit adds more to total revenue than to total cost, and Fourmet's profit increases. Therefore, Fourmet will continue to increase output until no more *extra*

profit can be made. This occurs at output rate of 10 unit per week where marginal revenue equals marginal cost. For output beyond 10 units per week, marginal cost exceeds marginal revenue and the firm would make a loss from those units of output. For example, producing the eleventh unit will incur a loss of $20, and profit falls from $250 to $230 per week.

Exhibit 7.5 **Short-run profit maximisation using the marginal analysis method**



In addition to comparing total revenue and total cost, the profit-maximising level of output can be found by comparing marginal revenue and marginal cost. As shown in part (a), profit is maximised at output level of 10 units per week where marginal revenue equals marginal cost.

In part (b) the average total cost (ATC) and average variable cost (AVC) curves are added. At output rate of 10 units per week, ATC is $50 per unit. So, Fourmet is making a profit of $25 per unit of output and a total profit of $250 per week (blue shaded rectangle).

Profit maximisation in perfect competition
Occurs when *MR* = *MC*.

We can now state the condition of **profit maximisation in perfect competition:**

> The competitive firm maximises profit by producing output up to the level where marginal revenue equals marginal cost.

Part (a) of **Exhibit 7.5** shows how use of the *MR* = *MC* rule can graphically determine profit maximisation output. As you can see, a horizontal line is drawn at the going market price of $75. Recall this line is the demand curve facing the competitive firm, and is also its MR curve.

Below 10 units of output, the *MC* curve is below the *MR* curve. Profit rises as Fourmet increases output because a rise in output adds more to revenue than it does to cost. Beyond 10 units of output, the *MC* curve is above the *MR* curve and the firm will make a loss if it produces more than 10 units. Profit is therefore maximised at the point where the *MR* curve intersects the *MC* curve.

Now we can return to a point briefly mentioned early in relation to why there are two profit maximisation levels of output. This occurs because we are using discrete output values. If we are dealing with continuous output values, then we can see from part (a) of **Exhibit 7.5** that below output level of 10 units, *MR* > *MC*. After 9 units, should the firm produce unit 9.5th? 9.6th ...? The answers are positive. So the firm should produce up to 9.99 units, which is effectively 10 units. That is why we said that no more extra profit can be made when the firm produces up to the level where *MR* = *MC*.

Now we can graphically study how much profit Fourmet makes. In part (b) of **Exhibit 7.5**, we add the *ATC* and *AVC* curves. At the profit-maximising level of output of 10 units, the vertical distance between the price of

$75 (*MR* curve) and the *ATC* curve is the *profit per unit*. Multiplying the per unit profit by the quantity of output gives the profit of $250 [($75 − $50) × 10 = $250]. The shaded rectangle in part (b) of **Exhibit 7.5** represents the maximum profit of $250 per week. We have arrived at the same profit maximisation amount ($250) as derived by comparing the total revenue and the total cost.

In summary

- A **price-taker** firm in perfect competition faces a perfectly elastic (horizontal) demand curve.
- The *total revenue–total cost method* is one way in which the firm determines the level of output that maximises profit.
- The *marginal analysis method* is a second approach to finding where a firm maximises profits. Using the marginal analysis method, a competitive firm maximises profit when it produces up to the level where *MR* = *MC*.

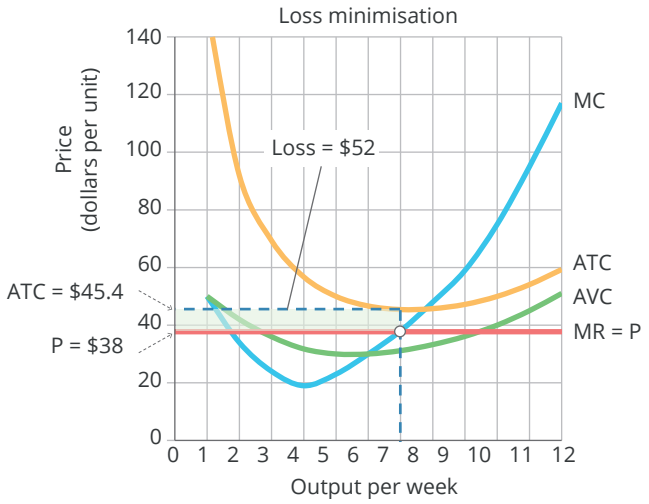
4 Short-run loss minimisation for a competitive firm

The competitive firm must take the price determined by market supply and demand forces. When a change in market conditions causes the market price to fall, the firm can do nothing but adjust its output to make the best of the situation.

A competitive firm facing a short-run loss

Suppose after the COVID-19 pandemic a decrease in the market demand for frozen meals causes the market price to fall to $38 per pack. As a result, the firm's *MR* curve shifts downward to the new position where the price is $38 per unit, as shown in **Exhibit 7.6**. In this case, there is no level of output at which the firm earns a profit because the market

Exhibit 7.6 **Short-run loss minimisation**



If the market price is less than the average total cost, the firm will produce a level of output that keeps its loss to a minimum. The *MR* = *MC* rule also applies to determine the loss minimisation output. This occurs at 7 units of output per week.

price (P) is below the *ATC* curve at any level of output. The firm must now decide whether to continue to operate or to shut down. If it continues to operate it must also decide at what level of output to produce to minimise loss.

The *MR* = *MC* rule used in the profit-maximisation case applies here as well. At a price of $38, *MR* = *MC* at 7 units per week. At this output level, *ATC* is $47.4, greater than the price of $38. Fourmet is making a per unit loss of $9.4 and a total loss of $52 ($9.4 × 7), which is equal to the shaded area. Fourmet is minimising its loss. You are encouraged to find out the losses at other levels of output, and check that the loss at output of 7 units is indeed the minimum. By calculating the total revenue and total cost at this level of output, we can also obtain the same total loss of $52 per week. However, why would Fourmet continue to produce if it is making a loss?

Note that although the price is not high enough to pay the average total cost, the price is high enough to cover the average variable cost (AVC). Since P > AVC, each unit sold also contributes to paying a portion of the average fixed cost, which is the vertical distance between the *ATC* and *AVC* curves. Even though it is making a loss in this situation, the firm should continue to operate since it can recover part of the fixed cost. If it were to shut down, the contribution that it is currently making to fixed cost would disappear and it would suffer a bigger loss of $100 per week, which is the fixed cost.

Global perspective

Applicable concepts: fixed costs, variable costs, profit, loss, entry and exit

How much does an Uber driver earn?

Getty Images/AFP/ROSLAN RAHMAN

Uber is the largest ride-sharing service provider in the US, with a market share of nearly 60 per cent as of 2019. It became a public company in 2019, with a 2020 market value of over US$80 billion, approaching the combined value of two of largest car manufacturers in the US – Ford and General Motors.[1] The idea of sharing the cost of driving was conceived in 2008 when Garrett Camp, one of its co-founders spent a small fortune hiring a private driver for New Year's Eve. While we marvel at his ingenuity and achievement of turning a small idea into a multi-billion-dollar international empire, have you wondered how its Uber drivers fare?

In 2018, there were more than three million Uber drivers globally, including over 900 000 in the US alone.[2] According to the Economic Policy Institute (EPI), the median wage for Uber drivers in the US after expenses and fees is $9.21 an hour, just slightly above the US minimum wage.[3] *The Guardian* reported high turnover rates of Uber drivers; for example, only 4 per cent of them continued driving for Uber after one year.

The situation in Australia for Uber drivers does not fare any better.[4] In a 2018 simulation study entitled 'Subsidising billionaires', Jim Stanford found that the average Uber driver's per hour net wage in the six capital cities is A$14.6, well below the statutory minimum wage of A$18.3 per hour. This earning is net of government tax, petrol, vehicle depreciation, insurance and other operation costs as allowed by the Australian Taxation Office.[5]

What do you think?

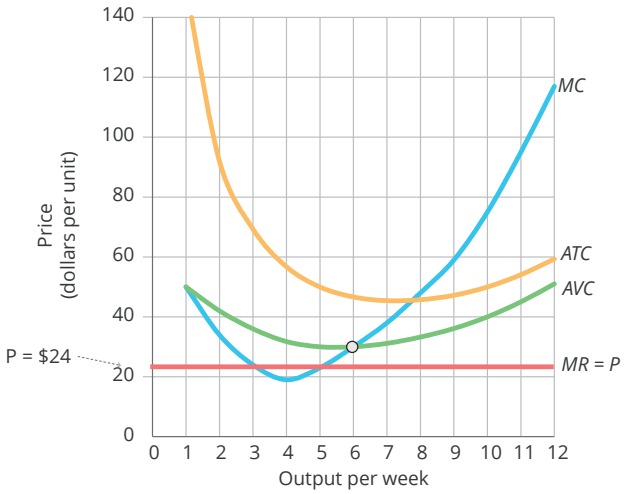
1. An Uber driver operates as a self-owned business. Of all the costs of running an Uber in Australia, which costs are fixed and which are variable?
2. In Australia, based on the simulation study presented above, has an average Uber driver been able to earn enough to cover their total cost of operation? Are they making an economic profit or loss?
3. Considering the characteristics of perfect competition (e.g., the numbers of drivers, ease of entry and exit) would you consider the ride-sharing industry come close to a perfectly competitive market?
4. Do you expect an Uber driver to make a profit in the long run? Why/Why not?

1 A. Reddy, 'Uber is going public at a $75.5 billion valuation. Here's how that stacks up',10 May 2019, https://www.businessinsider.com.au/uber-technologies-initial-public-offering-uber-valuation-comparison-2019-4
2 Ibid.
3 L. Mistel, 'Uber and the labor market', *Business Insider Australia*, 2018, https://www.epi.org/publication/uber-and-the-labor-market-uber-drivers-compensation-wages-and-the-scale-of-uber-and-the-gig-economy
4 M. Sainato, '"I made $3.75 an hour": Lyft and Uber drivers push to unionize for better pay', *The Guardian*, 22 March 2019, https://www.theguardian.com/us-news/2019/mar/22/uber-lyft-ipo-drivers-unionize-low-pay-expenses
5 J. Stanford 'Subsiding billionaires', Centre for Future Work at the Australia Institute, March 2018.

A competitive firm shutting down

What happens if the market price drops below the *AVC* curve, as shown in **Exhibit 7.7**? For example, if the price is, say, $24 per unit, should Fourmet produce some level of output? The answer is no. The best course of action is for the firm to shut down. If the price is below the minimum point on the *AVC* curve, the revenue from each unit produced (which is the same as the market price in perfect competition) cannot cover the variable cost per unit (i.e., *AVC*), let alone make a contribution to fixed costs. To operate under these conditions would involve total losses in excess of those incurred if the firm were to shut down. If the firm shuts down, the maximum loss it can incur is equal to the total fixed cost. Rather than losing more than the total fixed cost, if the market price is below the minimum *AVC*, the firm would be better off shutting down and producing zero output. Such a price is known as *shutdown price* in the short run.

Exhibit 7.7 **The short-run shutdown point**



CAUSATION CHAIN

Price is below minimum average variable cost → Firm will shut down

If the price falls below the minimum point on the AVC curve, the firm shuts down. The reason is that operating losses are now greater than the total fixed cost. In this exhibit, at the price of $24 per unit, which is below the AVC curve at any level of output, the firm shuts down.

In summary

- As well as being able to determine the profit maximisation output by applying the *MR* = *MC* rule, a loss-making firm can also determine its loss minimisation output by applying this rule.
- If the price is below the minimum point on the average variable cost curve, the firm shuts down to minimise its loss in the short run.

5 Short-run supply curves under perfect competition

In **Chapter 3**, we drew a firm's supply curve by stating that the higher the price the more output supplied. Now we can construct the competitive firm's supply curve using more precision and then derive the market supply for the frozen meals industry.

You're the economist

Whose happiness increases during happy hour?

At bars, cafes and coffee shops across the country there are regular happy hours when drinks and sometimes food can be bought at lower than normal prices. These happy hours typically occur at times when business is slow – often during late morning or late afternoon when customers are between meals. And happy hours are spreading beyond the realm of the hospitality industry with some domestic airlines now calling their time-specific discounts 'happy hour' deals. There are many ways in which the happy hour discount can be delivered, with one variant involving 'tossing the boss' where the toss of a coin determines whether full price is paid or the drinks are free. And, just as economists would predict, some patrons – especially those purchasing alcoholic drinks – endeavour to extend the happy hour by buying up big just as happy hour is about to end. Assume the average fixed cost for a cup of coffee, including apportioned rent for the premises, insurance and interest forgone, is $2.20. The AVC for each cup of coffee including coffee, water and labour is $1.80. Would this coffee shop owner be happier charging $2 during happy hour or would they be better off taking a break and shutting down, if they could, during happy hour?

Alamy Stock Photo/UrbanImages

The competitive firm's short-run supply curve

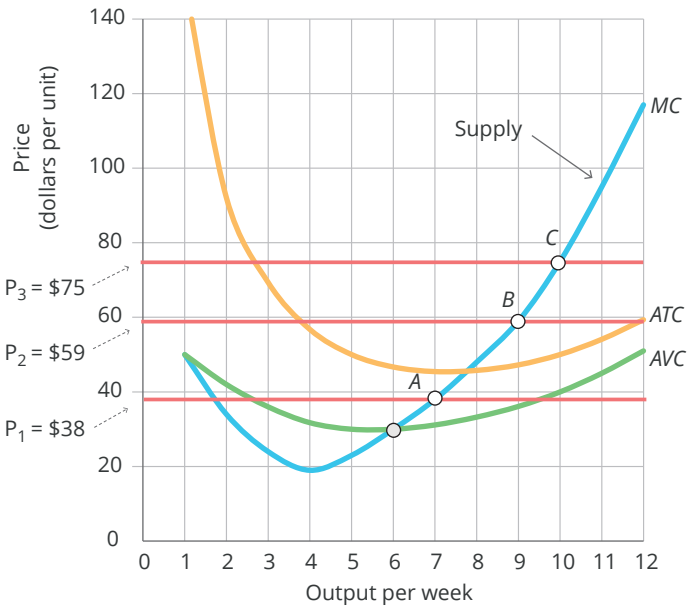
Exhibit 7.8 reproduces the cost curves from our Fourmet example.

Competitive firm's short-run supply curve The firm's marginal cost curve above the minimum point on its average variable cost curve.

Recall that for a competitive firm, its marginal revenue (MR) equals the market price (P). Therefore, in perfect competition the condition of profit maximisation can be restated as: P = MC. Suppose the market price for gourmet packs is $38 per unit ($P_1$). Profit maximisation (to be more correct in this case, loss minimisation) output occurs at 7 units per week where P = MC (point *A*).

If the price rises to $49, Fourmet earns an economic profit and maximise its profit by producing 9 units per week (point *B*). At a price of $75, profit maximisation occurs at 10 units of output per week (point *C*). As the market price increases, the firm's supply curve is traced by moving upward along its *MC* curve. Note that it is not true the firm's supply curve corresponds to the *whole* MC curve. As discussed, at any lower price than the minimum AVC, the firm cuts its loss by shutting down. The minimum point of the AVC curve is therefore the lowest point on the individual firm's short-run supply curve. We can conclude that the **competitive firm's short-run supply curve** is its marginal cost curve above the minimum point on its AVC curve.

Exhibit 7.8 The firm's short-run supply curve



The exhibit shows how the short-run supply curve for Fourmet is derived. When the price is $38, the firm will produce 7 units per week at point *A*. If the price rises to $59, the firm will move upward along its marginal cost curve to point *B* and produce 9 units per week. At $75, the firm produces 10 units per week. Thus, the firm's short-run supply curve is the marginal cost curve *above* its AVC curve.

The perfectly competitive industry's short-run market supply curve

Perfectly competitive industry's short-run market supply curve The supply curve derived from the horizontal summation of the short-run supply curves of all firms in the industry.

The **perfectly competitive industry's short-run market supply curve** is simply the horizontal sum of the short-run supply curves of all the firms in an industry.

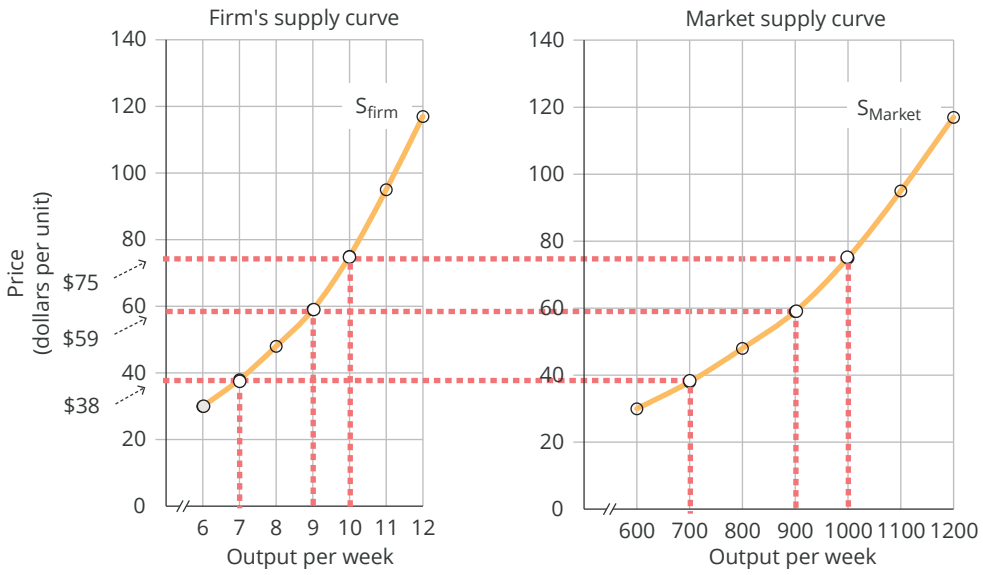
Although in perfect competition there are many firms of different sizes and utilising different technologies, we will suppose for simplicity that the industry has 100 firms of the same size and using same technology as Fourmet. Therefore, these firms have the same *MC* curve, and hence the same supply curve. **Exhibit 7.9** illustrates the construction of the market supply curve (S_{Market}). At a price of $38, the quantity supplied by Fourmet would be 7 units per week. Since there are 100 identical firms, the quantity supplied by the industry will be 700 units per week. Following this procedure for all prices, we generate the industry's short-run market supply curve.

Note that the market supply curve derived in **Exhibit 7.9** is based on the assumption that input prices remain unchanged as output expands. Shortly you will learn how changes in input prices affect derivation of the supply curve in the long run. First, however, we need to examine short-run and long-run equilibrium for a competitive firm.

In summary

- The **competitive firm's short-run supply curve** is the firm's marginal cost curve above its minimum point.
- The industry's short-run market supply curve is the horizontal summation of the short-run supply curves of all firms in the industry.

Exhibit 7.9 **Deriving the industry short-run market supply curve**



The short-run market supply curve for an industry is derived by the horizontal summation of quantities supplied at each price for all firms in the industry. In this exhibit, we assume there are 100 identical firms in the frozen meals industry. At $38, Fourmet supplies 7 units of output, and the quantity supplied by the industry is therefore 700 units. Other points forming the industry short-run market supply curve are obtained similarly.

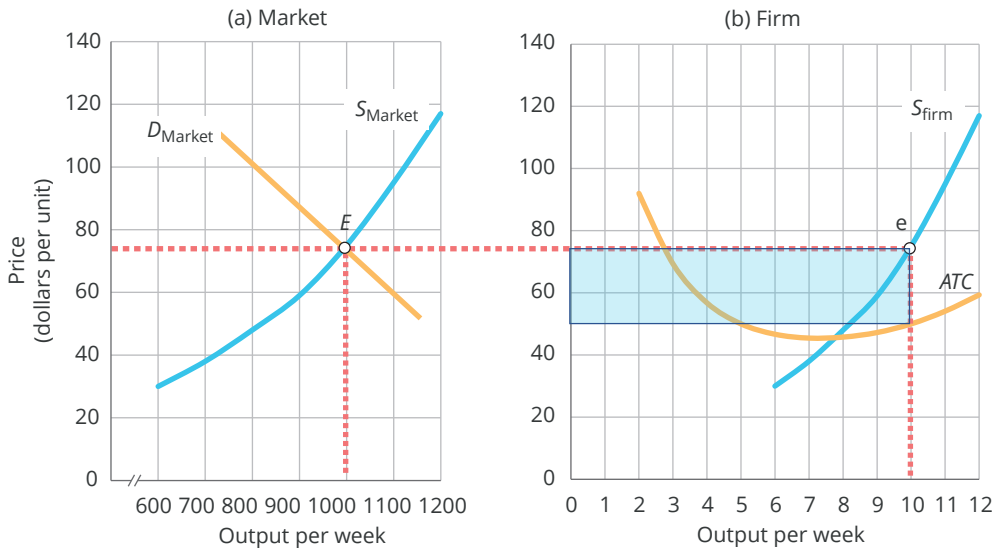
6 Short-run and long-run equilibrium for a competitive firm

Equilibrium is a situation in which there is no tendency to change. In the context of firms and markets this means a situation where the level of output and its price remain stable. For a market when demand equals supply, other things being unchanged, there is no tendency for the market output and price to change. The market is said to be in equilibrium. When a firm maximises its profit, it has no reason to alter its output decision. Therefore profit maximisation is the condition for a firm to be in equilibrium. We look first at short-run equilibrium in perfect competition, which is illustrated in **Exhibit 7.10**.

Short-run equilibrium for a competitive firm

Exhibit 7.10 shows (a) the market demand and supply in the competitive frozen meals industry and (b) Fourmet's supply and *ATC* curves. As shown in diagram (a), the interaction of the market demand and market supply determines the equilibrium market price at $75 per unit and equilibrium output of 1000 units per week (point E). At this price Fourmet maximises its profit by producing 10 units per week (point e). The market is in equilibrium at point E where demand equals supply, whereas the firm is in equilibrium at point e, maximising its profit. This state of short-run equilibrium will remain until some factor changes and causes a new equilibrium condition in the industry. As you will see shortly, given that this short-run equilibrium is one in which firms are earning an economic profit, some change will occur, and this short-run profit cannot be sustained in the long run.

Exhibit 7.10 **Short-run equilibrium in perfect competition**



Part (a) shows that short-run market equilibrium occurs at point E. The intersection of the market supply and demand curves determines the price of $75 facing the firm shown in part (b). Given this market price, the firm establishes its profit-maximising output at 10 units per week and earns an economic profit shown by the shaded area.

Long-run equilibrium for a competitive firm

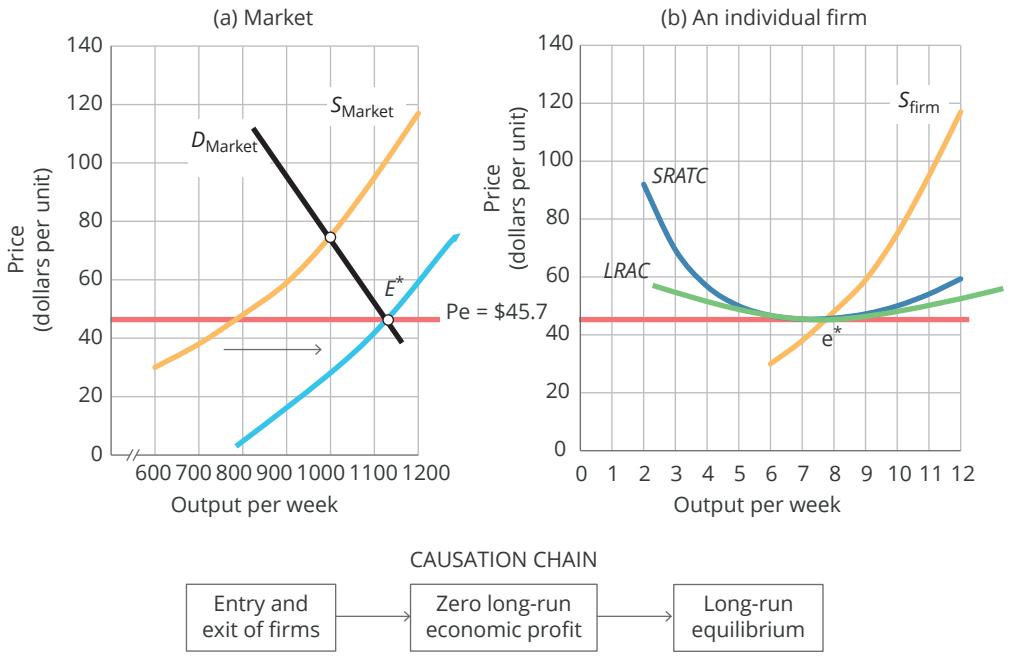
Economic agents respond to incentives. If firms in a competitive market are making economic profits as depicted in part (b) of **Exhibit 7.10**, they can only enjoy those profits in the *short run*. This is because in the long run new firms enticed by profits opportunities will enter the industry to compete for the profits. Entry of new firms will increase the market supply causing the market price to fall until economic profits are reduced to zero. When firm are earning zero economic profits, there will be no incentive for more new firms to enter the market and the industry will be in long-run equilibrium.

Exhibit 7.11 illustrates the long-run adjustment in the frozen meals industry. In the long run in response to the profit signal, new frozen meals manufacturers enter the industry, shifting the market supply curve to the right. The market reaches a new equilibrium point E* as shown in diagram (a). At this new equilibrium point, the market price is driven down to the minimum point of the short-run average total cost (SRATC) curve at $45.7. Diagram (b) shows that at this price, following the MR = MC rule, Fourmet produces an equilibrium output of 7.8 units per week (point e*). At this output level, Fourmet earns a zero economic profit. Other things being unchanged, there will be no more market forces to alter this equilibrium, and the industry is now in the *long-run* equilibrium.

The long-run adjustment is similar if there are economic losses in an industry. Therefore the discussion will be in brief and you are encouraged to model the adjustment process. In the long run some existing firms leave the industry, causing the short-run market supply curve to shift to the left and the market price to rise, reducing losses. This exit continues until economic losses are eliminated in the long run.

In part (b) of **Exhibit 7.11**, note that the long-run average cost (*LRAC*) curve is added to show an important point about productive efficiency of the competitive market. You can see that in the long-run price equals the minimum point of the *LRAC* curve as well as the minimum point of the short-run average cost curve. Given

Exhibit 7.11 **Long-run perfectly competitive equilibrium**



In the long run, entry of new firms will raise market supply, driving down market price until it is equal to the minimum ATC. Long-run market equilibrium therefore occurs at point *E**. At the minimum point of the ATC curve (point e*) the firm earns zero economic profit. Firm's long-run equilibrium therefore occurs at point e*.

the U-shaped *LRAC* curve, the firm is producing with the optimal (least cost) factory size. To see that this is optimal, consider firms with a larger or smaller factory size. In either case these firms would be at a point on the *LRAC* curve that involved higher average costs than at the minimum point. This could not be a point of long-run equilibrium, because these firms would seek to change the size of their factories in order to decrease their average costs and maximise their profits. Only when all firms in the industry are operating at the minimum point of the *LRAC* curve will long-run equilibrium be achieved.

These conditions for long-run competitive equilibrium can also be expressed as:

$$p = MR = SRMC = SRATC = LRAC$$

As long as the market conditions remain unchanged (e.g., consumer taste, technology, input prices and prices of related goods), there is no reason for a competitive firm to change its output level, factory size or any other aspect of its operation. Under long-run equilibrium conditions, there are neither positive economic profits to attract new firms to enter the industry nor negative economic profits to force existing firms to leave. In long-run, the adjustment process of firms moving into or out of the industry is complete. The price that prevails in a perfectly competitive market in long-run equilibrium is the lowest possible price that can be achieved. The market is operating efficiently, producing output at the lowest possible ATC and society's wellbeing is maximised.

The Invisible Hand

The market process we have just discussed is what Adam Smith refers to when he describes the market as the Invisible Hand. Through the interaction of buyers and sellers, as presented in **Chapter 3**, the market determines the equilibrium price and output, and the competitive equilibrium market outcome brings about maximisation

of social surplus. In this chapter we show that in the long run in response to price and profit signals firms will channel resources into or away from an industry with the important result that the output is produced at the lowest possible cost (the minimum LRAC), and the consumers pay the lowest possible price. What is even more amazing about the competitive market is that this efficient use of resource is achieved without the planning or coordination by the central government, and despite (because?) individuals' motivation is to advance their self-interest in the free market. We can conclude the competitive market forces are the Invisible Hand that enables resources to be utilised in the most efficient manner for the betterment of society.

Here we should also note that our long-run analysis should not be interpreted as meaning that the industry will stay at this long-run equilibrium outcome forever. This is because market conditions change continually, long-run equilibrium may only be achieved from time to time. At other times, following, say, a change in technology, the market will adjust and be moving towards another long-run equilibrium. Witness the advancement in semi-conductor technology and the continual fall in prices and rise in quality of computing and communication devices in the past decades. We won't say computers and mobile devices belong to the perfectly competitive industry. However, the fundamental characteristics of the consumer and producer are the same in this industry. Consumers seek convenience in conducting their life and businesses aim to be more productive in seeking profits. It is the same profit motives and market forces of competition that drives the industry to become more efficient and innovative in producing output that meet consumers' demand – the invisible hand.

Now as a practice, can you model the impacts on the frozen meals industry if consumers' demand for frozen meals falls after the pandemic? Next, we will present the industry's long-run supply curve and discuss how costs of production can alter in different ways when an industry expands in the long run.

Corporate social responsibility

Economics and ethics

Corporate social responsibility (CSR) involves the idea that firms have a responsibility to behave in ways that enhance environmental and social wellbeing. CSR is a topic that is examined in most business management courses taught in universities. While some firms might give particular emphasis to CSR in the belief that it enhances their reputation and thus the marketability of the goods and services they produce, others may simply believe that it is the right thing to do. It has occurred to economists and others that such behaviour might be conditional upon firms earning enough profit to fund such responsible behaviour.

So, for example, Bert van de Ven and Ronald Jeurissen in a paper in *Business Ethics Quarterly* argue that competitive markets can limit what firms can 'afford' when it comes to CSR. In this context, they ask if it is morally relevant that some firms face much more intense competition than others?[1] In the Australian context, they could well have asked whether orchardists who operate in a highly competitive market, and who have been found to have been underpaying workers, would be less likely to display CSR than the highly profitable big banks that, among other financial institutions, were the subject of a Royal Commission in 2018.

Given that perfect competition is the market structure in which competition is fiercest, do you think competitive firms show less CSR than firms able to earn higher profits? Do you think that the incidence of under-paying workers, cutting back on health and safety requirements or ripping off customers is more common in highly competitive industries?

[1] B. van de Ven and R. Jeurissen, 'Competing responsibly', *Business Ethics Quarterly*, Vol. 15, 2 April 2005, pp. 299–317, http://www.uky.edu/~jtgros2/responsibility.pdf, accessed 12 September 2017.

In summary

- **Short-run equilibrium** for a competitive firm occurs when the firm maximises profit or minimises losses by producing at output level where MR = MC.
- **Long-run equilibrium** occurs when the firm is producing at the lowest LRAC, earning zero economic profit.

7 Three types of long-run supply curve under perfect competition

In the long run *all* inputs are variable. Existing firms in an industry can react to profit opportunities by building larger plants, buying more land and equipment, to ensure that it is operating at the optimal scale. Profits also attract new firms to an industry, while losses cause some existing firms to leave the industry. This entry and exit of firms represent expansion and contraction of the industry, which can influence the shape of the long-run supply curve.

Competitive industry's long-run supply curve The curve that shows the quantities supplied by the industry at different equilibrium prices after firms complete their entry and exit.

The **competitive industry's long-run supply curve** shows the quantities supplied by the industry at different equilibrium prices after firms complete their entry and exit. In the following discussion, we focus on expansion of an industry. The argument is equivalent for a contracting industry. The shape of the long-run supply curve depends on the response of input prices as new firms enter the industry. The following sections discuss three possible cases. The key to understanding these three cases is to realise that as industry expands with entry of firms, the consequent increase in demand for inputs may cause the prices of these inputs to rise, fall or remain the same. We will consider the constant-cost industry first.

Constant-cost industry

Constant-cost industry An industry in which the expansion of industry output by the entry of new firms has no effect on the firms' cost curves.

In a **constant-cost industry**, input prices remain constant as new firms enter the industry. This implies that in a constant-cost industry the expansion of the industry has no effect on the firm's cost curves. For example, expansion of a grazing industry, be it sheep or cattle, may occur with no change in firms' costs. **Exhibit 7.12** illustrates the long-run supply curve of a constant-cost industry.

Part (a) of **Exhibit 7.12** shows the initial market equilibrium point E_1 with short-run market supply curve S_1 and market demand curve D_1. Now assume the market demand curve increases from D_1 to D_2 due to, for example, a change in consumers' taste in favour of the product. As a result, the short-run market equilibrium moves to E_2. Correspondingly, the equilibrium price rises from \$60 to \$80 and industry output increases from 60 000 to 70 000 units.

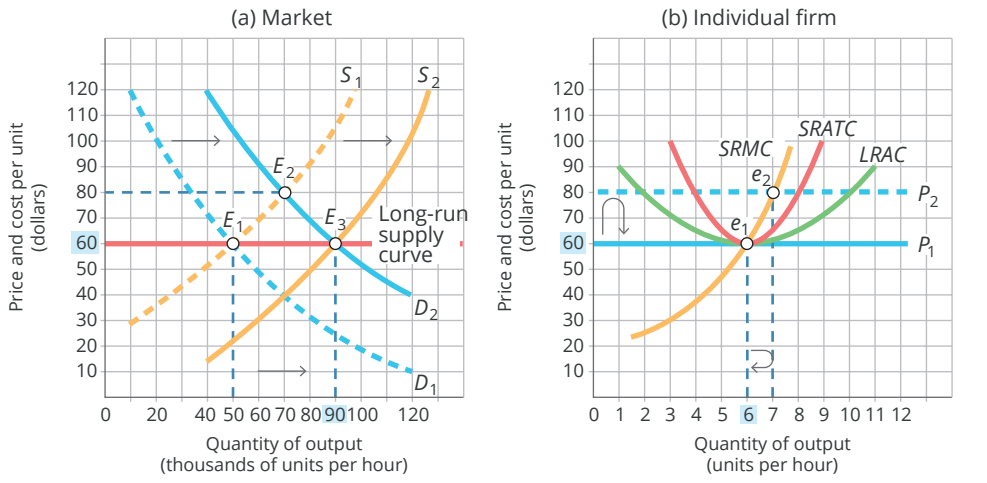
The short-run response of the typical firm in the industry to the price change is presented in part (b). As shown in part (b) of **Exhibit 7.12**, the firm takes the increase in price and adjusts its output from 6 to 7 units per month. At the higher price and output, the firm changes from earning a zero economic profit to making a positive economic profit because the new price is above the *SRATC* curve. All the other firms in the industry make the same adjustment by moving upward along their *SRMC* curves and make short-run profits.

In perfect competition, in the long run new firms are free to enter the industry in response to a profit opportunity, and they will do so. The addition of new firms shifts the short-run market supply curve rightward from S_1 to S_2 driving down the market price. Firms will continue to enter the industry until profit is eliminated when the market price is reduced to the minimum SRATC. This occurs at equilibrium point E_3, where market demand D_2 intersects market supply S_2. Thus, the entry of new firms has restored the initial equilibrium price of \$60. The firm responds by moving back downward along its *SRMC* curve until it once again produces 6 units and earns zero economic profit. Although the typical firm produces its original output of 6 units, the industry has more firms now and the industry output increases to 80 000 units per month.

As shown in the exhibit, the path of these changes in industry long-run equilibrium points traces a horizontal line, which is the industry's long-run supply curve.

In presenting the above adjustment process it is assumed that the input prices are unchanged and as a result firms' short-run cost curves are also unchanged. What is the likely cause of this outcome? If an industry's use of particular inputs is small in relation to the overall availability of these inputs in the economy, it is likely that expansion of the industry will have a negligible effect on the prices of these inputs. Thus, a situation of constant cost industry is likely to prevail.

Exhibit 7.12 Long-run supply in a constant-cost industry



CAUSATION CHAIN

Increase in demand sets a higher equilibrium price → Entry of new firms increases supply → Initial equilibrium price is restored → Horizontal long-run supply curve

Part (a) shows an industry in equilibrium at point E_1, producing 60 000 units per month and selling units at \$60 per unit. In part (b) the firm is in long run equilibrium, producing six units per month (e_1) and earning zero economic profit. Then market demand increases from D_1 to D_2 because of (say) a change in consumers' taste, and the equilibrium price rises to \$80. Industry output rises to 70 000 units per month (E_2) and the individual firm increases output to seven units per month (e_2). Firms are now earning an economic profit, which attracts new firms into the industry. In the long run, the entry of these firms causes the short-run supply curve to shift rightward from S_1 to S_2, and the price is reduced to \$60 and a new long run equilibrium point, E_3, is established in the industry. At E_3, industry output rises to 80 000 units per month and the firm's output returns to six units per month (e_1). As illustrated the expansion of the industry has no effect on input costs, the firm's short run cost curves in part (b) are unchanged and the industry's long run supply curve is horizontal.

Applicable concept: long-run perfectly competitive equilibrium maximises efficiency

Analyse the issue

How do consumers know that prices are not excessive?

It is sometimes thought that consumers cannot tell whether they are being ripped off by a firm because they are not privy to information about the firm's costs. This line of thought has it that the consumer would need to know details of the firm's costs so that they could be subtracted from the price charged to see whether the profit being made is excessive. Nothing could be further from the truth.

To see why, consider our discussion of long-run equilibrium under perfect competition. We have seen that if excess profits occur in a competitive industry, new entrants will appear and compete these

Getty Images Plus/iStock/jarino47

profits away. This means that if consumers observe strong competition in an industry they should deduce that excess profits are not being made and that prices are thus not excessive.

In practice it is usually not difficult to tell whether competition is strong. Just ask yourself about the characteristics of the industry in which the firm in question operates. If there are many firms in the industry, if the products sold are similar, and if it is easy for new firms to start up in the industry, then it is more than likely that the industry is highly competitive and that you are not being ripped off.

As former Chairperson of the Australian government's Productivity Commission, Gary Banks, has put it:

> the nice thing about competitive markets is that their effectiveness in disciplining prices does not depend on knowing how prices are *calculated;* only how they *compare* among alternative suppliers.

He continues:

> [n]o doubt, in the very short term, and especially for low cost items, consumers may not bother looking around much. But, in time, they could be expected to find out if their preferred outlet has been charging more than elsewhere (even if they weren't sure why) and to act accordingly. Most suppliers would anticipate and wish to avert such loss of business. The risk to a firm's reputation of consumers actually concluding that they had been *had* [ripped off] would only reinforce this.[1]

We can say that the competitive market is a most effective regulator of anti-competitive behaviour in the economy. Any government regulation required is essentially to ensure competitive market conditions.

What do you think?

1. What does the fact that competitive firms are sometimes called price-takers tell us about the ability of these firms to charge excessive prices?
2. Are you more likely to be ripped off by the local provider of lawn care services or by one of Australia's big four banks?
3. At times when prices for certain fresh seafood, such as prawns or filleted fish, are very high, some consumers argue that the fishers or the fish shop are ripping them off. What is wrong with this analysis?

[1] G. Banks, 'Competition is the best price regulator', Presentation to the Committee for Economic Development of Australia, Perth, 21 November 2000. (Co-authored with Lisa Gropp.) Reproduced in Gary Banks, *An Economy-wide View: Speeches on Structural Reform*, Productivity Commission, Melbourne, 2010, pp. 151–62.

Increasing-cost industry

Increasing-cost industry
An industry in which the expansion of industry output by the entry of new firms increases the firm's input costs.

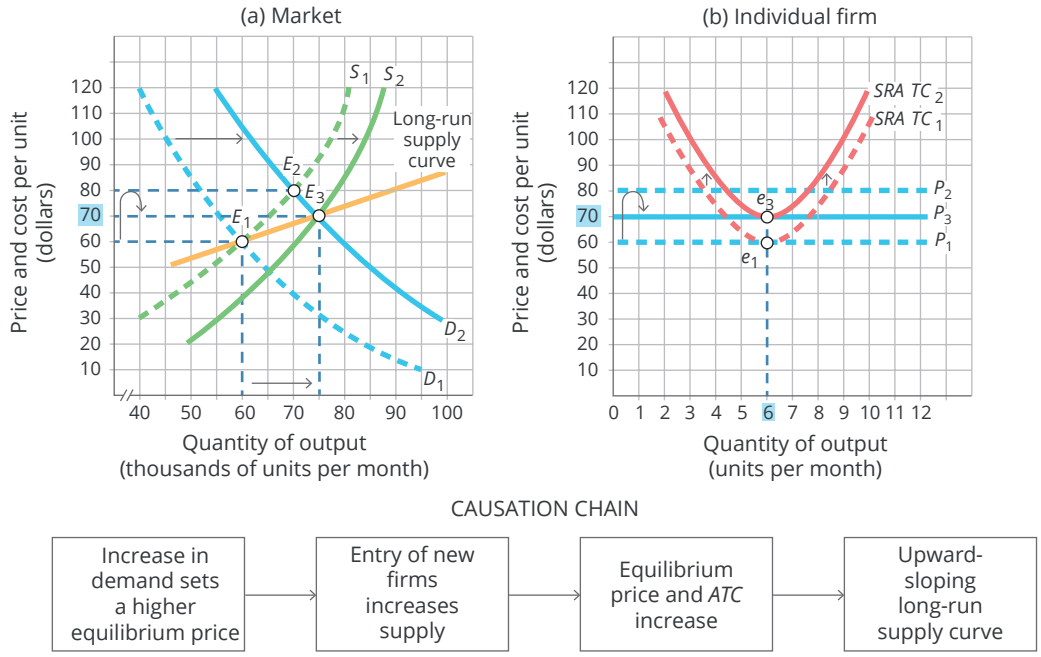
In Australia before the coronavirus pandemic hit its economy, the housing market experienced strong growth. The construction costs rose on average by 3.7 per cent per annum well above the annual inflation rate of 1.7 per cent. The increase in construction costs was due mainly to the rising labour cost as employment in the construction industry has risen over 14 per cent in the period between 2014 and 2019.[6] The construction industry is an example of an **increasing cost industry**. An increasing-cost industry is an industry in which the expansion of output by the entry of new firms increases the firm's input costs.

Exhibit 7.13 shows what happens in an increasing-cost industry when an increase in demand causes industry output to expand. In part (a), the industry is initially in long-run equilibrium at point E_1. As in the previous case, the demand curve shifts rightward from D_1 to D_2, establishing a new short-run equilibrium at E_2. This movement upward along short-run industry supply curve S_1 raises the price in the short run from $60 to $80, resulting in profit for the typical firm. Once again, new firms enter the industry and the short-run market supply curve shifts

[6] H. Dowling, 'Construction costs increasing, outpacing inflation', *Mortgage Business*, 7 November 2019, https://www.mortgagebusiness.com.au/breaking-news/13976-construction-costs-increasing-outpacing-inflation

rightward from S_1 to S_2. Part (b) of **Exhibit 7.13** shows that the response of the firm's *SRATC* to the industry's expansion differs from that in the constant-cost industry case. In an increasing-cost industry, due to increase in input costs the firm's *SRATC* curve shifts upward from $SRATC_1$ to $SRATC_2$, corresponding to the new long-run equilibrium at point E_3. At this final equilibrium point, the price is higher at \$70 than the initial price of \$60 and the industry output rises to 75 000 units. By connecting the long-run equilibrium points E_1 *and* E_3, we obtain the industry's long-run supply curve. For an increasing cost industry, its long-run supply curve slopes up to the right.

Exhibit 7.13 **Long-run supply in an increasing-cost industry**



This pair of graphs derives the long-run supply curve based on the assumption that input prices rise as industry output expands. Part (a) shows that an increase in demand from D_1 to D_2 causes the price to increase in the short run from \$60 to \$80; the market equilibrium shifts from E_1 to E_2. The typical firm represented in part (b) earns an economic profit, enticing new firms to enter the industry, causing an increase in industry supply from S_1 to S_2 and the market price to fall. As industry output expands, input prices rise, pushing up the firm's short-run average total cost curve from $SRATC_1$ to $SRATC_2$. As a result, a new long-run market equilibrium point E_3 as shown in part (a) is established. At E_3 the equilibrium price is \$70, which is above the initial equilibrium price, and the industry output is 75 000 units and firm's output returns to six units per month (e_3). The long-run supply curve for an increasing-cost industry is upward-sloping. Note the firm's SRMC curves are omitted for clarity in part (b).

Decreasing-cost industry

Decreasing-cost industry An industry in which the expansion of industry output by the entry of new firms decreases the firms' input costs.

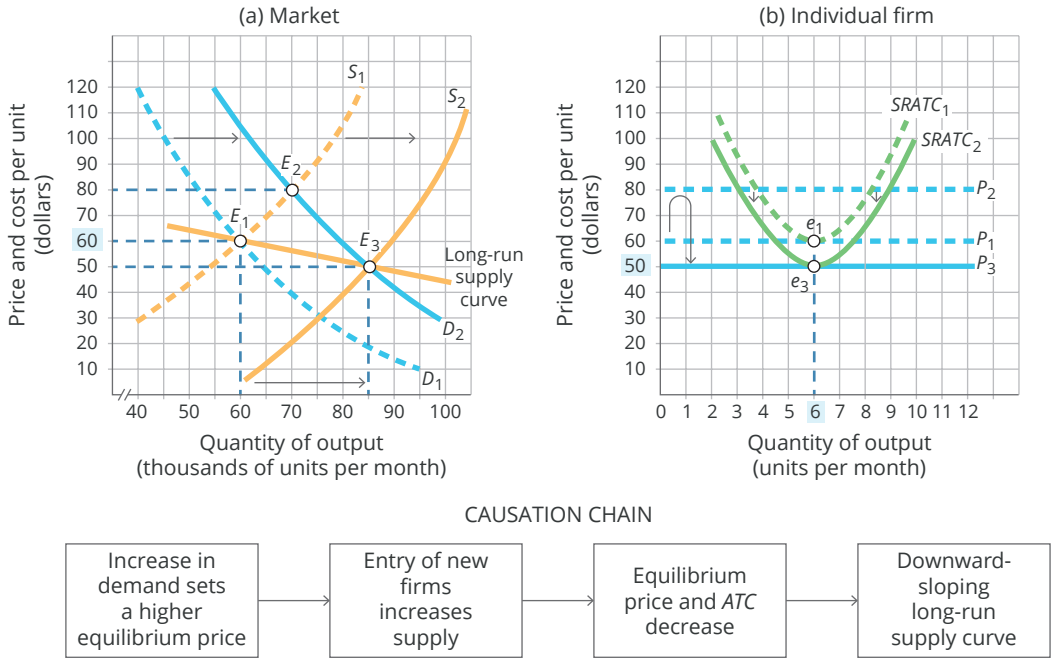
A **decreasing-cost industry** is an industry in which the expansion of industry output by the entry of new firms decreases the firm's input costs. For example, since the 90s as China joined the World Trade Organization and opened up its markets, southern China has become a manufacturing hub, exporting many labour-intensive goods to all parts of the world. As more joint ventures set up their factories there, the local governments spent a lot of money on infrastructure, hospitals and schools to support the booming local economy. The rise in economic activities also attracted local and overseas corporations to provide financial and other commercial services

to this region. These advantages greatly reduce the costs of running business for these manufacturers[7]. Since these advantages are external to the individual firm, they are known as *external economies of scale*. **Exhibit 7.14** illustrates the long-run adjustment process of an increase in demand for Chinese exports and the subsequent expansion of the manufacturing industry in southern China. As the industry expands, due to decrease in various input costs (e.g., transportation and financial costs) the SRATC curve of the typical firm falls. And the long-run supply curve of a decreasing-cost industry slopes downward to the right.

Note that in each of the three possible scenarios described above, the long-run industry supply curve is drawn by connecting the two long-run equilibrium points of E_1 and E_3. Equilibrium point E_2 represents short-run equilibrium before the entry of new firms reducing economic profits to zero in the long run.

Finally, it is important to understand that each of the three models presented above can be observed in the real world. Some industries have constant costs when they expand, while others may have decreasing costs or increasing costs. Only direct observation and economic analysis of the industry can tell us which type of industry it is.

Exhibit 7.14 **Long-run supply in a decreasing-cost industry**



The increase in industry demand shown in part (a) causes the price to rise to $80 in the short run ($E_2$). Temporarily, the typical firm illustrated in part (b) earns an economic profit. Higher profits attract new firms and market supply increases. As the industry expands, the average total cost curve for the firm shifts lower. The industry re-establishes long-run equilibrium at the lower price of $50, producing 85000 units (E_3) and the firm returns to six units per month (e_3). Connecting E_1 and E_3 gives us the long run supply curve. The long-run supply curve for a decreasing-cost industry is downward-sloping. Note the firm's SRMC curves are omitted for clarity in part (b).

[7] You may ask if the expansion of the manufacturing industries in southern China would also push costs of labour up. During the first decade of this phenomenal growth in manufacturing, there was an influx of migrant workers to this part of China from other provinces preventing wages from rising rapidly. However, more recently as the cheap labour supply dried up and the cost of living rose, the Chinese manufacturing industries have lost much of its competitive advantage gained from cheap labour.

Zero economic profits – why bother to produce?

Our discussion of long-run equilibrium shows that in the long run competitive firms will earn *zero* economic profit. You may wonder: If a firm earns zero profit, why would they bother to continue to produce at all?

How would you interpret a zero economic profit? It's not as bad as it sounds. Recall in **Chapter 6** we distinguish between economists' and accountants' definitions of profit. In economics, profit is defined as the difference between total revenue and total costs. Total costs include both explicit and implicit costs. The important implication is that zero (economic) profit means that the firm has earned sufficient revenue to cover all costs of production, including implicit costs. This means that the entrepreneur has earned enough to compensate for their own inputs (e.g., own labour and capital) and for risk-taking. Economists call this condition **normal profit**, which is the minimum returns to own inputs (including risk taking) necessary to keep a firm in operation. To reiterate, zero economic profit signifies that there is just enough total revenue to pay the owners for all explicit and implicit costs. To state it differently, there is no benefit from reallocating resources to another use.

Normal profit The minimum return to own inputs (including risk taking) necessary to keep a firm in operation. A firm that earns normal profit earns total revenue equal to its total opportunity cost (total explicit and implicit costs).

Applicable concept: the characteristics of perfect competition

Global perspective

Collaborative consumption facilitated by the internet increases competition in markets

The rise of the internet as a global means of communication has led many commentators to argue that the ability of internet users to connect with businesses and consumers anywhere in the world as easily as they might communicate with their next-door neighbour means that the potential for international competition between firms has increased enormously. In other words, the internet is increasing global competition and promoting the efficiency gains that are an acknowledged outcome of competitive markets. But more than that, the internet is facilitating what has become known as collaborative consumption, which vastly increases the number of suppliers in some markets. This increase in suppliers gives even further impetus to competition.

Way back in 1997, when the internet was in its infancy, the former deputy chairperson of the ACCC, Allan Asher, drew the following implications for global competition from the internet revolution.

> [I]t seems almost certain that on-line markets and other forms of electronic commerce will expand rapidly to the point where a truly global retail marketplace will emerge. Already some industries are feeling the effects of these new technologies, especially those which are essentially information and booking services ...
>
> [This] global electronic market will present both tremendous opportunities as well as some very real challenges for regulators, industry and consumers ...
>
> Electronic commerce has the potential to deliver significant gains to consumers in terms of price, quality and service through increased competition. This is likely to happen for two interrelated reasons – lower barriers to entry and increased numbers of suppliers competing in product markets ...
>
> Since the internet allows newer and smaller players to promote and sell products in direct competition with larger players, it will increase the number of competitors in the market. Consumers can now tap into a global market and are not bound to a restricted number of physically nearby suppliers – improved choice, price and quality should result.[1]

Asher's prediction, that the internet greatly reduces the cost of information connecting suppliers and potential customers and increases the number of competitors in the market, has become even more prescient with the recent explosion of sites promoting collaborative consumption. These sites, such as Uber, which allows almost anybody with a modern car to offer a taxi service, and Airbnb, which enables householders to rent out a room to travellers, have vastly

increased the number of suppliers in many markets. This has increased competition markedly in these markets. So much so that taxi operators who have paid governments large sums for restricted licences have staged protests against this competition. Not only do these cabbies face the problem of declining revenue from fares but the license owners are also facing significant losses from the fall in market value of their taxi licenses.

Similarly, Airbnb's entry into the hospitality industry has hotel operators shaking their heads while homeowners complain that their neighbourhoods have become hotel precincts full of strangers.

Getty Images/John van Hasselt – Corbis

What do you think?

1 How will the internet and collaborative consumption change markets in such a way that they increasingly conform to the three conditions of the perfectly competitive market?
2 Find out the changes to taxi license price since the legalisation of ride-sharing service (e.g., Uber and Lyft) in the country (or city) where you live. Are these observed changes in prices of taxi licenses what the perfect competition model in the long run would predict?

[1] Allan Asher, *International perspective: access to justice for consumers in the global electronic marketplace*, ACCC, 17 March 1997.

In summary

- A **constant-cost industry** is an industry in which expansion of total output involves neither an increase nor a decrease in each firm's average total cost.
- A **decreasing-cost industry** is an industry in which expansion of total output involves a decrease in each firm's average total cost.
- An **increasing-cost industry** is an industry in which expansion of total output involves an increase in each firm's average total cost.

Key concepts

Market structures
Perfect competition
Competition policy
Price-taker
Marginal revenue
Profit maximisation in perfect competition
Competitive firm's short-run supply curve
Perfectly competitive industry's short-run market supply curve
Equilibrium
Perfectly competitive industry's long-run supply curve
Constant-cost industry
Increasing cost industry
Decreasing-cost industry
Normal profit

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Distinguish the four types of market structures

- The four market structures are perfect competition, monopolistic competition, oligopoly and monopoly. The defining characteristics of a **market structure** include the number of firms in that market, the degree to which the products they sell are similar and the ease of entry into, and exit from, the market.

2. Describe the characteristics of perfect competition

- **Perfect competition** is a highly competitive market structure with many sellers and buyers. Each firm in the industry is very small relative to the market as a whole, all the firms sell a homogeneous product, and firms are free to enter and exit the industry.

3. Analyse how a competitive firm determines its short-run profit maximisation output

- A competitive firm is a **price-taker** facing a perfectly elastic (i.e., horizontal) demand curve. It can sell all it wishes at the prevailing market price, but it will sell nothing above the given market price. This is because so many competitive firms are selling an identical product at the going market price.
- Using the *total revenue–total cost method* the competitive firm selects the level of output that maximises profit. Profit reaches a maximum when the vertical difference between the total revenue and the total cost curves is at a maximum.
- The *marginal analysis method* is a second approach to determining a firm's profit-maximisation output. **Marginal revenue** is the change in total revenue from a one-unit change in output. Marginal revenue for a competitive firm equals the market price. The $MR = MC$ rule states that the firm maximises profit (or minimises loss) by producing the output up to the level where marginal revenue equals marginal cost. If the going market price is above the ATC curve, the competitive firm is making a profit.

4. Determine a competitive firm's short-run loss minimisation output

- If the going market price is below the ATC curve, the competitive firm is making a loss. A loss-making firm can also apply the $MR = MC$ rule to determine its loss minimising output. However, if the price is below the minimum point on the AVC curve, the $MR = MC$ rule does not apply and the firm shuts down to minimise its losses.

5. Derive the firm's supply and market supply under perfect competition in the short run

- The **competitive firm's short-run supply curve** is the firm's marginal cost curve above the minimum point of its AVC curve.
- The competitive industry's short-run market supply curve is the horizontal summation of the short-run supply curves of all firms in the industry.

6. Explain the difference between short-run and long-run equilibrium for a competitive firm

- A firm's **equilibrium** occurs when the firm maximises profit or minimises loss. Economic profit (or loss) exists only in the short run in a competitive market. Entry (or exit) of firms will remove any economic profit (or loss) in the long run.
- *Long-run competitive equilibrium* occurs when the firm earns a zero economic profit by producing where price equals minimum long-run average cost, which also equals the minimum short-run average total cost.

7. Distinguish the three types of long-run market supply curve under perfect competition

- A **constant-cost industry** is an industry in which expansion of output involves neither an increase nor a decrease in each firm's average total cost. This is the case when input prices remain constant. The long-run market supply curve in a constant-cost industry is horizontal.
- An **increasing-cost industry** is an industry in which expansion of output involves an increase in firm's average total cost. As the industry expands, the long-run equilibrium market price rises, resulting in an upward-sloping long-run market supply curve.
- A **decreasing-cost industry** is an industry in which expansion of output involves a decrease in firm's average total cost resulting in a downward-sloping long-run market supply curve. As industry output expands, the long-run equilibrium market price falls.

Study questions and problems

1 List and briefly describe the characteristics of the perfectly competitive market structure.

2 Check out the price of a cappuccino at a number of coffee shops in your city. Are the prices the same or different for the same size cup or mug? Do you think the coffee shop industry is perfectly competitive?

3 Discuss the following statement: 'In the real world there is no industry, which conforms precisely to the economist's model of perfect competition. This means that the model is of little practical value.'

4 Consider the following cost data for a competitive firm in the short run:

Output (*Q*)	Total fixed cost (*TFC*)	Total variable cost (*TVC*)	Marginal cost (*MC*)	Marginal revenue (*MR*)
1	$100	$120	$____	$____
2	100	200	____	____
3	100	290	____	____
4	100	430	____	____
5	100	590	____	____

a Complete the MC column. If the market price is $140, complete also the MR column in the above table.

b How many units of output will the firm produce in order to maximise profit in the short run at the market price of $140?

c Calculate the amount of economic profit or loss it makes in (b).

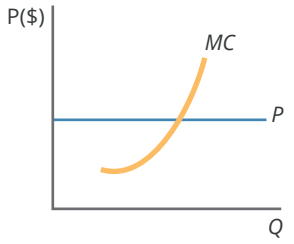
5 Refer to the same firm as in Question (4) and derive its short-run supply curve.

6 Indicate whether you agree or disagree with the following statement: 'When marginal revenue equals marginal cost, total cost equals total revenue and the firm makes zero economic profit'. Explain your answer.

7 A firm's current output is 1000 units per month, with a fixed cost of $2000 per month and a total variable cost of $3600 per month. The market price of this good is $4 per unit. Is the firm making a profit or a loss? Should the firm continue to operate or shut down? In such a situation, what would be your advice?

8 A competitive firm's marginal cost curve is presented in **Exhibit 7.15**. Given the market price of P, mark the firm's equilibrium output in the diagram. Is this firm making a profit or loss? Discuss by adding appropriate curves to the diagram.

Exhibit 7.15 A competitive firm



Answer to 'You're the economist'

Whose happiness increases during happy hour?

As long as price exceeds average variable cost, the coffee shop is better off operating than shutting down. Since $2 per cup more than covers the variable costs, the proprietor will be happier opening and charging happy hour prices than shutting down. The $0.20 remaining after covering variable costs can be put towards the $2.20 of average fixed costs. Were the shop to shut down during happy hour it could make no contribution to these overhead costs. If you said because they can get a happy hour price that exceeds their average variable cost, the happiness of coffee shop owners, just like that of their customers, increases during the happy hour, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 Which of the following goes closest to fitting the economist's model of perfect competition?
- **a** A suburban law firm.
- **b** A city-based firm of management consultants.
- **c** A house cleaning business.
- **d** The only hotel in a small country town.

2 If a competitive firm sells 4 units of output at a market price of $360 per unit, its marginal revenue per unit is:
- **a** $90.
- **b** $360.
- **c** more than $90, but less than $360.
- **d** cannot be determined.

3 Which of the following are constant at all levels of output under perfect competition?
- **a** Price and marginal cost.
- **b** Marginal cost and marginal revenue.
- **c** Price and marginal revenue.
- **d** Marginal revenue and average cost.

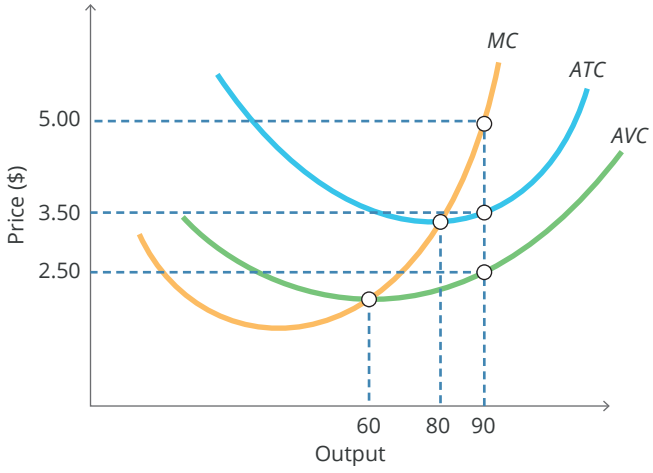
4 A competitive firm's supply curve follows the upward-sloping segment of its marginal cost curve above the:
- **a** average total cost curve.
- **b** average variable cost curve.

c average fixed cost curve.
d average price curve.

5 Why would a firm decide to continue to produce even though it makes losses by doing so?
a A firm would never produce when profit is negative.
b Because the price is high enough to cover the average total cost.
c Because the price is high enough to cover the average variable cost.
d Because the price is high enough to cover the average fixed cost.

6 Refer to **Exhibit 7.16**, and answer the following questions.
a If the price of the firm's product is $5 per unit the firm will produce:
i 60 units.
ii 80 units.
iii 90 units.
iv zero output – it should shut down.
b If the market price is $2.5 per unit, which of the following is true.
i The firm should continue to operate because it is earning an economic profit.
ii The firm should stay in operation for the time being even though it is earning an economic loss.
iii The firm should shut down since it is making an economic loss.
iv The market is in long-run equilibrium.

Exhibit 7.16 **A competitive firm's cost curves**



7 In a perfectly competitive market, all firms, in the long run, earn:
a positive economic profit.
b positive accounting profit.
c zero normal profit.
d zero economic profit.

8 In long-run equilibrium, the competitive firm's price is equal to which of the following?
a marginal revenue.
b minimum short-run average total cost.
c short-run marginal cost.
d all of the above.

9 If input prices rise as new, competitive firms enter an industry we know that this is
a an increasing-cost industry.
b a decreasing-cost industry.
c an industry with an upward sloping long-run supply curve.
d both an increasing-cost industry and an industry with an upward sloping long-run supply curve.

Monopoly

8

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Describe the characteristics of the monopoly market structure.
2 Determine the price and output decisions for a monopolist.
3 Understand how monopolists can engage in a practice known as price discrimination.
4 Evaluate the case against and for monopoly by comparing monopoly and perfect competition.

You probably know that from time to time McDonald's has promotions based on the well-known board game Monopoly. But you may not know that when this game was published by Parker Bros in the 1930s, it took off. This was the time of the Great Depression when a huge proportion of the workforce was unemployed and household incomes were falling year after year. It seems that the opportunity that Monopoly gave for people to earn make-believe money by exercising their economic power over other players was the ideal escape from the harsh world in which they lived during the Great Depression.

Alamy Stock Photo/CoCo Jones

No, this chapter is not about the game of Monopoly. Nevertheless, you will learn that the behaviour of real-world monopolies presented in this chapter has much in common with the behaviour of players in the game. Players try to win the game by gaining more market power and exercising this power as much as possible. They strategically purchase 'sites' on the board in the knowledge that they can gain maximum monopoly power by owning more than one railway station, more than one utility or more than one street in a given location. They can then gain additional power by 'erecting' houses and hotels on some of these sites. As the players' ownership of sites and improvements on them grows, their capacity to earn rents from these sites grows more than proportionately. This feature mirrors the real world where increased ownership of strategic resources and market share can earn monopoly profits for their owners. The winning player is the one who, as a result of the high degree of monopoly power he or she has, is able to bankrupt all other players.

In the previous chapter we studied perfect competition, which may be viewed as the paragon of economic virtue. Why? Under perfect competition there are many sellers, each lacking any power to influence price. Competition enables the consumers to enjoy output at lower prices and costs. Perfect competition and monopoly are polar extremes. The word *monopoly* is derived from the Greek words *mono* and *poleo*, which translate as 'single seller'. A monopoly has the market power to set its price and not worry about competitors. If you have ever been on a cruise ship or holidayed at a remote resort, perhaps you have noticed that there is just one outlet where you can buy alcohol. In this situation you may be paying higher prices for alcohol than would be the case if many sellers competed in this market.

This chapter explains why firms have great difficulty in entering markets dominated by monopolists. Applying the analytical tools introduced in previous chapters, we study how a monopolist determines what price to charge and how much to produce. Then we explore some of the interesting implications for monopoly power of the development of electric cars. The chapter ends with a discussion of the arguments for and against monopolies.

1 The monopoly market structure

Monopoly
A market structure characterised by (1) a single seller, (2) a unique product and (3) extremely difficult or impossible entry into the market.

Recall from the previous chapter that the market structure at the opposite extreme from perfect competition is **monopoly.** Monopoly is a market structure characterised by (1) a single seller, (2) a unique product, and (3) great barriers to entry into the market. Unlike the product of a competitive firm, there are no close substitutes for the monopolist's product. Monopoly, like perfect competition, corresponds only approximately to real-world industries. However, as is the case with perfect competition, it serves as a useful benchmark model. Let's look at a brief description of each monopoly characteristic.

Single seller

In perfect competition, many firms make up the industry. In contrast, a monopoly means that the single firm *is* the industry. The firm provides the total supply of a product in a given market. For example, the campus bookshop, the hospital cafe and the supplier of electric power to your home may be local monopolies. The only service station or hotel in a small, isolated country town and the only hot dog or liquor stand at a football game are also examples of monopolies. Nationally, the National Broadband Network (NBN) is a monopoly of provision of non-wireless high-speed internet network in Australia. In the real-world local monopolies are more common than national or world-market monopolies.

Unique product

A unique product means there are *no close substitutes* for the monopolist's product. Thus, the monopolist faces little or no competition. In reality, however, there are few products that do not have close or fairly close substitutes; for example, students can buy new and used textbooks from sources other than the campus bookshop, including online sites. Gas is a good substitute for electricity for cooking and heating. It is true that there are no good substitutes for electricity for lighting, and it is likely that the producer of electricity in your area has a monopoly in this application. Nonetheless, with domestic rooftop solar power becoming more viable over time as the cost of batteries for storage of power generated during daylight hours becomes cheaper, it is likely that the extent of the monopoly power of suppliers of reticulated electricity will diminish. With the advance of internet, we can source similar or even identical products from other parts of the world, fostering more competition with local or national monopolies.

Barriers to entry

In perfect competition, there are no constraints to prevent new firms from entering an industry. In the case of monopoly, there are extremely high barriers that make it very difficult or impossible for new firms to enter an industry. We will now look at the three major barriers that prevent new sellers from entering a market and competing with a monopolist.

Ownership of a vital resource

Sole control of the entire supply of a strategic input is one way in which a monopolist can prevent a newcomer from entering an industry. A famous historical example is Alcoa's monopoly in the US aluminium market from the late 19th century until the end of the Second World War. The source of Alcoa's monopoly was its control of bauxite ore, which is necessary to produce aluminium.

Rather than the vital resource being a mineral, it could be human resources (labour) over which the monopolist has control. For example, it would be very difficult for a new professional tennis organisation to compete successfully with the Association of Tennis Professionals (ATP), which has contracts with the best players and links to the most prestigious tournaments.

Legal barriers

The oldest and most effective barriers protecting a firm from potential competitors are the result of government ownership or the issuing of franchises and licences. This means that government operates the monopoly itself or permits a single firm to provide a certain product or service and excludes competing firms by law. For example, in many countries the reticulation of electric power, water and gas is often undertaken by monopolies established by national, state or local government. Many of Australia's state governments issue licences for monopoly casinos and have a monopoly in off-course betting on horse races. These gambling monopolies are usually supported by laws that apply severe penalties to persons setting themselves up in competition to the state-sponsored monopoly. Although governments often argue that these monopoly controls are designed to minimise the ill-effects of gambling, economists argue that the purpose of these laws is to guarantee the stream of revenue that government collects from these monopolists in the form of licence fees and taxes.

Government-granted licences restrict entry into some industries and occupations. For example, radio and television stations must be licensed. In most cases, doctors, lawyers, dentists, nurses, teachers, real estate agents, taxis, hotels and other professions and businesses are required to have a licence. In the case of the professions, the need for a licence does not preclude competition between licensed practitioners; however, it may serve to encourage them to form a guild or union, which then behaves as a monopolistic supplier on behalf of its members. One example is the Pharmacy Guild of Australia (PGA). The PGA is an organisation for the owners of pharmacies in Australia, and the pharmacists have their separate professional association: the Pharmaceutical Society of Australia. According to some critics, the PGA was formed primarily to protect the interest of its pharmacy owners by influencing the government to maintain restrictive rules and regulations through political donations and other lobbying activities.[1]

Patents and copyrights are another form of government-initiated barrier to entry. The government grants patents to inventors, thereby legally prohibiting other firms from producing the patented product for the duration of the patent (20 years in most countries). Copyrights give creators of literature, art, music and films exclusive rights to sell or license their works. The major purpose behind granting patents and copyrights is to encourage innovation and new products by guaranteeing, for a limited period, exclusive rights to profit from new ideas.

Economies of scale

Recall the concept of economies of scale from the chapter on production costs (**Chapter 6**). As a firm increases its scale of operation, the long-run average cost of production falls. **Exhibit 8.1** describes the market of town water. The firm's *LRAC* curve declines to the right. In **Chapter 7** our discussion of perfect competition shows that competition is good as it drives down the market price and cost of production for the benefit of the consumers. Would this be also true in the market of town water? As shown in the diagram, the firm will achieve a low average cost of production ($15 per unit) if it is the only supplier producing 100 units of output to meet the demand of the

[1] Read the following article: L. Russell, 'What is the Pharmacy Guild of Australia and why does it wield so much power?' *The Conversation*, 27 November 2019, https://theconversation.com/what-is-the-pharmacy-guild-of-australia-and-why-does-it-wield-so-much-power-127315

Global perspective

Applicable concepts: barriers to entry, patents, and economies of scale

Tesla: swapping one barrier for another?

In June 2014, Elon Musk, the CEO and co-founder of Tesla – a small but highly innovative car manufacturer in the US – announced that his company would be making all of its patents freely available. As well as producing a series of sexy battery-electric saloon cars with dazzling performance, Tesla also owns an increasing number of its own charging stations, which are the electric equivalent of the familiar petrol-retailing service stations.

An interesting and valuable characteristic of Tesla's technology is that it enables its electric cars to travel much further than its rivals before the batteries needs charging. This characteristic does much to reduce the degree of so-called 'range anxiety' – anxiousness about the possibility that the car's batteries will go flat before the driver's destination is reached. (Tesla cars are powered by batteries alone.) Tesla has also been working on a battery swap program that allows a new bank of batteries to be installed at one of its charge stations in less time than it takes to fill the tank of a conventional car.

Tesla's announcement that it would be making its patents freely available caused a flurry of wide-ranging explanations for this move. These included suggestions that Musk had seen the light and was simply spreading the love by embracing the idea that a world in which ideas are freely exchanged would be a better place.

Contrary to this view, and consistent with their assumption that individuals rationally pursue their own self-interest, economists looked for a more plausible explanation of Tesla's move. Of the many interpretations proffered by economists there are two that we shall look at here.

In order to understand these explanations, we need to be aware of one more important feature of Tesla's business. This is their investment in a huge battery plant in the US. This plant has become known as the 'Gigafactory'. With assistance from leading Japanese lithium-ion battery producer Panasonic, Tesla has built a huge plant employing 6000 people that will reap enormous economies of scale. It is estimated that these economies will reduce battery prices by at least 30 per cent, leading to a significant decrease in the price of new electric cars as well as a decrease in their overall running costs. In addition, this new plant will produce lower-cost batteries for a wide range of other applications. Tesla has plans to build more gigafactories around the world.

The first of the explanations for Tesla's release of its patents relates to its chain of charging stations and its proposal for these stations to include a high-speed battery-charging service. Economic commentators argue that if the process of charging batteries across all electric car manufacturers were to be standardised, then the company with charge stations that are specifically designed for this standardised process would soon have a monopoly or near monopoly in this area. If free access to its patents would encourage other electric car manufacturers to use Tesla's patented methods of charging batteries, this could lead to a vast expansion of Tesla's chain of charge stations. Tesla could then have a market advantage similar to that of Microsoft in the computing world. In this explanation of Tesla's decision to give free access to its patents, Tesla is in effect swapping its market-protecting patents for the protection that comes with ownership of a vital resource in the form of a network of charging stations that can service most if not all electric cars.

The second and more intriguing explanation is that Tesla is more interested in selling batteries than cars. It is argued that if, by making its patents freely available, Tesla can encourage increased manufacture of electric cars by its competitors then it will have a much larger market for its Gigafactory batteries. In particular, access to patents that allow its competitors to build cars with a much higher range would mean an increase in demand for electric cars as range anxiety diminished. In this scenario, Tesla is swapping the patent protection it currently gains in its car manufacturing business for protection in the form of economies of scale in its battery plant. From a purely economic perspective, if Tesla's strategy means that it gains more profits from its Gigafactory than it loses from its car factory then the strategy has been successful.

Getty Images/The Washington Post

Moreover, with control over the manufacture of batteries, which are one of the most important inputs into electric car production, it is possible that Tesla's car business will flourish because it will have a cost advantage over its rival car manufacturers. Furthermore, decisions made by the UK and France in 2017 to ban new diesel and petrol cars from 2040 will provide a great impetus to development of battery-powered electric vehicles.

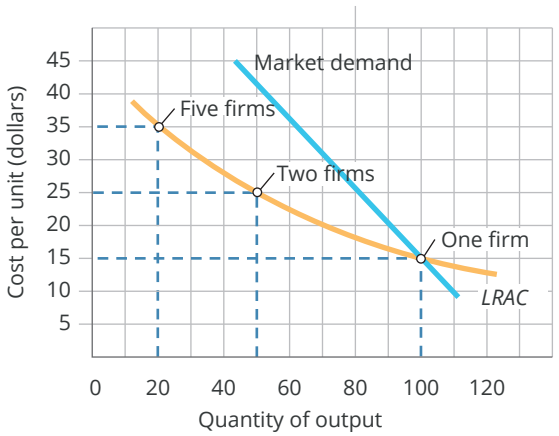
It will take some years for the effects of Tesla's move into large-scale battery manufacturing to become apparent. Only then will we be able to see whether the explanations for this move that have been proffered by economists turn out to be correct.

whole market. If a second firm enters the market, assuming the market is equally shared by the two firms, each firm will produce 50 units of output at the average cost of $25 per unit. The two-firm average cost is much higher than one-firm average cost because water supply involves very high fixed cost of installation of underground pipe network. A second firm means a second set of distribution network providing half of the market's demand, thus increasing the average cost of output. More firms and competition will push the average cost even higher. This means that a monopoly can emerge naturally because the full benefits of scale economies can only be enjoyed if the firm is the sole supplier in the market.

Economists call the situation in which one seller emerges in an industry because of economies of scale a **natural monopoly**. A natural monopoly is an industry in which the long-run average cost of production declines over the range of output where the entire market demand is met. As a result, a single firm can supply the entire market demand at a lower unit cost than two or more smaller firms. Utilities, such as electricity, gas and water, are examples of natural monopolies.

Natural monopoly An industry in which the long-run average cost of production declines throughout a large range of output. As a result, a single firm can supply the entire market demand at a lower cost than two or more smaller firms.

Exhibit 8.1 **Minimising costs in a natural monopoly**



In a natural monopoly, a single firm can produce at a lower average cost than two or more firms in an industry. This condition occurs because the *LRAC* curve for the firm decreases over the relevant range. For example, one firm can produce 100 units at an average cost of $15. Two firms in the industry can produce 100 units of output (50 units each) for an average cost of $25, and five firms can produce the same total output for an average cost of $35.

In the past, in order to prevent exploitation of consumers by these powerful natural monopolies, it was usual for them to be government-owned. Today, it is increasingly common to have these industries in private hands, with a regulatory authority such as the Australian Competition and Consumer Commission (ACCC) or the New Zealand Commerce Commission (NZCC) charged with preventing these private firms from exploiting their monopoly power. These experiments with privatisation have not always been successful as is attested by

an increasing number of disenchanted customers, especially price-hike-weary electricity consumers, voicing their dissatisfaction with their private suppliers. In some cases the regulatory authority limits exploitation by requiring these monopolists to give competitors access to their distribution network. For example, gas producers that own gas pipelines in Australia may be required to give access to other producers, as may mining companies and state governments that own railway lines.

In summary

- **Monopoly** is a single seller supplying the entire output of an industry. The demand curve that it faces is the entire industry demand curve for the good or service it sells.
- Barriers to entry to an industry include sole control of vital inputs, legal barriers and economies of scale.
- A **natural monopoly** arises because of the existence of economies of scale in which the *LRAC* curve falls over the relevant range of output as production increases.

2 Price and output decision for a monopolist

In order to study the behaviour of a monopolist, we need to understand a major difference between perfect competition and monopoly, which is the demand curve faced by the firm. Recall from the previous chapter that the competitive firm is a *price taker*. This is because it faces a horizontal demand curve and has no influence over the prevailing market price. The competitive firm takes the market price and select the output that maximises its profit. In contrast, a monopolist is a **price maker**. A price maker is a firm that faces a downward-sloping demand curve. This means that a monopolist can set the price of its product with its corresponding level of output to maximise its profit, rather than simply determining its production level in response to the going industry price. We will apply the marginal analysis approach to a hypothetical kitchenware company, We-are-Kitch, to model the behaviour of a monopolist.

Price maker
A firm that faces a downward-sloping demand curve and can therefore choose among price and output combinations along the demand curve.

Demand and marginal revenue of a price maker

Suppose technologists at a firm called We-are-Kitch invent a kitchen appliance they call 'Pizzaronto', which anyone can easily install in their kitchen. Once installed, the appliance transforms a set of pizza ingredients – including flour, yeast, tomato and toppings – into a ready-to-bake pizza that can be simply popped into a hot oven. The government grants We-are-Kitch a patent and the company becomes a monopolist selling these new appliances. Because of this barrier to entry, We-are-Kitch is the only seller in the industry. Although other firms try to compete with this invention, they create poor substitutes and fail. Because We-are-Kitch is the only supplier of Pizzaronto in the market, the market demand curve and the demand curve facing We-are-Kitch are identical.

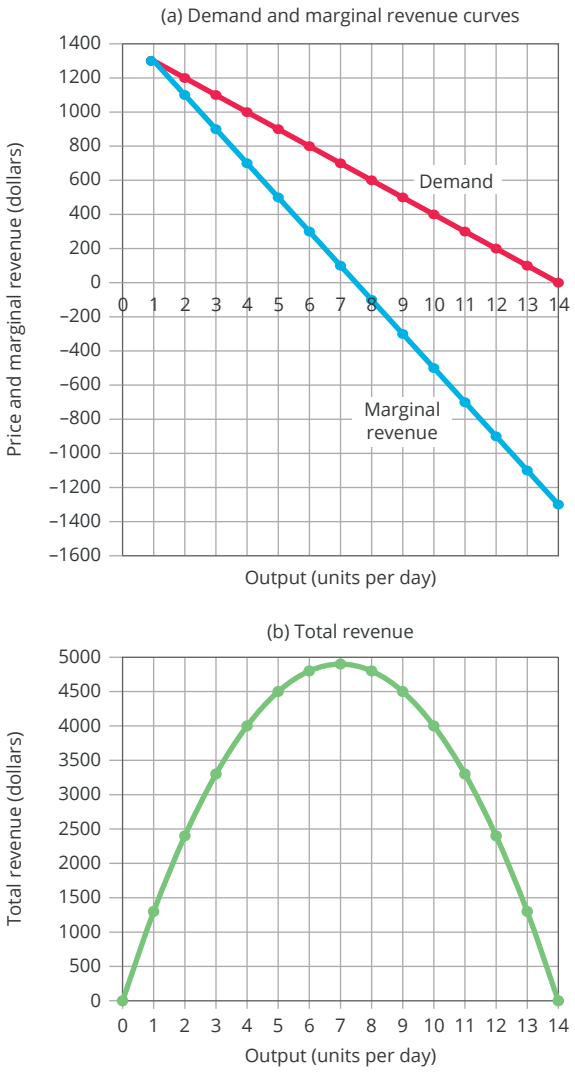
Part (a) of **Exhibit 8.2** illustrates the demand and the marginal revenue curves for We-are-Kitch. Facing a downward sloping demand curve, We-are-Kitch can sell more output at a lower price. Column 3 presents the firm's total revenue (price times quantity), as shown graphically in part (b) of **Exhibit 8.2**. If We-are-Kitch charges a price above $1300, consumers purchase zero units and, therefore, total revenue is zero. To sell one unit, We-are-Kitch must lower the price to $1300 and total revenue rises from zero to $1300. Because the marginal revenue is the increase in total revenue that results from a one-unit change in output, the marginal revenue from selling the first unit of output is $1300 ($1300 – $0). To sell two units, the monopolist must lower the price to $1200 for *both* units and total revenue rises to $2400. The marginal revenue of the second unit is $1100 ($2400 – $1300). Note that the marginal revenue from selling the second unit is less than the price of $1200 received from that unit. The marginal revenue from selling the third unit is $900, less than its selling price of $1100. Similar calculations produce the rest of the table. Note that the marginal revenue is below price except for the first unit of output. This outcome is different from the case of the competitive firm, for which marginal revenue and price are the same.

Now we can examine the relationship between total revenue and output. You can see that, starting from zero output, as the price falls and sales increase, total revenue rises until it reaches a maximum at 7 units of output (the elastic range) and then total revenue falls after that (the inelastic range). Graphically, this is shown as the 'revenue hill' drawn in part (b) of **Exhibit 8.2**. The explanation was presented earlier in **Chapter 5**.

The firm will never operate on the inelastic range of its demand curve that corresponds to negative marginal revenue. The reason is that it would be silly for a firm to incur the extra cost of increasing output only to find that sales revenue (total revenue) falls (i.e., negative MR). In our example, We-are-Kitch would not charge a price lower

Exhibit 8.2 **Demand, marginal revenue and total revenue for We-are-Kitch**

Output per day	Price	Total revenue	Marginal revenue
0	1 400	0	–
1	1 300	1 300	1 300
2	1 200	2 400	100
3	1 100	3 300	900
4	1 000	4 000	700
5	900	4 500	500
6	800	4 800	300
7	700	4 900	100
8	600	4 800	−100
9	500	4 500	−300
10	400	4 000	−500
11	300	3 300	−700
12	200	2 400	−900
13	100	1 300	−1 100
14	0	0	−1 300



Part (a) shows the relationship between the demand and the marginal revenue curves. The *MR* curve is below the demand curve except for the first unit of output. Between one and seven units of output $MR > 0$; beyond seven units of output $MR < 0$.

The relationship between demand and total revenue is shown in part (b). When the price is \$1400, Q = 0, total revenue is zero. When the price is set at zero, Q =14, total revenue is also zero. In between these two extreme prices, the price of \$700 maximises total revenue. This price corresponds to seven units of output.

than $700 and produce an output greater than seven units per day. But what price will the monopolist charge and how much output will it produce if it wishes to maximise profit? We will now look at the answer to this question.

Monopoly in the short run

Like all firms, a monopoly will determine its profit maximisation output by considering the revenue and cost of production. The table in part (a) of **Exhibit 8.3** presents We-are-Kitch's costs of production at various levels of output. Notice that for simplicity the marginal cost (MC) of production is assumed constant at $500 per unit with a fixed cost of $600. Since MC is constant, AVC equals MC and so the AVC and MC curves overlap. Although AVC curve is horizontal, the average total cost (ATC) curve slopes down to the right since ATC = AVC + AFC and average fixed cost (AFC) slopes down to the right. The firm's ATC, AVC and MC curves together with its demand and MR curves from **Exhibit 8.2** are shown in part (a) of **Exhibit 8.3**. Part (a) illustrates a situation in which We-are-Kitch can earn an economic profit because part of its ATC curve is below the demand curve, a point that we will clarify later. As with the competitive firm, we apply the marginal analysis method to determine We-are-Kitch's profit maximisation output. A monopolist maximises profit by producing the quantity of output where *MR* = *MC* and charging the corresponding price on its demand curve. In this case, 5 units is the quantity at which *MR* = *MC*. As represented by point *A* on the demand curve, at 5 units of output, the price charged is $900 per unit.

At the output of 5 units the *ATC* is $620 (point B). Because price is greater than the *ATC* at the *MR* = *MC* output, the monopolist earns an economic profit of $280 per unit. At the daily output of 5 units, total economic profit is $1400 per day, as shown by the shaded area.

We can now see that if the demand curve is above the ATC curve over a certain range of output, the firm can make a profit by producing within this range. To maximise this profit it selects a quantity of output where *MR* = *MC*, as discussed.

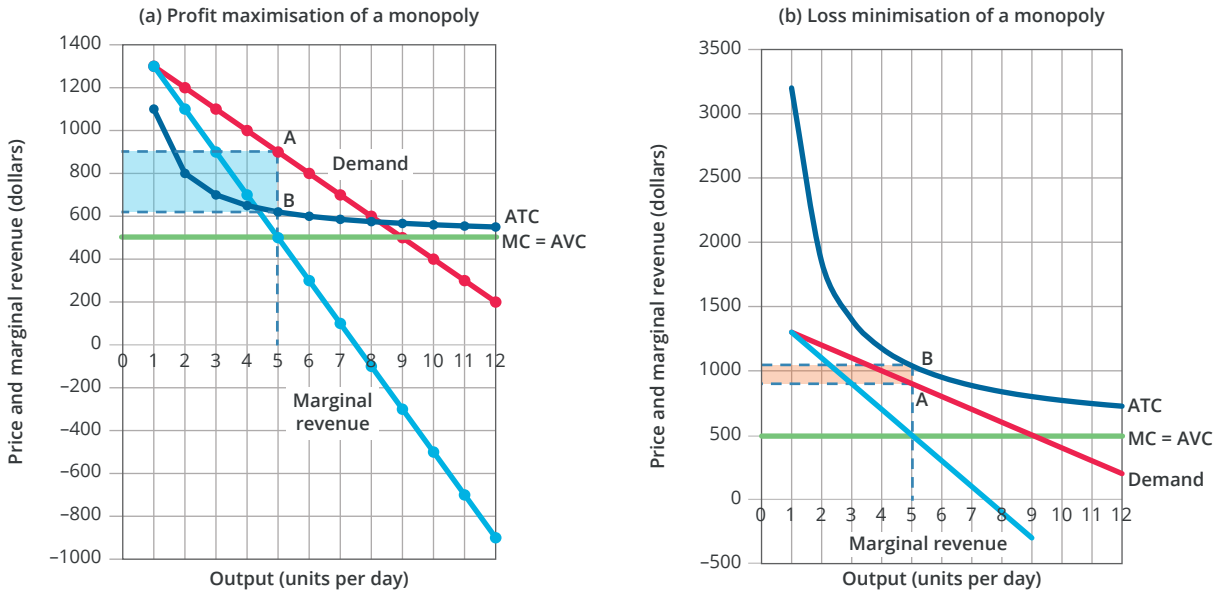
It is important to also note here that, unlike the competitive equilibrium where MU = P = MC, the monopolist charges a price that is greater than marginal cost (P > MC). In other words, since P = MU, at the profit-maximisation output of 5 units, MU > MC. It means that the firm could create more value by increasing output beyond 5 units as discussed in **Chapter 4** and as a result increase social surplus. If you are wondering why the monopolist would not increase output even though price is greater than marginal cost, think for a moment about the motive of the firm, which is to maximise its own profit. As you can see from part (a) of **Exhibit 8.3**, if the monopolist were to increase output beyond the level where *MR* = *MC*, marginal revenue will be below the marginal cost, resulting in a fall in profit. As you will see shortly, the fact that the profit-maximising monopolist charges a price greater than marginal cost has important implications for the question of whether monopoly compromises economic efficiency.

Observe that a monopolist charges neither the highest possible price nor the revenue-maximising price. In part (a) of **Exhibit 8.3**, $900 is not the highest possible price. Because We-are-Kitch is a price maker, it could have set a price above $900 and sold less output than 5 units. However, the monopolist does not maximise its profit by charging the highest possible price. Now note that 5 units is *not* the output that maximises total revenue. If this monopolist did choose to maximise revenue by producing 7 units, you can see that it would fail to maximise its profit because it would not be following the *MR* = *MC* profit-maximising rule. The only case in which a monopolist would maximise total revenue at the same time as it maximised profit would be when its marginal cost of production was zero. A monopolist producing with zero marginal cost is an unlikely case.

The fact that a firm has a monopoly does not guarantee economic profits. A monopolist has no protection against changes in demand or cost conditions. Suppose that following a spate of sabotage by Italian slow food advocates, We-are-Kitch faces much higher insurance charges causing big increases in its fixed cost to $2700. Part (b) of **Exhibit 8.3** presents the firm's new ATC schedule and ATC curve. Note that the *ATC* curve is now everywhere higher than the demand curve. This implies that P < ATC at any level of output and the firm will make a loss. The best We-are-Kitch can do is to minimise its loss by producing where *MR* = *MC*. In this example, the marginal cost curve has not changed because the increase in insurance costs represents a change in fixed cost only. Since the marginal cost and demand curves are unchanged, the loss minimisation output is the same as before at 5 units

per day, charging a price of $900 (point A). At output of 5 units, the *ATC* is $1040 (point *B*) and We-are-Kitch takes a loss of $140 per unit, and a total loss of $700 per day as represented by the shaded area ($140 × 5 units). In the short run, even though We-are-Kitch makes a loss, it will continue to produce as the price is above AVC and can cover part of the fixed cost.

Exhibit 8.3 We-are-Kitch's costs of production



(a) Profit maximisation

Output per day	Total cost	ATC	MC
0	$ 600	–	$500
1	1 100	$1 100	$500
2	1 600	800	500
3	2 100	700	500
4	2 600	650	500
5	3 100	620	500
6	3 600	600	500
7	4 100	586	500
8	4 600	575	500
9	5 100	567	500
10	5 600	560	500
11	6 100	555	500
12	6 600	550	500

(b) Loss minimisation

Output per day	Total cost	ATC	MC
0	$2 700	–	–
1	3 200	$3 200	$500
2	3 700	1 850	500
3	4 200	1 400	500
4	4 700	1 175	500
5	5 200	1 040	500
6	5 700	950	500
7	6 200	885	500
8	6 700	837	500
9	7 200	800	500
10	7 700	770	500
11	8 200	745	500
12	8 700	725	500

Part (a) illustrates We-are-Kitch maximising profit by producing 5 units of output, which corresponds to the intersection of the marginal revenue and the marginal cost curves. The price the monopolist charges is $900, which is point A on the demand curve. Because the price ($900) is above the *ATC* ($620) at the output of 5 units, the monopolist earns an economic profit of $1400 per day, represented by the shaded area.

In part (b), We-are-Kitch makes a loss. Here the *ATC* curve is higher than it was in part (a) because of the much higher fixed cost of $2700, with the result that the demand curve lies below the *ATC* curve at all output levels. Since its MC is unchanged, its MR = MC output is the same as before. It minimises its loss by producing 5 units of output, charging $900 (point A). At 5 units, ATC is $1040 per unit which is above the price of $900. The shaded area shows the firms' loss at $700 per day.

What if demand goes down and the *demand* curve is below the *AVC* curve for a monopolist? As under perfect competition discussed in **Chapter 7**, the monopolist will shut down. To operate would only add further to losses.

Monopoly in the long run

In perfect competition, economic profits are impossible in the long run. The entry of new firms into the industry drives the product's price down until economic profits are eventually reduced to zero. High barriers to entry, however, protect a monopolist. If We-are-Kitch is making a profit as shown in part (a) of **Exhibit 8.3**, it will continue to enjoy that profit in the *long run*.

In the long run, the monopolist has increased flexibility. The monopolist can alter its plant size in order to lower cost, just as a perfectly competitive firm can. But like competitive firms, if We-are-Kitch is making losses in part (b) of **Exhibit 8.3**, it will not remain in business in the long run when losses persist – regardless of its monopoly status.

In reality, no monopolist can depend on barriers to protect it fully from competition in the long run. One threat is that entrepreneurs will find innovative ways to compete with a monopoly. For example, We-are-Kitch must fear that other firms will use their ingenuity to develop a better and cheaper complete pizza-making appliance. To dampen the enthusiasm of potential rivals, one alternative is to sacrifice short-run profits to secure greater profits in the long run. Returning to part (a) of **Exhibit 8.3**, the monopolist might wish to charge a price below $900 and produce an output greater than 5 units per day. In this way it would send a signal to potential rivals that it is willing to sacrifice profits in the short run to discourage the entry of rival firms. A decision by a firm to discourage potential competition by temporarily reducing prices is known as *predatory pricing*.

In summary

- A **price maker**, such as a monopolist, faces a downward-sloping demand curve.
- The marginal revenue is below the demand curves for a **price-maker** (except for the first unit).
- The monopolist produces the profit-maximising output where the *MR* and the *MC* curves intersect like the competitive firm, and locates the profit-maximising price on its demand curve.
- In the long-run, the monopolist can enjoy an economic profit because of barriers to entry. But changes in market conditions (e.g., a fall in demand) can result in the monopolist making a loss.

3 Price discrimination

Price discrimination
This occurs when, for the same product, a seller charges different customers different prices not justified by cost differences.

Our discussion so far has assumed that the monopolist charges each customer the same price. What if We-are-Kitch decides to try to sell identical Pizzaronto units for, say, $900 to restaurant owners and $500 to householders? Why will it do so? Under certain conditions, we will see that a price maker may practise **price discrimination** to increase profit. Price discrimination occurs when, for the same product, a seller charges different customers' different prices not justified by cost differences. By practising price discrimination the seller is able to increase the revenue it receives above the level that would prevail if all buyers were charged the same price. Examples of price discrimination are plentiful; some hairdressers charge female customers a higher price than male customers; and you can often pay a much lower price for a take-away pizza if you present a discount coupon.

Conditions for price discrimination

Not all sellers can engage in price discrimination. The following three conditions must exist before a seller can price discriminate.

1 The seller is a price maker and therefore faces a downward-sloping demand curve. As you will see in the next chapter, this means monopoly is not the only market structure in which this practice may appear.

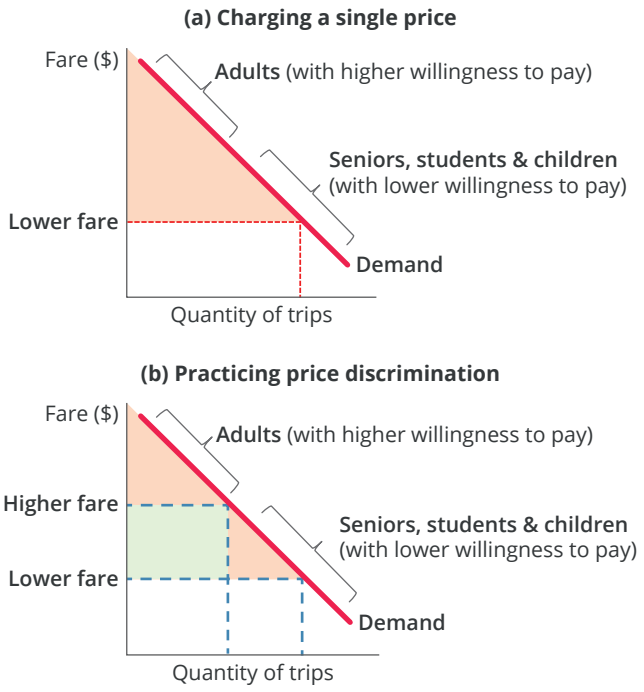
2 The seller is able to segment the market by distinguishing between different groups of consumers. This separation of buyers will be shown to be based on different willingness to pay.[2]

3 It is impossible or too costly for customers to resell the product. In most cases, services bought by one customer cannot be re-sold to another. For example, a husband cannot pay a lower price and have a haircut *for* his wife.

The providers of passenger train services meet the conditions for price discrimination. First, because a downward-sloping demand curve is faced, a lower (concession) fare will increase the quantity of tickets sold. Second, a passenger's age, or status as a student, can easily be determined and verified, allowing the rail authority to classify different groups of consumers with different willingness to pay. Children, seniors and students usually have lower incomes and a lower willingness to pay than adults of working age. Third, the nature of the service and the conditions of travel prevent reselling of train tickets. A child, senior or student cannot buy a rail ticket at a low price and sell it to an adult for a higher price. Children are visibly younger than working-age adults, and seniors and students are required to carry identification that indicates their eligibility for the concession fare.

Exhibit 8.4 illustrates how the rail authority practises price discrimination. To understand the implication of a downward sloping demand curve, imagine we line up all train passengers on a demand curve in descending

Exhibit 8.4 **Price discrimination**



To maximise profit, the rail authority separates travellers into two groups: those eligible for concession fares and those not. Therefore, the rail authority sets a full fare for trips by working-age adults and a concession fare for trips by children, seniors and students. By using price discrimination to extract consumer surplus (green rectangle) from working-age adults, the rail authority earns a greater profit than it would by charging a single price to all passengers.

[2] There are many ways to price discriminate. In this chapter we will only consider between-group price discrimination based on willingness to pay. It is noted that we can also analyse between-group price discrimination based on different elasticities of demand

order of willingness to pay for the service. Working age adults tend to occupy the upper part of the demand curve (the higher value customers) since they have higher willingness to pay, whereas seniors, children and students sit on the lower part of it (the lower value customers).

To attract more passengers, the rail authority can charge a lower fare *for all*, if it does not practise price discrimination. As shown in part (a) the consumers will enjoy a large consumer surplus (orange triangle). But one may ask: Why do we charge the working age adults (higher value customers) a lower price when they are willing to pay more for the service? Realising that the rail authority can price discriminate, setting a higher (full) fare for working-age adults and a lower (concession) fare for children, seniors and students. By charging working-age adults a higher fare, the rail authority can extract some of their consumer surplus (green rectangle) as shown in part (b), and earn a higher profit.

In the You're the Economist section, you will have opportunities to find out more about other interesting methods of price discrimination.

Economics and ethics

Price discrimination vs price fixing

Examples of price discrimination abound. Cinemas offer lower prices for children than for adults, although the cost of providing a seat is the same for both. Hotels, restaurants and hairdressers often give discounts to senior citizens – and not because they eat less or have less hair! Local authorities, such as city and shire councils, often give pensioners a discount on their rates bill.

A common reaction to price discrimination from those who do not benefit from it is that it is unfair. After all, if it costs the same to provide a meal to a pensioner as it does to anyone else, why should the hotel or restaurant give a discount to the pensioner? But look at the other side of price discrimination. First, the seller is pleased because price discrimination increases profits. Second, many buyers benefit from price discrimination by not being excluded from purchasing the product. In **Exhibit 8.4**, price discrimination makes train travel possible for concession fare travellers who could not afford to pay a higher fare. Price discrimination allows retired people to enjoy hotels and restaurants they could not otherwise afford, it allows more children to go to the cinema, and it enables pensioners to more easily afford the costs of living in their own homes.

Here we must distinguish between *price discrimination* and *price fixing*. By and large, economists do not support firms abusing their market power by charging customers higher prices through collusion or other means, rather than being constrained by the competitive market forces, as occurs with perfect competition. In its 2018 report on the financial sector competition,[1] the Productivity Commission found that the four big banks in Australia abused their market power to achieve big profits by using strategies such as opaque pricing and commissions, and selling high margin products that are of no or little benefits to customers. This is price fixing, not price discrimination. Nonetheless, if price discrimination means that lower prices are available to the needy when they use public transport, visit the hairdresser or use doctors' and lawyers' services, then, from an ethical perspective, there is some merit in the practice.

[1] Australian Government Productivity Commission, 'Competition in the Australian financial system' 3 Aug 2018, https://www.pc.gov.au/inquiries/completed/financial-system#report

You're the economist

What do publishers, car makers, IBM and flat-pack furniture deliverers have in common?

There is a long-standing tradition of publishers releasing first editions of books in hard cover. Sometime later when hard cover sales have dwindled, the publisher then releases a cheaper paperback edition. When you pay a higher price to buy the 'luxury' or 'special' model of a new car you typically find that a number of features not offered in the 'standard' model will be incorporated into the vehicle. In 1990 IBM released a new model LaserPrinter E, a cheaper version of its popular LaserPrinter. The cheaper model prints at about half of the speed of the more expensive model. You will think LaserPrinter E is cheaper in price because it is cheaper to make. What you may not know is that the two models are almost identical in their technology and component parts, and LaserPrinter E is actually slightly *more expensive* to make. Why?

IKEA, the world's best-known retailer of flat-pack furniture, will arrange delivery of your purchase to your home if you wish. At its Perth and Adelaide stores IKEA's preferred carriers charge a flat rate for delivery anywhere within the metropolitan area, regardless of the distance of delivery address from the IKEA store or warehouse. While each of these examples relates to very different industries, there is something they have in common. What is it?

In summary

- **Price discrimination** allows the monopolist to increase profits by charging different groups of buyers different prices based on their willingness to pay, rather than charging a single price to all buyers.

4 Comparing monopoly and perfect competition: the case against and for monopoly

Now that the basics of the two extremes of perfect competition and monopoly have been presented, we can compare and evaluate these market structures. This is an important assessment, because the contrast between the disadvantages of monopoly and the advantages of perfect competition is the basis for many government policies, especially those enshrined in competition policy. To keep the analysis simple, we will assume that the monopolist charges a single price rather than engaging in price discrimination.

The monopolist as a resource misallocator

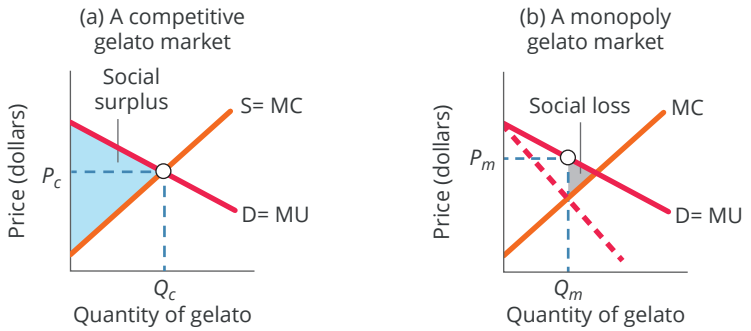
Recall the discussion of market efficiency from **Chapter 3** and the explanation in **Chapter 7** as to why perfect competition is efficient. Efficiency exists when a market utilises resources to create the greatest social surplus. Social surplus is maximised when output is produced up to the point where marginal utility equals marginal cost ($MU = MC$). In the absence of externality competitive equilibrium is efficient. Why? This is because demand is determined by marginal utility and supply by marginal cost. In competitive equilibrium where demand equals supply, the condition of $MU = MC$ is also satisfied. Therefore social surplus is maximised in competitive equilibrium in the absence of externalities.

On the other hand, a monopolist restricts output and charges a price above marginal cost to maximise own profit. Therefore in monopoly equilibrium MU > MC. The condition of maximisation of social surplus is not satisfied.

To illustrate this important idea, let us revisit the gelato market from **Chapter 3** and consider two scenarios: (a) a competitive gelato market and (b) a monopoly gelato market. Part (a) of **Exhibit 8.5** shows a competitive gelato market producing equilibrium output (Q_c) and setting a competitive price (P_c). At Q_c, $MU = MC$ and social surplus (shaded area) is maximised.

In contrast, consider a gelato market with only one producer as shown in part (b) of **Exhibit 8.5**. This could happen if Gela bought all gelato shops in the town and became the only seller of gelato cones. In a monopoly, the market demand curve *is* the monopolist's demand curve. The monopolist's MR curve lies below the demand curve. To maximise profit, the monopolist sets $MR = MC$ by restricting the output to Q_m (monopoly output) and raising the price to P_m (monopoly price). At the monopoly output of Q_m, marginal utility is greater than marginal cost ($MU > MC$). In this situation, there are some consumers who value gelato at more than its marginal cost of production. In spite of valuing it at more than the marginal cost of production, they are unable to purchase

Exhibit 8.5 **Comparing perfect competition and monopoly**



In (a) the intersection of market supply (S) and market demand (D) establishes the equilibrium price of P_c and the equilibrium quantity of Q_c. At Q_c MU = MC, and social surplus (blue area) is maximised. Now assume the industry suddenly changes to a monopoly as shown in (b). The monopolist produces the *MR* = *MC* output of Q_m, which is less than Q_c. By restricting output to Q_m, the monopolist is able to charge the higher price of P_m and maximise its profit. Under monopoly, inefficiency occurs because social surplus is no longer maximised and a social loss (pink area) is resulted.

more of it. Had the monopolist used more resources and produced additional units up to the competitive output (Q_c), social surplus would be maximised. But the monopolist's objective is to maximise own profit rather than social surplus. Therefore, it restricts output to Q_m to maximise profit resulting in a *social loss* – the grey-shaded triangle in part (b). Observe also that while the producer surplus is increased, the consumer surplus is reduced since the consumers pay more for less. (Can you indicate the producer surplus and consumer surplus in the diagram?) Thus, in monopoly not only are resources misallocated, but there is also a redistribution of income from the consumers to the firm.

Anything good about monopoly?

So far, a case has been made against monopoly and in favour of perfect competition. Now it is time to pause and summarise the economist's case against monopoly, as follows:

- A monopolist gouges (overcharges) consumers by charging a higher price than would be the case under perfect competition.
- Because a monopolist's profit-maximising output is less than that for a competitive industry, too few resources are allocated to production of the product. Stated differently, the monopolist is responsible for a misallocation of resources because it charges a price greater than marginal cost by restricting output. In competitive industries, output is produced where marginal utility is equal to marginal cost and the result is maximisation of social surplus and an optimal allocation of resources.
- Monopolists earn long-run economic profit, which contrasts with the zero economic profit in the long run for a competitive firm.
- Monopoly alters the distribution of income in its favour.

However, not all economists agree that monopoly is bad. Economists Joseph Schumpeter and John Kenneth Galbraith have praised monopoly power. They have argued that the rate of technological change is likely to be greater under monopoly than under perfect competition. Their view is that monopoly profits afford giant monopolies the financial strength to invest in the well-equipped laboratories and skilled labour for research and development (R&D) necessary to create technological change. Monopoly also means the monopolist can capture all the profits from innovations. Therefore, it is not only able to but also has the incentive to invest in R&D.

The counter-argument is that monopolists are slow to innovate. Freedom from direct competition means the monopolist is not motivated and therefore tends to stick to the 'conventional wisdom'. As Nobel laureate Sir John Hicks put it, 'The best of all monopoly profit is a quiet life'. In short, monopoly offers the opportunity to relax a bit and not worry about the 'rat race' of technological change.

What does research on this issue suggest? Not surprisingly, many attempts have been made to verify or refute the effect of monopoly on technological change. Unfortunately, the results to date have been inconclusive. It may well be that a mix of large and small firms in an industry is optimal for creating competition to drive down prices and incentives for technological innovations.

Nevertheless, the pervasiveness of competition policy in free-market economies around the world does suggest that there is widespread support for the view that monopoly power should be curbed. Recall that in the previous chapter we noted that many of the regulatory activities of bodies like the ACCC or the NZCC are designed to limit unethical behaviour in the business sector when big firms abuse their market power – behaviour that Adam Smith might have described as displaying unenlightened self-interest.

Before we end this chapter, recall that one of the barriers to entry discussed earlier was economies of scale. If monopoly exists because of economies of scale (natural monopoly), then it does not necessarily follow that more competition will produce more output for less. For a similar argument, a merger may benefit the community if it can reduce its costs of production by reaping the benefits of scale economies. Any attempt by the competition policy agency to break up a natural monopoly into smaller firms or to ban mergers may result in an increase in production costs because economies of scale have been lost. This situation is particularly relevant to Australia's agricultural industries facing a small domestic market when they need scale economies to lower its costs of production to be competitive in today's global market. How to balance the pros and cons of concentrated market power? This issue will be explored in the 'Analyse the issue' section and further discussed in the next chapter.

Applicable concepts: competition versus monopoly, economies of scale, and competition policy

Analyse the issue

Does size matter?

Over the past three decades, the Australian government and many Australian firms have pursued the goal of globalisation. This has led to increasing integration of the Australian economy into the world economy. It has also resulted in a number of unforeseen developments, such as the desire of some Australian firms to shift their headquarters offshore to take advantage of more favourable tax regimes, and the emergence of arguments to the effect that Australia's competition policy should be less committed to the ideal of high levels of competition in the Australian economy. This latter development reflects the increasing competition faced by Australian firms in international markets.

The argument based on increasing international competition promotes the idea that Australian firms competing in world markets must get bigger or get out: if Australian firms are prevented by competition policy from merging with other Australian firms, there is no way that they can achieve sufficient size to be competitive in world markets. In pushing this argument, the business community points to examples like Qantas, which has a large chunk of the Australian market but less than 3 per cent of the world market; or Australia's big four banks, which are nevertheless small by international standards.

In the past, the response from the ACCC to these arguments has been as vehement as the arguments themselves. Here is part of what the ACCC chairperson at the time, Professor Allan Fels, had to say about the issue in a speech to the Australia–Israel Chamber of Commerce in March 2001:

> Australia is far from alone in its interest in mergers, and the Commission's critics should realise that governments worldwide have created strong laws to prevent the creation of cosy cartels. They have also empowered strong anti-trust agencies to enforce laws to protect consumers and small business from anti-competitive mergers.

> If the big business critics were serious about promoting the development of large internationally competitive Australian companies [they] would acknowledge the benefits of competition. Would Australia's big companies be internationally competitive if they had to secure their raw materials, such as coal and petrol, from a monopoly supplier? How would they fare if they had to export their goods through a monopoly transport company and raise finance from a monopoly bank? Would they be better off if they purchased their products from a monopoly retailer, 'Colesworth'?[1]

Acknowledging that the foundations of competition policy are to be found in economic theory, Professor Fels concluded his speech by remarking that '[t]he anti-competitive conduct provisions of the *Trade Practices Act* [now known as the *Competition and Consumer Act 2010*], including the merger provisions, are an attempt to enact economics as law'.

In 2014, Murray Goulburn, Australia's biggest dairy producer with one-third of industry output, was effectively blocked from buying Warrnambool Cheese and Butter Company because of the adverse effect of reduced competition in the Warrnambool areas on dairy farmers, as argued by the government authority. In response to the Takeover Panel's decision. Senator Barnaby Joyce questioned what is more important – a 'competitive industry' or an internationally 'competitive firm', calling for a review of the competition law to 'support the creation of national champions in industries across Australia'.

Fairfax Media/Sasha Woolley

In a report commissioned by the Business Council of Australia (BCA) from McKinsey & Company, the BCA made its position on this issue very clear:

> [w]hen economies of scale are required to compete globally, Australia's relatively small market size can be an impediment, especially if merger and acquisition rules prevent the consolidation required to match international rivals.[2]

Will the benefit of more competition be outweighed by the productivity gain when a firm is allowed to grow in size by merger? It is always a delicate balancing act. More recently, in early 2020, the ACCC strongly opposed the proposed merger of TPG Telecom and Vodafone claiming 'competition is lost when incumbents acquire innovative new competitors'.[3] Justice Middleton, however, decided to give the merger the green light, arguing that the merged company will have more capacity in investing its 5G network to 'become a more effective competitive constraint on Telstra and Optus'.[4] The ACCC decided not to appeal against the Federal Court's decision. The merger went ahead.

What do you think?

Using the information provided, answer the following questions:

1. Using appropriate cost concepts, explain why some Australian firms may need to become big if they are to compete internationally.
2. Elsewhere in his 2001 speech, Professor Fels pointed out that an overwhelming majority of merger applications were approved by the ACCC, although the TPG-Vodafone merger application was not one of them. Why do you think the High Court approved the TPG–Vodafone merger against the ACCC's advice?

[1] A. Fels, Chairperson ACCC, 'Mergers and big business', speech presented at boardroom lunch, 'Mergers and market power', 15 March 2001, Sydney.

[3] J. Lydon, D. Dyer and C. Bradley, 'Compete to prosper: Improving Australia's global competitiveness', McKinsey & Company, 2014, https://www.mckinsey.com/global_locations/pacific/australia/en/latest_thinking/compete_to_prosper, accessed 15 September 2014.

[4] ACCC, 'ACCC will not appeal Federal Court's decision to allow TPG-Vodafone merger', 5 March 2020, https://www.accc.gov.au/media-release/accc-will-not-appeal-federal-court%E2%80%99s-decision-to-allow-tpg-vodafone-merger

[5] J. Fernyhough, 'ACCC savaged in TPG-Vodafone judgment', *Financial Review*, 4 March 2020, https://www.afr.com/companies/telecommunications/accc-savaged-in-tpg-vodafone-judgment-20200304-p546uz

In summary

- When compared to perfect competition, a monopolist charges a higher price than a competitive firm.
- Monopoly is inefficient as social surplus is not maximised.
- Monopoly may be beneficial to consumers if it enjoys significant economies of scale to lower cost of production and to produce more for less.

Key concepts

Monopoly | Natural monopoly | Price maker | Price discrimination

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Describe the characteristics of the monopoly market structure

- **Monopoly** is a single seller supplying the entire output of an industry. The demand curve that it faces is the market demand curve. The monopolist sells a unique product, and high barriers to entry protect it from competition.
- Barriers to entry that prevent new firms from entering an industry are (1) ownership of an essential resource, (2) legal barriers and (3) economies of scale.
- A **natural monopoly** arises because of the existence of economies of scale in which the *LRAC* curve falls continuously over the output range that meets the market demand. Economies of scale allow a single firm to produce at a lower cost.

2. Determine the price and output decisions for a monopolist

- **Price-maker** firms, such as a monopolists, face a downward-sloping demand curve.
- The marginal revenue for a monopolist is below the demand curve (except for the first unit) of output.
- The **monopolist**, like the competitive firm, locates the profit-maximising output where the *MR* and the *MC* curves intersect, and set a price on its demand curve.
- In the long-run the monopolist can continue to earn an economic profit because of barriers to entry. If, in the long run, demand and cost conditions prevent the monopolist from earning at least a normal profit, it will leave the industry.

3. Understand how monopolists can engage in a practice known as price discrimination

- **Price discrimination** allows the monopolist to increase profits by charging different buyers different prices, rather than charging a single price to all buyers. Three conditions are necessary for price discrimination: (1) the demand curve faced by the firm is downward-sloping, (2) buyers with different willingness to pay can be identified and separated, and (3) buyers are prevented from reselling the product at a higher price than their purchase price.

4. Evaluate the case against and for monopoly by comparing monopoly and perfect competition

- When compared to perfect competition, monopoly disadvantages are: (1) a monopolist charges a higher price than a competitive firm; (2) resource allocation is inefficient because the monopolist produces less output than if competition existed; and (3) monopoly transfers income from consumers to producers.
- Monopoly through economies of scale can produce at a lower cost than smaller firms in perfect competition. The monopolist may also be more willing and more able to invest in R&D because they can capture all profits from technological innovations.

Study questions and problems

1 Using the three characteristics of monopoly, explain why each of the following is a monopolist in relation to at least some of the output it produces:
- **a** The only casino in an Australian capital city.
- **b** The provider of the Sydney Harbour Bridge climb.
- **c** The NBN Company.

2 In spite of the finding that monopolists charge higher prices and produce less output than competitive firms, some economists argue that monopolists benefit society (they raise economic welfare). What are some of the reasons that these economists give for taking this position?

3 Explain the conditions required for a price-maker to be able to engage in price discrimination. Give an example of an Australian price-maker – large or small – that practises price discrimination.

4 Explain why a monopolist would never produce in the inelastic range of the demand curve.

5 **a** Use the following demand schedule for a monopolist to calculate total revenue and marginal revenue.
b The firm's marginal cost is presented in the last column of the table. Determine the profit maximising output and price of the monopolist. Do you know if the firm is making a profit or a loss?

Price	Quantity demanded	Total revenue	Marginal revenue	Marginal cost
$5.00	0	-	-	-
4.50	1			1.50
4.00	2			2.00
3.50	3			2.50
3.00	4			3.00
2.50	5			3.50
2.00	6			4.00
1.50	7			4.50
1.00	8			5.00
0.50	9			5.50

6 Make the assumption that marginal cost for the monopolist in question (5) is zero like the cases described in the Global Perspective box in **Chapter 6**. Given the data from the demand schedule in question (5), what price will the monopolist charge, and how much output should the firm produce?

7 To what extent do you think an operating system such as Microsoft Windows displays the characteristic of a natural monopoly?

8 Suppose the lawn-mowing industry approximates a perfectly competitive industry. Suppose also that a single firm buys all the assets of the lawn-mowing firms and establishes a monopoly. Contrast these two market structures

with respect to price, output and allocation of resources. Draw a graph of the market demand and market supply for lawn-mowing services before and after the takeover. Would you expect the monopoly to maintain its monopoly power in the long run?

9 Typically, retail prices for clothing and shoes of a particular brand and style are the same regardless of size. Why do you think this occurs when larger sizes use more raw material in their production and are, *ceteris paribus*, more costly to produce? Are firms that sell these products at a single price regardless of size engaging in price discrimination?

Answer to 'You're the economist'

What do publishers, car makers, IBM and flat-pack furniture deliverers have in common?

Price discrimination occurs when different customers are charged different prices where the difference in price does not reflect differences in cost, or where customers are charged the same price although costs differ. The publishers' practice of initially releasing a hardback edition enables them to charge a much higher price to those buyers who must get their hands on a copy of the new book as soon as possible. These are customers who place a high value on being able to read the book early, high up on the demand curve. There are buyers who are willing to wait and buy the much cheaper paperback later and place a smaller value on early reading of the book. The benefit for the publisher lies in the fact that the higher price charged for the hardback is proportionately much greater than the higher cost of publishing it. In other words, the publisher gains greater profit per book for the hardback than for the paperback by extracting consumer surplus from the high-value consumers.

The same principle applies to laser printers. Consider two type of users – business and household customers. The business customers are willing to pay a much higher price for speed than household customers. If IBM charge all customers a high price for LaserPrinter, they cannot sell it to household customers. Only if they can charge two different prices for the two different groups! How? They found a solution to this pricing problem – inserting chips into LaserPrinter to slow down its printing speed, branding the slower version as LaserPrinter E and selling it at half of the price of LaserPrinter.

IKEA also uses price discrimination to set its delivery charges. Usually a furniture delivery company will charge customers based on distance travelled, that is, customers who live further from the store would pay higher delivery prices. In the IKEA example, all customers pay the same delivery charge for the same amount of purchase as long as they live within the metro area. *Effectively*, IKEA is practising price discrimination – setting a lower delivery charge to customers living further away from IKEA store on a per kilometre basis. In so doing, it can attract more distant customers to the store by charging them a lower delivery price, and at the same time without lowering its delivery charge for nearby customers.

If you said that each of the pricing practices discussed above involves price discrimination that often enables firms to increase their profits, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 A monopolist sets the:
- **a** highest possible price on its demand curve.
- **b** price that maximises total revenue.
- **c** price at which marginal revenue equals price.
- **d** price at which marginal revenue equals marginal cost.

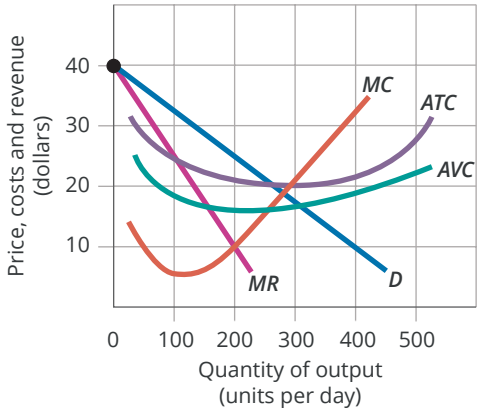
2 To maximise profit or minimise loss, the monopolist in **Exhibit 8.6** should set its price at:
- **a** $30 per unit.
- **b** $25 per unit.
- **c** $20 per unit.
- **d** $10 per unit.
- **e** $40 per unit.

3 As shown in **Exhibit 8.6**, the profit-maximising or loss-minimising output for this monopolist is:
- **a** 100 units per day.
- **b** 200 units per day.
- **c** 300 units per day.
- **d** 400 units per day.

4 As shown in **Exhibit 8.6**, this monopolist:
- **a** earns positive economic profit.
- **b** should shut down in the long run.
- **c** earns zero economic profit.
- **d** should shut down in the short run.

5 Which of the following statements is true?
- **a** All monopolies are created by the government.
- **b** The monopolist never takes a loss.
- **c** *Ceteris paribus*, a monopolist charges the same price as a competitive firm.
- **d** None of the above statements is true.

Exhibit 8.6 **Profit maximising for a monopolist**



6 It is a characteristic of a natural monopoly that:
- **a** the long-run average total cost curve continues to decline as its output rises.
- **b** the short-run average total cost curve continues to decline as the firm expands its output.
- **c** the long-run average total cost curve is constant throughout the full range of its output.
- **d** the long-run average total cost curve increases as the firm expands its output.
- **e** both a and b above are correct.

7 A monopolist:
- **a** will never shut down because it always earns economic profits.
- **b** will shut down if price is below average total cost.
- **c** will shut down if its demand curve lies wholly below the average total cost curve.
- **d** would shut down if its demand curve lies wholly below the average variable cost curve.

8 Compared to a perfectly competitive industry, other things being equal, which of the following is true about a monopoly?
- **a** Both consumer surplus and producer surplus are larger in monopoly.
- **b** Both consumer surplus and social surplus are smaller in monopoly.
- **c** Consumer surplus is lower but producer surplus is larger in monopoly.
- **d** Both b and c are true.

9 Which of the following involves price discrimination?
- **a** A department store has a '25 per cent off' sale.
- **b** A wine wholesaler in Adelaide charges less for its products in Bourke (in the far west of New South Wales) than in Adelaide.
- **c** Manufacturers of cars in South Korea sell these cars at higher prices in Australia than in South Korea.
- **d** Your phone plan has higher charges for international calls than for domestic calls.

Monopolistic competition and oligopoly

9

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Describe the characteristics of the monopolistic competition market structure.
2 Know why we see the monopolistically competitive firm as a price maker.
3 Understand price and output decisions for a monopolistically competitive firm.
4 Compare monopolistic competition and perfect competition.
5 Describe the characteristics of the oligopoly market structure.
6 Apply the game theory to examine strategic decision-making for an oligopolist.

With increasing health consciousness, the health snack food industry in Australia has experienced strong annual growth of 2.7 per cent over the five years through to 2020.[1] Further boosted by the rise of disposable income, consumers are willing to pay premium prices for these health snacks; profit margins are high. Because it is relatively easy to set up to cater for the niche segments of market, profit expectation has therefore attracted many small-scale operators into this industry over this period, producing health snacks ranging from protein bars to natural health drinks. Due to ease of entry, there are no dominant players in this industry.

One of the brands that you like most is Vitality, specialised in dried nuts and fruits, distributed through supermarkets and health food stores and is currently enjoying good profits. The company Vitality Ltd does not fit either of the two extreme models studied in the previous two chapters. Instead, Vitality's characteristics are a blend of monopoly and perfect competition. For starters, similar to a monopolist, the demand curve Vitality faces is downward-sloping. This means Vitality is a *price maker* because it can consider charging a higher price, which means that although it would lose some customers, many loyal customers would keep returning. The reason is that Vitality's products are different from those of its competitors. It makes its products different from its competitors' by using advertising and by combining different ingredients to create distinct tastes and nutritional values. In short, like a monopolist, Vitality has a degree of *market power*. But like a competitive firm, it faces a lot of competition and must share the market with many other health snack food producers. Will health snack food producers like Vitality continue to enjoy profits in the long run in this keenly contested market?

iStock.com/RosetteJordaan

[1] Ibisworld, 'Industry at a glance', 2020, https://my.ibisworld.com/au/en/industry-specialized/od5486/industry-at-a-glance

While Vitality operates in a highly competitive environment, the small number of fuel companies (e.g., BP and Caltex) have much larger market power, dominating the petrol retail industry. Vitality and BP compete in different market structures, both of which fall between perfect competition and monopoly. One of these market structures is *monopolistic competition* and the other is *oligopoly*. Vitality operates in the former, while BP belongs to the latter. This chapter will look at the behaviour and performance of firms in these two market structures, and compare them with perfect competition and monopoly presented in the previous two chapters.

1 The monopolistic competition market structure

In the previous two chapters we have looked at the market structures at each end of the continuum presented in **Exhibit 7.1**. In this section we look at monopolistic competition, which is closest to perfect competition.

With the model of perfect competition being seen as a somewhat unrealistic portrayal of small firms in real-world markets, there commenced in the 1920s and 1930s a movement to develop a new, more realistic model. The most prominent contributors to the model were Joan Robinson from Cambridge University in Cambridge, England, and Edward Chamberlin from Harvard University in Cambridge, Massachusetts, US. In 1933, these authors published different books dealing with what they both described as the economics of imperfect competition. It is their economics of imperfect competition that is today described by the model of monopolistic competition.

Monopolistic competition A market structure characterised by (1) many small sellers, (2) a differentiated product and (3) easy market entry and exit.

Economists define **monopolistic competition** as a market structure characterised by (1) many small sellers, (2) a differentiated product and (3) easy market entry and exit. Monopolistic competition fits numerous real-world industries. In fact, many of the businesses you deal with on a regular basis – the local florist, your favourite restaurant, the phone-case retailer and the bakery – are all monopolistic competitors. Monopolistic competition is the market structure in which we find more firms than in any other structure. Let's look at each of the characteristics of monopolistic competition in turn.

Many small sellers

Under monopolistic competition, as under perfect competition, the existence of a large number of firms means that each firm has a very small market share and no single firm can influence the market outcome. Vitality, described in the chapter introduction, is an example of a monopolistic competitor. Although Vitality faces keen competition in the health snack food market, it can set prices slightly higher than those of rival brands without fear of losing all its customers, as would occur if it was a perfect competitor. As you will see, it is able to do this because it sells a differentiated product.

Differentiated product

Product differentiation The process of creating real or apparent differences between goods and services.

The key feature of monopolistic competition is **product differentiation**. Product differentiation is the process of creating real or apparent differences in the product. A differentiated product means there are close, but not perfect, substitutes for the firm's product. While the products of each firm are very similar, the consumer views them as somewhat different. There may be numerous small operators producing health snack food in a given market, but they are not all the same. They differ in type, taste, nutritional value and quality of products sold.

Product differentiation can be real or imagined, but it does not matter if it is only imagined, so long as consumers believe such differences exist. For example, many customers think Vitality makes the best quality organic fruits and nuts bar in town even though other shops offer similar products. The importance of this consumer viewpoint is that Vitality can increase prices if it wishes because many of its customers will be willing to pay a slightly higher price to buy from them. The greater the perceived superiority of Vitality products, the

greater will be its capacity to raise its prices without significant loss of customers. This gives Vitality the incentive to try to differentiate its products and services further by, for example, undertaking expenditure on packaging and by developing a marketing campaign using photos of their staff tending herbs in a garden.

The example of Vitality makes it clear that, under monopolistic competition, rivalry centres not only on price competition but also on **non-price competition**. Non-price competition is the situation in which a firm competes using advertising, packaging, product development, quality and service, rather than lower prices. Non-price competition leads to important differences between monopolistic competition, perfect competition and monopoly. Under perfect competition, there is no non-price competition because the product is identical for all firms. Likewise, the monopolist may have little incentive to engage in non-price competition because it sells a unique product.

Non-price competition The situation in which a firm competes using differences in advertising, packaging, product development, quality and service, rather than lower prices.

Easy entry and exit

Unlike a monopoly, firms in a monopolistically competitive market confront low barriers to entry. Someone who wants to enter the health snack food industry can get a loan, lease a workshop and start producing health snack food without too much trouble. In monopolistic competition markets are very easy to enter. Just consider how anyone with a small amount of capital, the willingness to work hard and a modicum of common sense can start up a business, such as a bike shop, a lawn care service, a plant shop or a home maintenance service.

In summary

- **Monopolistic competition** is a market structure characterised by (1) many small sellers, (2) a differentiated product and (3) easy market entry and exit.

2 The monopolistically competitive firm as a price maker

Whereas the perfectly competitive firm is a price taker, the monopolistic competitor is a *price maker*. The primary reason is that its product is differentiated. This gives the monopolistically competitive firm some control over its price. If it decides to raise its price it finds that, rather than losing all its customers as would be the case under perfect competition, a degree of loyalty to its product or service means that some customers will remain steadfast. On the other hand, a decision by a monopolistic competitor to lower its price will result in an increase in sales. This contrasts with the perfectly competitive case where the firm can sell as much as it likes at the going market price, and thus has no incentive to reduce price. This control over price means that, as for a monopolist, the demand curve for a monopolistically competitive firm are downward-sloping. However, the existence of close substitutes causes the demand curve for the monopolistically competitive firm to be more elastic than the demand curve for a monopolist.

Advertising pros and cons

Before presenting the complete graphical models for monopolistic competition, let's pause to examine further the topic of advertising. As explained at the beginning of this chapter, a distinguishing feature of a monopolistically competitive firm is that it engages in non-price competition, which may involve the use of advertising to differentiate its product. The desirability of this advertising is a subject that has exercised the minds of economists for decades.

Recall production is a value creating process. Critics of advertising argue that the main purpose of advertising is to persuade or mislead consumers into buying something they do not need, without creating any value in the process. From society's viewpoint, the resources used in advertising could be used for schools, hospitals, food, clothing or other useful products and services. On the other hand, proponents of advertising cite its many benefits. Those on this side of the debate see ads as infusing the product with its characteristics (luxury cars for sophisticated people, off-road vehicles for adventurous types) and providing worthwhile information. Advertising informs consumers of the target market for the product and the availability of different products, as well as their advantages. So while advertising means that the product may cost a little more, this information saves consumers money and time. By providing more information on availability, price and quality, ads may also increase price competition among sellers. Finally, protagonists of advertising argue that consumers are rational and cannot be fooled by advertising. If a product is undesirable, customers will not buy it. By contrast critics of advertising argue that it is so powerful it can lead consumers to make purchase decision that causes harm to themselves. Perhaps the most obvious example of this view is found in the continual calls for the promotion and advertising of junk food to be controlled in an endeavour to fight the obesity epidemic.

As long as there is advertising there will be continuing debate about its value to society. Nevertheless, a measure of the concern that society has about the ill-effects of misleading advertising can be found in the regularity with which the Australian Competition and Consumer Commission (ACCC) imposes severe penalties on firms that contravene the *Competition and Consumer Act 2010* by engaging in false or misleading advertising. Later in this chapter you will learn that advertising to differentiate a product is also a key characteristic of many firms in the oligopoly market structure.

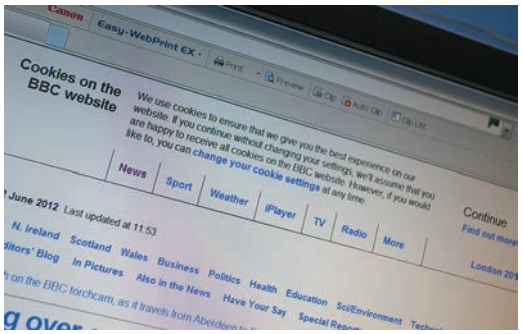
Global perspective

Applicable concepts: product differentiation and ethics of advertising

Social networking sites: becoming the dominant medium for advertising

Traditional media, such as newspapers and television, are struggling to even maintain advertising revenue as consumers turn to the internet for news and entertainment. On the other hand, social networking sites are becoming an increasingly important means by which advertisers reach prospective customers. They do this by tapping into the popularity of sites such as Facebook, LinkedIn and Twitter. As you know, sites such as these enable individuals to interact globally using a variety of means including sharing of personal profiles, games and the posting of photos and videos.

The challenge for entrepreneurs who initially develop these sites is to earn profits by both increasing their popularity and by differentiating their product to create innovative ways of attracting advertisers. Search engines have become a highly successful business model. If someone searches for 'what's on at the movies', sponsored ads for local movie theatres appear on the screen. Now suppose a social networking site offers an international movie distributor the opportunity to pay for a page where you and your friends can play games and complete quizzes based on the latest blockbuster movie. If it has played its cards right, not only will the distributor have created a heightened awareness of the new movie but it will also have a database containing the identities of thousands of potential customers who can be alerted to new movies when they are about to be released. This differentiation of its site has helped the social network entrepreneur to increase its profits as well as those of the movie distributor.



Alamy Stock Photo/Jeffrey Blackler

However, for what are often ethical reasons, not all ideas are winners. In the early days Facebook implemented a system that informed friends whenever a member purchased something from online retailers. Consumers complained that this was an invasion of privacy and this element of product differentiation was discontinued. Do you sometimes tear your hair out when you can't skip an ad that precedes a video or when you are distracted from your search by images running down or across the screen? Do these approaches to product differentiation sometimes backfire when you avoid sites with these annoying characteristics? In addition, not all advertisers will be drawn to social networking sites. In 2012, General Motors in the US decided to cancel its Facebook advertising because it thought it had little effect on car purchases.

What do you think?

As part of their drive to differentiate their product, internet companies use a range of ways of exploring the characteristics of users. These include the use of cookies, which can reveal every keystroke you use on your computer while on certain sites. What are the ethical issues that these practices raise? Do you think most users of the internet are aware of just how much information about them can be gathered by companies?

In summary

- The monopolistically competitive firm is a price maker. The demand curve it faces is less elastic than for a perfectly competitive firm and more elastic than for a monopolist.

3 Price and output decisions for a monopolistically competitive firm

Now we are prepared to analyse the short-run and long-run market outcomes for monopolistic competition. The short-run analysis resembles that of monopoly. In the long run, however, entry by new firms leads to market outcomes similar to a competitive market.

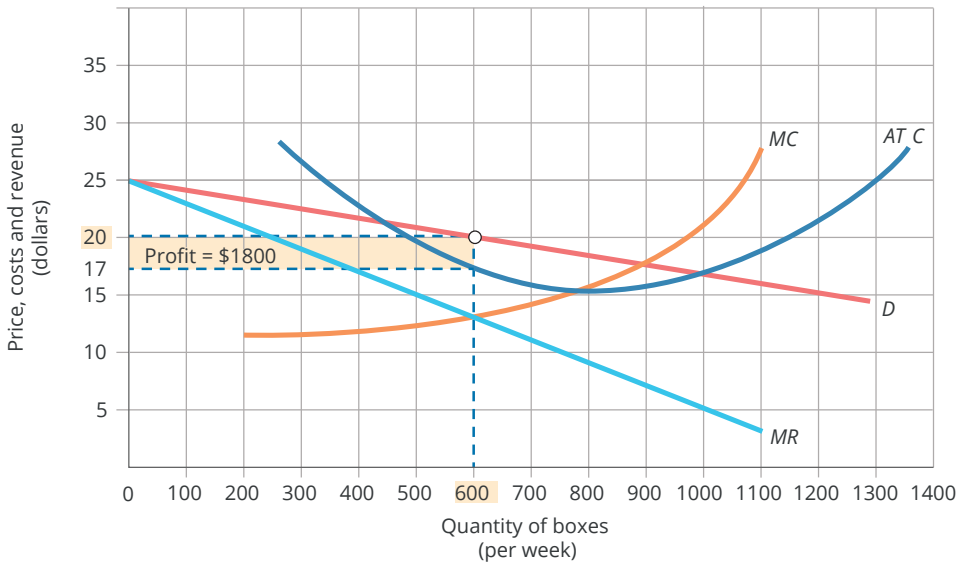
Monopolistic competition in the short run

Exhibit 9.1 shows the short-run equilibrium position for Vitality, a typical firm under monopolistic competition. As explained earlier, the demand curve faced by the firm slopes downward because customers perceive that Vitality's products are different from its competitors'. Vitality's customers are attracted by the particular nutritional value and unique taste of its products. These non-price factors differentiate Vitality's products and allow them to set their own price rather than 'take' the price that would be determined in the market if it were perfectly competitive. In the simple analysis that follows, we assume that Vitality sells only one product – boxes of protein bars. **Exhibit 9.1** presents Vitality's demand curve and the associated marginal revenue curve for its protein bars, and its costs curves.

Like firms in any market structure, the monopolistically competitive firm maximises short-run profit by following the $MR = MC$ rule. In this case, the marginal cost and the marginal revenue curves intersect at an output of 600 boxes per week. The profit-maximising price per box of $20 is given by the point on the demand curve corresponding to this level of output. This price exceeds the ATC of $17 per box, giving an economic profit of $3 per box. Vitality thus earns a short-run weekly economic profit of $1800.

You can see that this short-run outcome, where price is in excess of marginal cost, looks just like that of the monopolist.

Exhibit 9.1 **A monopolistically competitive firm in the short run**



Vitality is a monopolistically competitive firm that maximises short-run profit by producing the output where marginal revenue equals marginal cost. At the profit-maximising output of 600 boxes sold per week, the price of $20 per box is determined by its demand curve. Given the firm's costs, output and prices, Vitality will earn a weekly short-run profit of $1800.

As is also the case under monopoly and perfect competition, if the price equals *ATC*, the firm earns a short-run normal profit; if the price is below the *ATC* curve, the firm suffers a short-run loss; and if the price is below the *AVC* curve, the firm shuts down. In the next section we will turn to the long run, which is just like perfect competition where ease of entry into the market ensures that short-run economic profits are competed away.

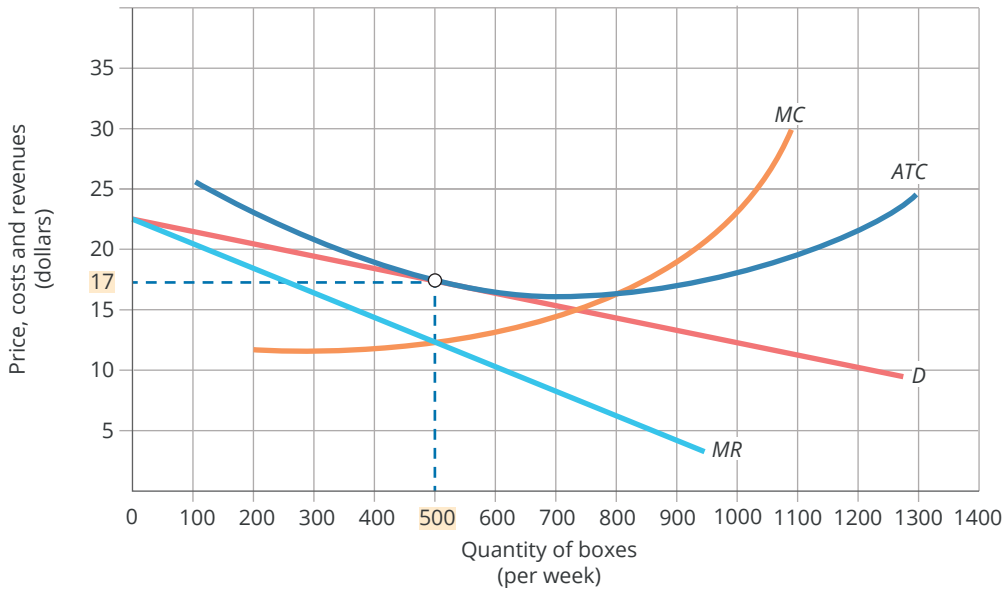
Monopolistic competition in the long run

Unlike a monopolist, the monopolistically competitive firm will not earn an economic profit in the long run. Rather, like a perfectly competitive firm, the monopolistically competitive firm earns only zero economic profit (or only normal profit) in the long run. The reason is that due to easy entry, short-run profits attract new firms into the industry. When Vitality earns a short-run profit, the entry of new firms will shift Vitality's demand curve to the left as some of its market share is taken away by new firms seeking profit.

The leftward shift in the firm's demand curve continues in the long run until the firm earns zero economic profit, which occurs when the demand curve is tangent to the *ATC* curve. **Exhibit 9.2** shows the long-run equilibrium condition for Vitality. Note that at the output of 500 boxes per week where the demand curve is tangent to the ATC curve, $MR = MC$.[2] Recall that $MR = MC$ is the profit maximisation condition for a firm. This means that at the profit maximisation output, $P = ATC$ and the firm is earning zero economic profit. Once long-run equilibrium of zero economic profit is achieved in a monopolistically competitive industry, there is no incentive for new firms to enter. You can see that this outcome (zero economic profit) is similar to the long-run outcome for a firm in perfect competition.

[2] It is no coincidence that when the demand curve is tangent to the *ATC* curve, $MR = MC$. This can be proved mathematically.

Exhibit 9.2 **A monopolistically competitive firm in the long run**



In the long run, profitability of existing firms encourages the entry of new firms, which decreases the demand for Vitality's products shifting its demand curve to the left. In the long run, Vitality earns zero economic profit at a price of $17 per box and produces an *MR* = *MC* output of 500 boxes per week.

Facing keen competition in the long run, Vitality, as well as other brands, may try to recapture market share by advertising, improving its quality, and using other forms of non-price competition. If this does happen, these firms may succeed in moderating the decrease in demand for their product; but the flip side is that they will see their average costs increase. In the above analysis, we ignore the possibility of firms engaging in increased non-price competition in these circumstances so that we can concentrate on the effects of increased competition on the demand curve alone. Therefore, in a dynamic market a monopolistic competitor firm like Vitality may continue to earn economic profit over time by becoming more efficient and innovative in producing its output.

You should now be able to analyse the long-run outcome for monopolistic competitors incurring short-run losses rather than profits. See if you can describe the shift in the demand curve that will occur when firms exit the industry in the long run, and the consequent change in price. One thing you can be sure of is that in long-run equilibrium, the firms remaining in the industry will be earning zero economic profit.

In summary

- Short-run equilibrium for a monopolistic competitor can yield *economic losses*, *zero economic profits* or *economic profits*. In the long run, monopolistic competitors make zero economic profits.

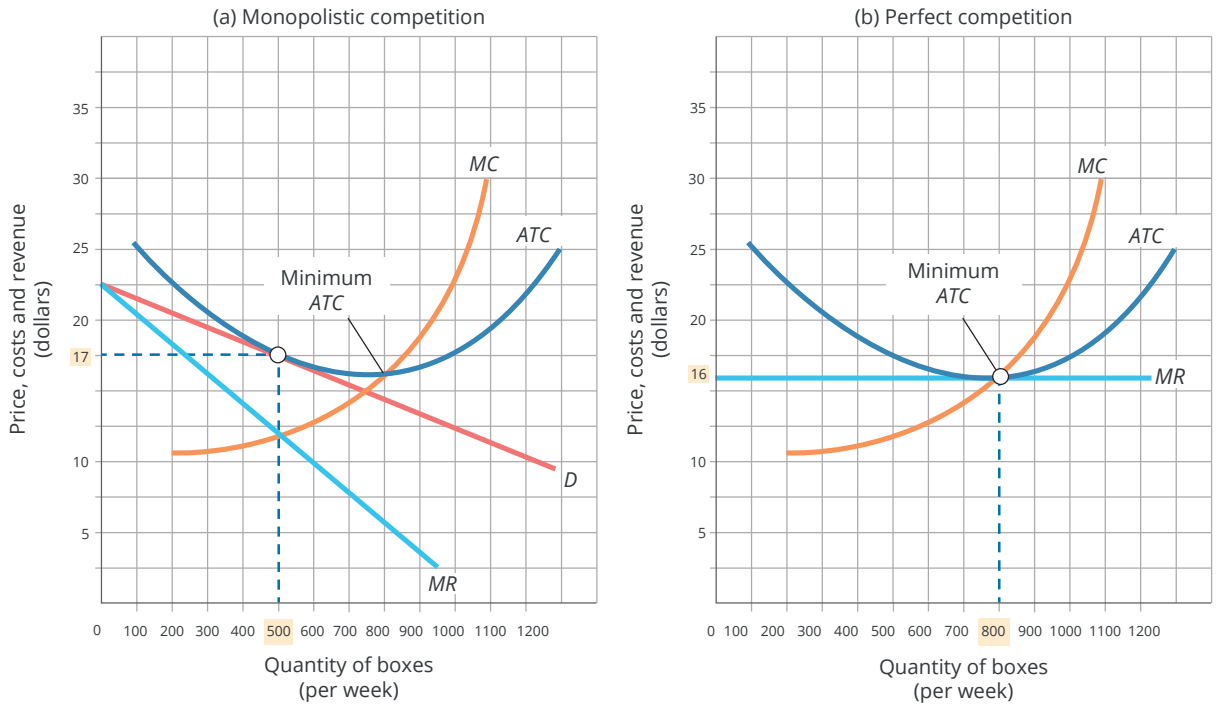
4 Comparing monopolistic competition and perfect competition

Some economists argue that the long-run equilibrium state for a monopolistically competitive firm, as shown in **Exhibit 9.2**, results in sub-optimal economic performance. Other economists contend that the benefits of a monopolistically competitive industry outweigh the costs. In this section, we again use the perfect competition model to help us understand both sides of the debate.

Is the monopolistic competitor a resource misallocator?

As in the case of monopoly, the monopolistically competitive firm fails the standard efficiency test. **Exhibit 9.3(a)** reproduces the long-run condition from **Exhibit 9.2** for comparison. In **Exhibit 9.3(b)**, it is assumed that the market for health snack food is perfectly competitive. For ease of exposition, it is also assumed that Vitality's costs remain unchanged when the market structure changes to become perfectly competitive. Let us first consider a competitive health snack food market. Because Vitality is a perfectly competitive firm, its product is identical to all other firms' in the industry. Vitality becomes a *price taker*, facing a horizontal demand curve with its marginal revenue equal to the market price. In the long run, the market forces (entry and exit of firms) will set the equilibrium price at the minimum *ATC* of $16 per box. At this long-run equilibrium price Vitality produces 800 boxes per week, earning zero economic profit.

Exhibit 9.3 **A comparison of monopolistic competition and perfect competition in the long run**



In part (a), Vitality is a monopolistically competitive firm that sets its price at $17 per box and sells 500 boxes per week. As a monopolistic competitor, Vitality earns zero economic profit in the long run *but* does not produce at the lowest point on its *ATC* curve.

Under conditions of perfect competition in part (b), Vitality becomes a price taker rather than a price maker. Here the firm faces a horizontal demand curve at the long-run equilibrium price of $16 per box. The output is 800 boxes per week, which corresponds to the lowest point on the *ATC* curve. Therefore, under prefect competition the price is lower and the output is higher than when Vitality operates as a monopolistically competitive firm.

A comparison of parts (a) and (b) of **Exhibit 9.3** reveals several important points. First, both the monopolistic competitor and the perfect competitor earn zero economic profit in the long run. Second, the long-run equilibrium output of the monopolistically competitive firm is to the left of the minimum point on the *ATC* curve. Like a monopolist, the monopolistically competitive firm charges a higher price and produces less output than a perfectly competitive firm. Moreover, at the lower level of output of 500 units per week, $MU > MC$ and the social surplus is not maximised.

The criticism of monopolistic competition, then, is that firms produce too little output at a higher price and cost, thus wasting society's resources in the process. With perfect competition, each firm would produce a greater output at a lower price and with a lower average cost.

Although monopolistic competition fails the standard efficiency test, it does have its benefits. These benefits arise from the greater consumer choice available under monopolistic competition compared with perfect competition. If there are many health snack food producers offering differentiated products, this gives consumers much more choice than if the producers were all the same. If you do not like Vitality's range of products, you may be able to find a better brand. Then again, other buyers may find that the products supplied by Vitality's competitors are more to their liking. Most of us appreciate being able to exercise choice in this way.

All things considered, we can be pretty sure that in the case of most goods and services, consumers would prefer the variety that comes with product differentiation in monopolistically competitive markets rather than be faced with no choice in a perfectly competitive market in which all firms sell identical products. Although product differentiation means lower output and higher prices, it would seem that the benefits of variety may outweigh the benefit from lower prices that would result from production of identical goods and services in a perfectly competitive market. It seems that governments see it this way too. Can you imagine a competition policy that proposed penalties for small firms that attempted to use product differentiation and advertising to make their products or services more attractive to consumers?

Lastly, as discussed in **Chapter 8** with monopoly, the monopolistically competitive firm can, because it faces a downward-sloping demand curve, try to practise price discrimination. In order to do this it must also fulfil the three conditions for price discrimination. Examples of price discrimination by monopolistic competitors include hairdressers that charge lower prices for seniors, and bike shops that charge children lower prices to service their bikes.

In summary

- Comparing monopolistic competition with perfect competition, we find that the monopolistically competitive firm does not achieve allocative efficiency.
- To the extent that the benefit of product differentiation outweighs the cost of higher price and cost, a monopolistic competitor may not be a resource misallocator.

5 The oligopoly market structure

Now we turn to oligopoly, a market structure in which a few large firms dominate the market. Many industries, such as steel, aluminium, motor vehicle manufacturing, airlines, petrol retail, banking, insurance, pharmaceuticals and tobacco, are best described as oligopolistic. This is the 'big business' market structure, in which firms often aggressively compete by using forms of non-price competition, such as costly marketing campaigns and image-building exercises.

Oligopoly
A market structure characterised by (1) few sellers, (2) either a homogeneous or a differentiated product and (3) barriers to market entry.

Economists define an **oligopoly** as a market characterised by (1) few sellers, (2) either a homogeneous or a differentiated product and (3) considerable barriers to market entry. Since the number of oligopolistic firms in the economy is relatively small, between them they have a large share of total output produced in the economy. Let's examine each of the characteristics of oligopoly in turn.

Few sellers

Oligopoly is competition 'among the few'. We use the terms 'Big Three' or 'Big Four' to mean that three or four firms dominate an industry. Economists measure the extent to which the largest firms in an industry dominate the market by using a measure called **concentration ratio**. This measure indicates the percentage of total sales in the industry that are generated by its largest firms. In Australia during 2020, the two largest grocery chains, Coles and Woolworths, had a combined market share of nearly 60 per cent, the largest four petrol retailers supplied over 70 per cent of the market, and the big four banks had 75 per cent of the market for small businesses and household loans.

Concentration ratio
A measure that indicates the percentage of total sales in the industry generated by its largest firms.

Because there are only a few firms in the market, when one firm attempts to increase output by lowering its price, it will have a big impact on the sales and profits of its competitors. The competitors may choose to compete or not. The outcome of any pricing decision by an oligopolist will therefore depend on how its competitors react to the price change. As a result an important feature of firms in an oligopoly market structure is that they tend to be mutually interdependent. **Mutual interdependence** is a condition in which an action by one firm may cause a reaction on the part of other firms. When Virgin considers lowering its fares or offering services on new routes, it must anticipate how Jetstar will change their fares or routes. Therefore, the decisions under oligopoly are more complex than under perfect competition, monopoly or monopolistic competition.

Mutual interdependence
A condition in which an action by one firm may cause a reaction on the part of other firms.

Homogeneous or differentiated product

Under oligopoly, firms can produce either a homogeneous or a differentiated product or service. The aluminium ingots produced by Alcoa are virtually identical to the ingots manufactured by the Aluminium Corporation of China. The oil sold by Saudi Arabia is very similar to the oil from Iran. Similarly, raw zinc, copper and steel are standardised products. But banking services produced by the big four Australian banks are differentiated products. Domestic air travel, soft drinks, tobacco products and insurance are also differentiated services or products sold by the oligopoly firms. Not only does product differentiation enable firms to gain loyalty from consumers but it can also be used as a medium for the execution of price discrimination. For example, banks have lower fee structures for students and pensioners, and supermarket chains charge different prices for the same product to different customers using personal promotion or discount offers via emails. If you use your supermarket loyalty card when shopping groceries, you must have received those promotion and special offer emails.

Barriers to entry

Similar to monopoly, formidable barriers to entry in an oligopoly market protect firms from new entrants. These barriers include very large financial requirements to enter the market, control of an essential resource by existing firms, patent rights held by existing firms, and other legal barriers, such as licences required by operators of banks and casinos. But the most significant barrier to entry in an oligopoly market structure is *economies of scale*. For example, larger car manufacturers achieve lower average total costs than those incurred by smaller firms. Consequently, the world motor car manufacturing industry has moved from having hundreds of firms prior to the Second World War to having no more than 20 significant players today.

Which model fits the breakfast cereal market?

You're the economist

You love cereals for breakfast. Do you have a big range of cereals to choose from for your breakfast each day? Some of the cereals you may have at home include Corn Flakes, Cheerio's, Nutrigrain, Rice Bubbles, Coco Pops and Vita Brits to name but a few. There are many different brands of the same product – breakfast cereal – available at the supermarket. Each brand is clearly differentiated from the others. Is the breakfast cereal a monopolistic competition or oligopoly industry?

Getty Images/Universal Images Group/ Jeff Greenberg

In summary

- **Oligopoly** is a market structure characterised by (1) few sellers, (2) a homogeneous or a differentiated product and (3) significant barriers to entry. When introducing new products or deciding on pricing strategies, oligopolies are **mutually interdependent**.

6 Price and output decisions for an oligopolist

Mutual interdependence among firms in an oligopoly makes this market structure more difficult to analyse than perfect competition, monopoly or monopolistic competition. Nevertheless, we can be sure that the market power that oligopolists derive from significant barriers to entry means that, like the monopolist, they can earn economic profits in the long run. However, the price–output decision of an oligopolist is not simply a matter of charging the price where $MR = MC$. Making price and output decisions in an oligopoly is like playing a game of chess game in which one player's move depends on the anticipated reactions of the opposing player. Indeed, the behaviour of oligopolists is so like that of players of a game that this market structure is often analysed by a specialised subdiscipline called **game theory**. As you will see shortly, this theory has evolved to help analyse and explain how players make strategic production and pricing decision.

Game theory
The study of strategic decision-making in situations where the players are inter-dependent.

Game theory

Let us look at a simple scenario. You and your competitor are the only suppliers in the market, a situation known as **duopoly**, facing two possible market outcomes: (1) both you and your competitor make $8 million of profits, and (2) both make $16 million of profits. Duopoly is an oligopoly market structure characterised by there being just two sellers. Being profit maximiser you prefer more profit to less and therefore would pick outcome (2). So will your competitor. If both you and your competitor prefer outcome (2) to (1), will outcome (2) occur? Before we examine this scenario in more detail, we will present the famous *prisoner's dilemma* created by Albert Tucker to illustrate strategic decision-making that can lead to logical but seemingly counter-intuitive outcome.

Duopoly
An oligopoly market structure characterised by there being just two sellers.

> Two criminals, David and Peter, caught by the police during a routine road-side operation, are suspected of committing a serious bank robbery. They did, in fact, commit the crime. If convicted, they will face long jail sentences of 5 years. The prosecutor, however, has only enough hard evidence to convict them of a minor offense – possession of a firearm, for which the penalty is one year in jail.

Each criminal is held in separate cell and told that they can confess or deny the bank robbery charge. If one cooperates and confesses and the other denies, the one who confesses will be granted the status of accomplice witness and set free, and his partner will receive the more severe penalty of 20 years imprisonment. If both confess, they will get a plea bargain agreement, and each will get an intermediate sentence of five years.

This scenario typifies a game of strategic decision-making. In game theory the two criminals are called *players*. Each player has to decide on an action – a *strategy*, by considering the other's action. The *playoff* of the game – the outcome a player receives from playing the game – is presented in **Exhibit 9.4**. For example, if Peter denies and David confesses, David will be set free and Peter will spend 20 years in jail.

Exhibit 9.4 **Playoffs matrix for the prisoner's dilemma**

		Peter	
		Confess	Deny
David	Confess	(5, 5)	(0, 20)
	Deny	(20, 0)	(1, 1)

As you can see it is in their joint interest for both to deny, in which case the prosecutor can only charge them with a lesser crime for which the penalty is one year in jail. But to protect their self-interest, both Peter and David have a dominant strategy of Confess. Therefore, (Confess, Confess) is the Nash equilibrium. Both end up spending five years in jail.

Given the two individuals cannot communicate with each other, how will they play the game? Let us consider David. If Peter denies, David clearly should confess and will go free. On the other hand, if Peter confesses, David should also confess in order to get a lesser punishment of 5 years imprisonment. Therefore, in this game David has a **dominant strategy** of confessing. That is, he would always confess whether Peter confesses or denies. A dominant strategy is an action a player will always take regardless of the other player's action.

Dominant strategy
An action the player will always take in game theory regardless of the other player's action.

It is easy to see that from Peter's point of view, he also has a dominant strategy, which is Confess. The equilibrium outcome of this game is both will confess and get a sentence of five years. This outcome (Confess, Confess) is known as **Nash equilibrium**, named after the mathematician John Nash, who received a Nobel Prize in economics in 1994 for his contribution in non-cooperative games. In game theory the Nash equilibrium is the combination of strategies where neither player has any incentive to change strategies given the strategy of his competitor. The outcome (Confess, Confess) is an equilibrium outcome. It is because given Peter's chosen strategy to confess, David will put himself into a worse situation of 20 years imprisonment if he changes his strategy from Confess to Deny. Therefore, David will have no incentive to change his strategy. The same can be said of Peter.

Nash equilibrium
The combination of strategies in a game where neither player has any incentive to change strategies given the strategy of his competitor.

You may notice that the combination of strategies that will advance their joint interest is for both players to Deny, in which case they will each get a sentence of one year. However, to protect their self-interest, both will *strategically* choose Confess, each ending up spending five years in jail. This game illustrates the dilemma facing the 'prisoners', arising out of the conflict between joint interest and self-interest.

Let us now look at a real-world prisoner's dilemma situation. Consider a small town in regional Victoria in Australia where BP and Caltex are the only petrol retailers. In the real world, when deciding on their pricing policy, firms have a range of prices to choose from. For simplicity let us assume that BP and Caltex are faced with only two options: a High or Low price. **Exhibit 9.5** presents the playoff of their decisions, which is known to both firms. It shows that if they cooperate and both charge a high price, they act together as a monopolist.

Since demand for petrol is inelastic, they will make a big joint profit of $32 million if they cooperate. Each enjoys a profit of $16 million per year, assuming the duopolists have equal market shares. However, if they compete and both charge a low price, their profits will be reduced to $8 million a year each. BP petrol is a close substitute of Caltex petrol. If one petrol retailer charges a high price whereas its competitor charges a low price, the retailer charging a low price will take away most of the customers and make a large profit of $20 million, and the one charging a high price will make a much smaller profit of $2 million.

Exhibit 9.5 **Playoffs matrix for petrol duopoly**

		Caltex	
		High	Low
BP	High	(16, 16)	(2, 20)
	Low	(20, 2)	(8, 8)

If both Caltex and BP charge a high price, they behave like a monopoly and each earns a big profit of $16 million per year. On the contrary, if they compete by charging a low price, they will make a small profit of $8 million. If one firm charges a high price, it will lose customers to its low-price competitor and make a very small profit of 2 million, whereas its competitor will earn a big profit of $20 million.

In Australia and most developed countries, there are competition laws to prevent anti-competition behaviour, such as price fixing and collusion. In the petrol retail market, it is illegal for the two firms to communicate with each other and engaged in explicit collusive behaviour. Each firm makes their pricing decision independently, just like David and Peter in the prisoner's game. We can see BP and Caltex face the same dilemma as David and Peter. BP and Caltex can make a larger combined profit if both charge a high price. However, using the same thinking discussed above, both players have a dominant strategy of low price for self-interest. Hence, (Low, Low) will be the equilibrium outcome, where each firm earns a small profit of $8 million a year. As discussed, (Low, Low) is a Nash equilibrium because once settled in this equilibrium neither firm has any incentive to change their strategy. For example, will BP be benefitted by changing its price from Low to High? It won't because it will lose customers to Caltex and end up earning a smaller profit.

It is worth mentioning that a game can be played in many ways. The above illustrates the simplest type of game, known as one-shot simultaneous game. A game can be played sequentially whereby one player takes the first move, followed by the second player. In other situations, a game can be played multiple times, called a repeated game. Interested readers can consult texts on game theory. In the next section we apply game theory to some real-world situations.

Collusion and price leadership

The petrol retailers know it is not to their advantage to compete and start a price war. If they can communicate in some way without breaking the law, they would prefer to cooperate and charge a high petrol price to make more profits. There is a way. Petrol prices are not the same in different areas within a city. If you live in Perth, Australia, have you visited *FuelWatch*, an industry petrol pricing website for Western Australia, to find out where to get cheaper petrol?

While you may benefit from the petrol price information published in this website, do you know that the major petrol retailers and other small retailers also use this information for communicating with each other to coordinate price movements, engaging in **tacit collusion**? Tacit collusion is an informal arrangement for firms to raise profits by fixing prices. To analyse how it works and the outcome of this price coordination, we will look at another phenomenon in the petrol retail market in Australia – the fuel price cycle.

Tacit collusion An informal arrangement for firms to fix prices.

You're the economist — Is price leadership illegal?

The *Competition and Consumer Act 2010*, which is administered by the ACCC, prohibits contracts, arrangements or understandings that have the purpose or effect of fixing, controlling or maintaining prices. Do you think that this provision means that the ACCC would consider taking action against a firm like a telco or a bank if it were to make a public announcement foreshadowing future pricing policy?

If you visit the ACCC's petrol price cycle website,[3] you will notice that petrol prices in capital cities in Australia follow cyclical movements. Different cities exhibit different cyclic patterns, some being more regular than others. For example, as of 2020, Perth has a regular weekly cycle whereas in Sydney, Melbourne and Brisbane the cycles have less regularity and longer durations. In Perth, the retail price for E10 peaks on Wednesday and gradually falls to a low on Tuesday. Then the price rises again to a high on Wednesday the following day, starting another price cycle. How are such concerted price movements between fuel companies coordinated? And what is the motivation behind it?

Price leadership
A pricing strategy in which a dominant firm sets the price for an industry and the other firms follow.

Without formal agreement, the petrol retailers can play a game of follow-the-leader that economists call **price leadership**, which is a pricing strategy in which a dominant firm sets the price for an industry and the other firms follow. It started in 2010 when BP, the dominant petrol retailer in Perth, first introduced the Thursday price jump.[4] It performed the role of the price leader and other competitors used the Thursday jump as the focal point to set their market prices. The result is reduced competition, higher overall fuel prices and more profits to fuel companies. Since then petrol retailers in Perth have basically eliminated price wars. Economists Leigh and Triggs estimated that for the period from 2013 to 2016 the fuel companies profit margins tripled and the consumers on average paid an extra 10 cents per litre for petrol.[5] This practice of price leadership continues today; the only difference is that the fuel price cycle peaks on Wednesday instead of Thursday.

You may wonder why the ACCC did not do anything about this situation. While price fixing is illegal, there is no law against high prices. It is difficult for the authority to prove firms collude to fix prices, and for years the ACCC have unsuccessfully brought price fixing charges against petrol retailers.[6]

Economics and ethics — ATM fees for 'foreign' transactions

Along with smaller banks, each of the Australian big four banks (ANZ, Commonwealth, NAB and Westpac) has a network of ATMs from which customers can, *inter alia*, withdraw cash. Historically these banks did not charge their own customers a fee for withdrawals from their own ATMs. However, each bank did levy a charge on customers of other banks who used their ATMs. In these instances, each of the big four banks charged a $2 fee for a so-called 'foreign' transaction on their ATM. This fee had been roundly criticised by consumer groups who said it represented unethical gouging on the part of banks whose cost of such a transaction was estimated by the Reserve Bank of Australia (Australia's central bank) to be 77 cents in 2014.[1] In other words, banks were marking up their cost by a massive 160 per cent.

1 See G. Lekakis, 'Banks break silence on "indefensible" ATM fees', *The New Daily*, 11 January 2015, http://thenewdaily.com.au/money/finance-news/2015/01/11/bendigo-bank-reviews-indefensible-atm-fees/, accessed 25 October 2017.

3 ACCC, 'Petrol price cycles', at https://www.accc.gov.au/consumers/petrol-diesel-lpg/petrol-price-cycles

4 'A few big firms: Concentrated markets and a lack of competition are damaging the Australian economy', *The Monthly*, 17 May 2017, https://www.themonthly.com.au/blog/andrew-leigh-and-adam-triggs/2017/17/2017/1495011536/few-big-firms

5 Ibid.

6 P. Coorey, 'ACCC boss Rod Sims says can't do much about petrol price', *Australian Financial Review*, 31 October 2018, https://www.afr.com/politics/accc-boss-rod-sims-says-cant-do-much-about-petrol-price-20181031-h17bcv

All this changed on 24 September 2017 when the Commonwealth Bank announced that it was scrapping the fee. It was well prepared for this move with the announcement being followed within hours by prominent internet advertising and by full-page ads in the press. Following this move by the Commonwealth Bank, the other three big banks dropped the fee within 24 hours.

Do you think the Commonwealth Bank's abolition of the ATM fee for 'foreign' transactions constituted price leadership? If the Commonwealth had increased its ATM fee do you think the other banks would have followed suit?

The cartel

The price leadership model we have discussed assumes that firms do not explicitly collude to avoid price competition. Instead, firms avoid price wars by informally playing by the established follow-the-leader pricing rules as in the petrol retail market. Another way to avoid price wars is for oligopolists to formally agree to act in concert. Instead of setting prices independently, firms openly agree with one another to cooperate to form a monopoly called a **cartel**. A cartel is a group of firms formally agreeing to control the price and the output of a product. The goal of a cartel is to reap monopoly profits by restricting output and raising prices. Although cartels are illegal in Australia and in most other developed countries, this does not mean that they do not exist covertly in these countries. Nonetheless, cartels are more common in some developing countries and also across international boundaries where there is no legal authority to regulate them. The best-known cartel is Organization of Petroleum Exporting Countries (OPEC) founded in early 1960s. The stagflation in the 1970s was a direct result of the three-fold increase in crude oil prices in early 1970s and another big hike in late 1970s, both initiated by OPEC by reducing cruel oil production. The members of OPEC divide crude oil output among themselves according to quotas openly agreed upon at meetings of the OPEC oil ministries. Saudi Arabia is the largest producer and has the largest quota. Let's see how these cartels work using a simplified scenario.

Cartel
A group of firms formally agreeing to control the price and the output of a product.

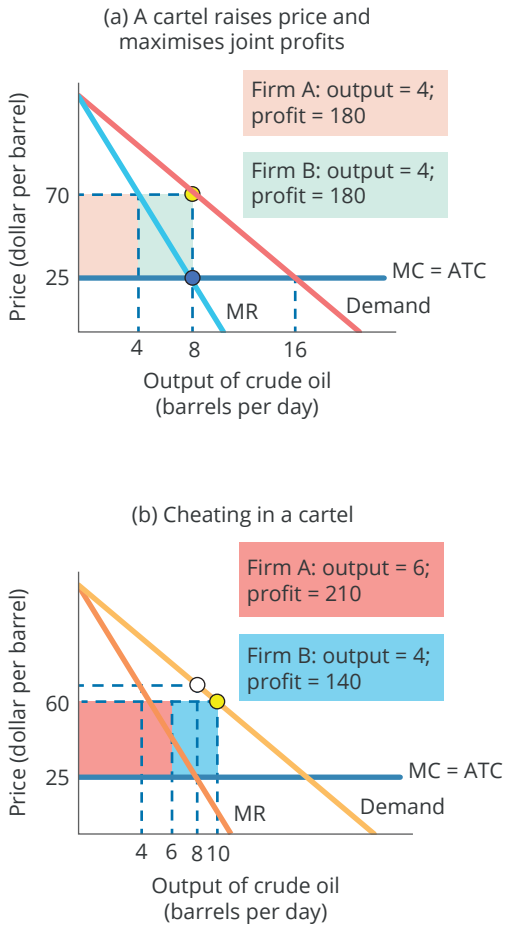
Using **Exhibit 9.6**, we can demonstrate how a cartel works and why cheating is a problem in a cartel. Our analysis begins before oil-producing countries have formed a cartel. To simplify calculations, we will assume that there are two firms and the marginal cost of oil production is constant with zero fixed cost. Given these assumptions, the *MC* and *ATC* curves are the same as shown in **Exhibit 9.6(a)**. If the oil industry is initially perfectly competitive, the market supply curve is the marginal cost (*MC*) curve. Competition will drive the market price to $25 per barrel where demand equals supply (*MC*). At the competitive equilibrium output of 16 million barrels per day, firms earn zero economic profit. There is thus a strong incentive for the firms to attempt to increase profitability by organising a meeting to establish a cartel to restrict output.

Now assume the cartel is formed. The members of the cartels will behave as a monopoly. As a monopoly, it will set a quota to reduce its total output to 8 million barrels per day where $MR = MC$, and charge $70 per barrel to maximise profit. Suppose the production capacities of the two firms are equal. They will agree on a quota of 4 million barrels per day each, which is half of the industry quota of 8 million. At the cartel price, each firm earns an economic profit of $180 million per day, a much better outcome than zero economic profit under perfect competition.

But what if one firm decides to cheat on the cartel agreement by stepping up its output while other firms stick to their quotas? Suppose Firm A secretly raises its output to 6 million barrels per day. The market output will go up to 14 million barrels per day, depressing the crude oil price to $60 per barrel. At a lower profit margin of $35, Firm B will suffer with a reduced profit of $140 million per day if it sticks to its agreed quota. But by expanding its output the cheating firm can increase its profit to $210 million per day ($35 x 6 million).

There is thus a strong incentive for cartel members to cheat. Was there evidence of cheating in the oil industry? Based on data collected on the set quotas and actual market output since late 1980s, CNN Business in 2016 reported that the *actual* OPEC oil output had been consistently above the OPEC quota output. This shows that some OPEC members *always* cheated in this period due to domestic political and economic expediencies.[7] Hence, the cartel is inherently unstable.

Exhibit 9.6 **Why a cartel member has an incentive to cheat**



Part (a) shows that if firms are operating in a perfectly competitive industry the market equilibrium output will be 16 million barrels per day at a price of $25 per barrel with each firm making zero economic profit. A cartel can agree to restrict output by setting a quota at 8 million barrels per day, raising the price of oil from $25 to $70 per barrel. As a result of this quota, each firm earns a daily profit of $180 million (= ($70 − 25) x 4 million). However, part (b) shows that if Firm A cheats on the cartel agreement, it will increase its output to 6 million barrels per day and earn a total profit of $210 million (= ($60 − 25) x 6 million).

[7] H. Long, 'Remember: OPEC almost always cheats', *CNN Business*, 12 December 2016, https://money.cnn.com/2016/12/12/investing/opec-oil-output/index.html

Applicable concepts: game theory, duopoly & interdependence **Analyse the issue**

Airlines play chicken

In both the prisoner's dilemma and petrol service station examples, there is one Nash equilibrium. However, you might have noticed that if the payoffs are altered, there may not be an equilibrium and in some cases more than one equilibrium. The game of chicken is an example of a game with multiple equilibria.

The game of chicken is derived from the real-world situation where each of two individuals has the choice of continuing with an action (e.g., two drivers driving toward each other on a collision course), which will seriously harm both parties, or discontinuing the action. Discontinuation of the action will minimise harm to both parties. If discontinuation does occur, the party that initiates the discontinuation is described derogatively as being 'chicken'. Children and teenagers play many variations of this game, continuing on the same path until they bump into each other or until one of them (the 'chicken') swerves away. Let us apply this game of chicken to the following business decision facing the two airliner companies.

The winter sun warms the scantily clad bodies on the pearl-white beach, palm trees sway in the gentle breeze and the nightlife in the bar is really something. It's a great place to be, but this holiday destination has no regular passenger flights – holiday makers must drive here or take a coach.

Shutterstock.com/GaudiLab

To see how the problem of limited travel modes to this growing tourist destination could be relevant to the duopoly model, imagine a situation where two established airlines are each preparing to introduce a regular service to this destination. They are doing this because there is already an airstrip near the beach used to freight tropical fruit and they have calculated there will be sufficient demand for this route to enable only one of them to service it profitably. They also believe that if one of them decides not to enter this route, this company will be left behind and will have great difficulty establishing itself on this route in the future. Given this background, let us assume that the following outcomes are possible:

1 Both airlines continue with their plans and both suffer large losses because there is insufficient demand for both of them to cover high fixed costs of operation.
2 Both airlines discontinue their plans to introduce the new service and stick with existing routes. As a result both suffer a blow to their reputation because they are seen to have pulled out of their well-publicised plan to introduce the new service. Both must embark on a public relations campaign to restore their credibility.
3 One airline discontinues its development of the new service while the other continues with the plan. The airline that discontinues avoids the loss it would have incurred if both airlines had continued with their plans; but because it has 'chickened out' of its well-publicised plan it suffers a blow to its reputation and must incur the cost of a public relations campaign to restore its image. The other airline continues with its plan. Because it is the only provider of services on this new route it is able to reduce its average fixed cost and earn reasonable profits. In addition, its reputation and public standing are greatly enhanced because it has successfully implemented its well-publicised plan to offer this new service.

In short, if both airlines continue with the plan they both lose big-time (no one chickens out and the crash occurs), each making a loss of $3 million. If they both pull out, each suffers a blow to its reputation (they have both chickened out), making a reputation loss of $1 million. If one pulls out and the other continues, the one that pulls out suffers a significant blow to its reputation (loss of $2 million) while the other earns good profits and enhances its reputation (one pays the cost of being chicken while the other benefits from not chickening out) making a gain of $4 million. You can see how these outcomes mirror the real-world game of chicken played by children and teenagers. **Exhibit 9.7** shows the playoff matrix for this game of chicken.

Exhibit 9.7 **Pay-off matrix for airlines playing chicken**

		Airline B	
		Chickens out	Continues
Airline A	Chickens out	(-$1m, -$1m)	(-$2m, +$4m)
	Continues	(+$4m, -$2m)	(-$3m, -$3m)

It is clear that there is a high degree of interdependency in this game – the outcome for each airline's decision depends on what its rival does in response.

What do you think?

1 (a) If Airline A continues with the plan, what will Airline B do? Explain why the outcome is a Nash equilibrium?
(b) If Airline B continues with the plan, what will Airline A do? Explain why the outcome is a Nash equilibrium?
2 Is there an incentive for the two airlines to collude? Why? What might prevent such collusion?

An evaluation of oligopoly

Oligopoly is much more difficult to evaluate than other market structures. The game theory just presented does not give a definite answer to the question of efficiency under oligopoly. Depending on the assumptions made, an oligopolist can behave much like a perfectly competitive firm in some situations, or more like a monopoly in others. Nevertheless, let's look at a comparison of outcomes under perfect competition compared with an oligopoly selling a differentiated product.

First, the price charged for the product will be higher than under perfect competition. The smaller the number of firms there are in the oligopoly and the more difficult it is to enter the industry, the greater the oligopoly price will be in comparison to the perfectly competitive price.

Second, an oligopoly is likely to spend money on advertising, product differentiation and other forms of non-price competition. These expenditures can shift the demand curve to the right. As a result, both price *and* output may be higher under oligopoly than under perfect competition.

Third, in the long run, a perfectly competitive firm earns zero economic profit. The oligopolist, however, can charge higher prices and earn economic profits because it is more difficult for competitors to enter the industry.

Overall, we can conclude that, as is the case with monopoly, the market power wielded by oligopolists results in less competition and misallocation of resources.

In summary

- **Non-price competition** includes advertising, packaging, product development and quality differences.
- **Game theory** analyses the strategic decisions of players when the outcome for each is dependent on the behaviour of others.
- *Collusion* is a tacit agreement for oligopolists to fix prices to raise profit. **Price leadership** one method to achieve collusion occurs when a dominant firm in an industry raises or lowers price, other firms follow suit.
- A **cartel** is a formal agreement among firms to raise prices and set output quotas. A cartel is inherently unstable.
- Comparing oligopoly with perfect competition, we find that the oligopolist possesses large market power, tends to allocate resources inefficiently, charging a higher price and restricting output.

Key concepts

Monopolistic competition
Product differentiation
Non-price competition
Oligopoly
Concentration ratio
Mutual interdependence
Game theory
Duopoly
Dominant strategy
Nash equilibrium
Tacit collusion
Price leadership
Cartel

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Describe the characteristics of the monopolistic competition market structure

- **Monopolistic competition** is a market structure characterised by (1) many small sellers, (2) a differentiated product and (3) easy market entry and exit. Given these characteristics, firms in monopolistic competition are price makers but, because they are so small in relation to the market, they cannot affect the *market* outcome.

2. Know why we see the monopolistically competitive firm as a price maker

- Therefore, unlike perfect competition where the firm's demand curve is horizontal, the monopolistically competitive firm faces a downward sloping demand curve and is a price maker. Since there are many competitors, the demand curve it faces is more elastic than for a monopolist.

3. Understand price and output decisions for a monopolistically competitive firm

- The same $MC = MR$ rule can be used to determine short-run equilibrium for a monopolistic competitor. In the short run, a monopolistic competitive firm can earn economic losses or economic profits. In the long run, however, monopolistic competitors make zero economic profits due to ease of exit and entry.

4. Compare monopolistic competition and perfect competition

- Comparing monopolistic competition with perfect competition, we find that the monopolistically competitive firm does not achieve allocative efficiency. It charges a higher price, restricts output and does not produce where average costs are at a minimum. However, the gain from product differentiation in monopolistic competition may outweigh this social loss.

5. Describe the characteristics of the oligopoly market structure

- **Oligopoly** is a market structure characterised by (1) few sellers, (2) a homogeneous or a differentiated product and (3) significant barriers to entry. Oligopolies are **mutually interdependent** because the outcome of an action by one firm depends on its rival's reaction.

6. Apply the game theory to examine strategic decision-making for an oligopolist

- **Game theory** analyses the strategic decisions of players when the outcome for each is dependent on the behaviour of others. One such strategic decision is collusion to overcome the prisoner's dilemma.
- **Price leadership** is a strategy to coordinate price movements when a dominant firm in an industry raises or lowers price, and other firms follow suit. It is method to achieve tacit collusion to avoid price competition for increasing profits under oligopoly.

- A **cartel** is a formal agreement among firms to raise prices and set output quotas. The goal is to maximise profits, but firms have an incentive to cheat, which is a constant threat to a cartel.
- Comparing oligopoly with perfect competition, we find that the oligopolists tend to allocate resources inefficiently by restricting output.

Study questions and problems

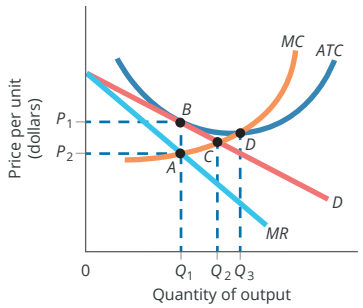
1 Suppose that in long-run equilibrium the minimum point on the *ATC* curve for a chair manufacturer's product is $200 per chair. Under conditions of monopolistic competition, will the long-run price of a chair be above $200, equal to $200 or less than $200? Explain your answer.

2 Explain the concept of a concentration ratio. Is the concentration ratio in a monopolistically competitive industry likely to be higher than for a perfectly competitive industry? What about the concentration ratio for an oligopolistic industry?

3 Using real-world examples discuss the characteristics of the oligopoly market structure.

4 **Exhibit 9.8** represents a monopolistically competitive firm in long-run equilibrium.
- **a** Which price represents the long-run equilibrium price?
- **b** Which quantity represents the long-run equilibrium output?
- **c** At which quantity is the *ATC* curve at its minimum?
- **d** Is the long-run equilibrium price greater than, less than or equal to the marginal cost of producing the equilibrium output? Shade the social loss of monopolistic competition in the long run.

Exhibit 9.8 **A monopolistically competitive firm in long-run equilibrium**



5 Draw a graph that shows how a new advertising campaign conducted by a monopolistically competitive firm would affect its demand curve and its short-run cost curves. (*Hint*: assume that advertising is a fixed cost.) Consider whether the combined effect of the change in the demand and cost curves is to raise or lower profits, *ceteris paribus*.

6 The table below shows the possible profit outcomes (in million dollars) of two firms, Giant and Big, adopting different pricing strategies. The first number listed represents the profit for Big and the second for Giant. Suppose they make their pricing decisions simultaneously.

		Giant	
		High	**Low**
Big	**High**	(5, 6)	(0, 8)
	Low	(7, 1)	(2, 3)

- **a** Is there is a dominant strategy for Big?
- **b** Is there is a dominant strategy for Giant?
- **c** What is the Nash equilibrium?

7 The Prisoner's Dilemma is perhaps the most famous idea in game-theory. Apply the Prisoner's Dilemma to one real-world behaviour of firms.

8 Why is mutual interdependence important under oligopoly but not so important in monopolistic competition or perfect competition?

9 Why is it that equilibrium for a monopolistically competitive firm is different from that of a perfectly competitive firm when both can earn economic profits in the short run but not in the long run?

Answers to 'You're the economist'

Which model fits the breakfast cereal market?

The fact that there are many differentiated products does not necessarily mean that there are many firms competing for your attention in the breakfast cereal aisle at the supermarket. Although there are a number of small suppliers, most of the cereals you see are produced by only three companies: Kellogg's Australia (which has nearly 40 per cent of the market), Nestlé Australia and Sanitarium Health Food. So, even though you see lots of different products in the supermarket, this doesn't mean the market structure is monopolistic competition. And with young adults shunning breakfast in ever-increasing numbers, this situation is unlikely to change any time soon. If you said the breakfast cereal industry is an oligopoly, YOU ARE THINKING LIKE AN ECONOMIST.

Is price leadership illegal?

In media release 254/99 issued by the ACCC on 23 December 1999, the then chairperson of the ACCC, Professor Allan Fels, said that:

> [b]anks, like other businesses, run the risk of breaching the *Trade Practices Act 1974* [now known as the *Competition and Consumer Act 2010*] if they 'telegraph' possible rises in interest rates to their competitors – and especially when they refer publicly to specific numbers such as 'up to 0.5 percentage points'.

He concluded by saying that:

> [t]he ACCC will maintain a close watch on bank behaviour in the coming period to ensure that there is no collusion.[8]

Similarly, in 2014 it was reported in the *Sydney Morning Herald* that the ACCC was looking at the major telcos Telstra, SingTel-Optus and Vodafone Hutchison to see if comments made by their executives constituted price signalling that resulted in smartphone price increases.[9] On the other hand, price leadership as practised in the petrol retail markets has been around for many years, the ACCC tried but found it difficult to present a strong case against this collusive behaviour. However, if you said the ACCC would consider taking action against banks or telcos that overtly engage in price leadership, YOU ARE THINKING LIKE AN ECONOMIST.

8 D. Ramli, 'Watchdog pricks up ears at telcos' price-signalling possibilities', *Sydney Morning Herald*, 19 May 2014.

9 Ibid.

Multiple-choice questions

1 Which of the following is correct?
- **a** Oligopolists must sell a differentiated product.
- **b** Monopolistic competitors can earn economic profits in the long run.
- **c** Game theory involves interdependency between firms.
- **d** Collusion between firms makes the market more efficient.

2 Which of the following industries is the best example of monopolistic competition?
- **a** Retail jewellery.
- **b** Household insurance.
- **c** Supply of mobile phone services.
- **d** Washing machine manufacturing.

3 Which of the following is *not* a characteristic of monopolistic competition?
- **a** A large number of small firms.
- **b** Economies of scale.
- **c** Easy market entry.
- **d** A differentiated product.

4 A monopolistically competitive firm will:
- **a** maximise profits by producing where *MR* = *MC*.
- **b** earn an economic profit in the long run.
- **c** shut down if price is less than average total cost.
- **d** do all of the above.

5 The theory of monopolistic competition predicts that in long-run equilibrium, a monopolistically competitive firm will:
- **a** produce the output level at which price equals marginal cost.
- **b** operate at the minimum average total cost.
- **c** produce more output than if it were perfectly competitive.
- **d** produce the output level at which price equals average total cost.

6 Referring to the following playoff matrix of a game, determine which of the following statements is correct. (The first number listed represents the outcome for Player 1 and the second for Player 2.)

		Player 2	
		Left	**Right**
Player 1	**Up**	(7, 9)	(10, 8)
	Down	(5, 6)	(3, 4)

- **a** There is no Nash equilibrium.
- **b** (Up, Right) is the Nash equilibrium.
- **c** There are two Nash equilibria for this game.
- **d** (Up, Left) is the Nash equilibrium.

7 Which of the following about simultaneous game is true? A simultaneous game is a game where:
- **a** each player makes his choice after his competitor's choice.
- **b** players make their choices at the same time.
- **c** players always have dominant strategies.
- **d** there is a Nash equilibrium.

8 In a game with two players, a player has a dominant strategy when:
- **a** the player's chosen strategy always maximises the combined benefit of the players.
- **b** the player's chosen strategy always maximises the social surplus
- **c** the player can always have a strategy that gives her better outcome than the other player in the game
- **d** the player has only the same best response regardless of the other player's strategy

9 If oligopolists successfully form a new cartel:
- **a** prices will be higher.
- **b** output will fall.
- **c** it is unlikely that members will try to cheat by secretly lowering prices.
- **d** prices will be higher and output will fall.

10 Microeconomic policy issues: Climate change and behavioural economics

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Evaluate how successive Australian governments have proposed very different policy approaches to the task of tackling climate change and carbon emissions.
2 Understand how economics incorporate social preferences in the rational choice model to explain behaviour.
3 Explain non-rational behaviour by combining psychology and economics.

Shutterstock.com/Sergey Panychev

A constant theme runs through the previous chapters of this book. This theme, which is central to the purpose of economics as a discipline, involves the idea that a knowledge of economics can help our understanding of markets and assist in the development of policies designed to increase the wellbeing of the community. In the first part of this chapter we look closely at the microeconomic issue of climate change and the part that policy can play in curbing these emissions in an efficient way.

Up until now, we have utilised economic concepts and theories to explain and predict a range of human behaviour pertaining to material requisites. In our explanation, we assume the motive behind our decision-making is to maximise self-interest. However, you must have observed that a lot of behaviours are influenced by other concerns than own profits and material wellbeing. Can economics explain these other-regarding behaviours? Furthermore, the rationality principle requires the economic person to be able to calculate the costs and benefits of each decision and make consistent decisions. The world we live in is full of uncertainty and our emotion and values creep in, influencing our decisions. Furthermore, most all of us do not possess the mental capacity to perform complex economic calculus. In the second part of this chapter, we will look at behaviours that violate or appear to violate the basic principles of the rational choice model, on which our economic analysis is based. In the past few decades, economics has been greatly enriched by the insights from cognitive and neural psychological research into decision-making. By combining psychology and economics, behavioural economics provides a more realistic model of human behaviour to not only more systematically explain a wider range of human behaviour, but also inform government to design and

implement public policies more effectively. We have seen the establishment of behavioural economics teams in government departments for this purpose, for example, the Behavioural Economics Team of the Australian Government (BETA). As you will see in this chapter, by recognising our various behavioural biases, the health authorities can design better public health preventive measures in combating COVID-19.

1 Climate change

As was explained in **Chapter 4**, economic analysis provides the tools that enable us to understand the concept of externalities and the ways in which government policies help us to deal with them. We saw that negative externalities, which impose costs on third parties not involved in a market activity, are tackled by government through a range of policies designed to reduce the output of the externality-creating activity.

Climate change, with its link to global warming, is perhaps the most important issue involving negative externalities in the world today. Although the extent of, and reasons for, climate change are highly contentious topics, there is nonetheless strong support around the world for the view that climate change is the result of human activity. There is, however, no international consensus on how to deal with the problem. Nor is there consensus at home: whereas previous Labor governments put Australia at the forefront of nations calling for worldwide acceptance of policies to tackle the problem head on, the Coalition government which took office in 2013 was much more cautious about tackling climate change. In the 2020 Paris Climate Meeting, Australia was urged to set a more ambitious target of net zero emissions by 2025, which required a much deeper cut than the Morrison government's pledged emission reduction by 26 to 28 per cent on 2005 level by 2030.

Climate change Long-term changes in the world's climate that are widely believed to involve global warming, and which are also widely believed to be caused by human activity.

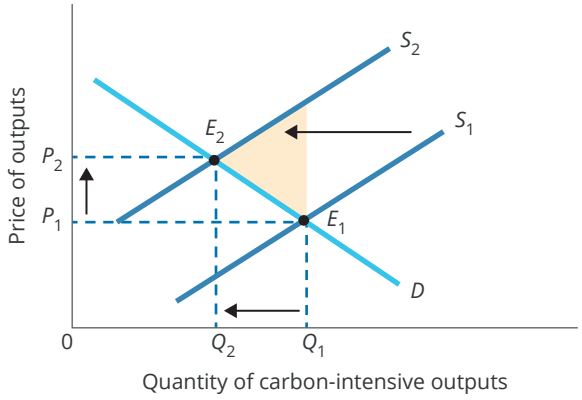
In the discussion that follows we concentrate on negative externalities arising from emissions of the pollutant that is widely argued to be the most important contributor to climate change. This pollutant is carbon dioxide (CO_2), which is mainly produced by the burning of fossil fuels. **Exhibit 10.1**, which has many similarities with part (a) of **Exhibit 4.6** in **Chapter 4**, shows the market for outputs that are currently produced by processes that involve the burning of large quantities of fossil fuels. The kinds of outputs we are thinking about include electricity, much of which comes from coal-fired power stations; steel, which uses metallurgical coal in its production; and road transport, which uses vast quantities of refined crude oil. These industries are central to the performance of modern economies. In **Exhibit 10.1** the market demand curve D and market supply curve S_1 establish the market equilibrium E_1 in the free market for carbon-intensive outputs. This is an inefficient outcome because no account is taken of the negative externalities this market produces. S_1 does not include the external costs imposed on the public by carbon-intensive industries because the carbon emitters are not paying the public for the damage caused by their emissions. The discharge of CO_2 and other pollutants into the atmosphere imposes an external cost, raising health care costs, contributing to global warming and, in general, eroding quality of life. If the external costs of carbon emissions were included in the industry supply curve in **Exhibit 10.1**, the relevant supply curve would become S_2. Supply curve S_2 therefore tells us the full cost of carbon-intensive product to the society (or social cost). The efficient equilibrium point would be E_2, which is the intersection of the market demand curve D and supply curve S_2, and the efficient level of output is Q_2.

It is now clear why the free market equilibrium E_1 is inefficient. This is because for output beyond Q_2 the social cost is greater than its social benefit, resulting in social loss. If there is no government intervention, at the free market equilibrium output Q_1 the social loss is represented by the shaded triangle.

In short, because the industry supply curve S_1 does not take into account the external costs, they are also not included in the price, P_1, of the carbon-intensive outputs produced by carbon emitters. The absence of the external cost of carbon emissions in the price of carbon-intensive outputs means that firms produce more outputs, such as electricity, steel and road transport services, than is socially desirable.

To the extent that the main culprit in climate change is carbon emissions, climate change can be brought under control if carbon-intensive output is reduced to the efficient level Q_2. But how can governments best go about achieving this outcome?

Exhibit 10.1 **Carbon emissions as an externality**



Resources are over-allocated to carbon-intensive activities at inefficient equilibrium E_1 because carbon-emitting firms do not include the external cost of CO_2 in their cost calculations. Supply curve S_2 includes the external costs of carbon emissions. If firms are required to purchase equipment to reduce carbon emissions, to purchase permits allowing them to emit, or pay a carbon tax on their emissions, that fully reflects these external costs, the economy can move towards the efficient equilibrium of E_2.

There are a number of ways in which the market failure associated with excessive carbon emissions can be addressed. These include: subsidies for renewable (clean) energy, a carbon tax, an emissions trading scheme and abatement subsidies. We look at each in turn.

Renewable (clean) energy subsidies and targets

A diverse range of government policies aimed at increasing the proportion of energy produced by non-polluting means has been implemented in Australia over the past two decades. Often, these policies are developed in the context of state and federal renewable energy targets (RETs) that set out the desired proportion of clean energy production to be achieved by a given future date; for example, a target may suggest that 30 per cent of production will come from clean energy sources by 2030. The fact that these policies have been freely introduced at both state and federal levels has meant that a hotchpotch of approaches can now be observed across the nation. Although very diverse in their nature and quantum, these policies, which have emphasised government subsidies to installers of clean energy generation equipment, such as solar panels and wind turbines, have had most influence on the electricity generation sector. By 2019, 21 per cent of Australia's electricity was produced by renewable energy sources, mainly by wind (7%), solar (7%), and hydro (5%) generators.[1] While the effect of subsidisation of clean energy is to reduce emissions of carbon during the generation process, such emissions still occur during the process of manufacture and installation of solar panels and wind turbines. However, as we will discuss in the 'Global perspective box' in this section, such emissions are usually ignored by advocates of so-called green energy.

Knowing that there is a budgetary limit to the extent to which carbon pollution can be reduced by subsidising clean energy, until the time when the production of renewable energy becomes more technologically effective and economically viable, governments in Australia and around the world have looked at other ways of reducing

[1] Australian Department of Industry, Science, Energy and Resources, 'Australian electricity generation – fuel mix', https://www.energy.gov.au/data/australian-electricity-generation-fuel-mix, accessed 12 January 2021.

carbon consumption. These other ways include *carbon taxes* and *emissions trading schemes*, which are discussed in the Australian context in the following two sections.

Carbon taxes

In 2011, the Gillard government announced that it would introduce a carbon tax, however this tax was abandoned by the Abbott government, which was elected in late 2013. A **carbon tax** involves government levying a tax per tonne of CO_2 emissions, which is taken to fully reflect the external cost imposed on society by carbon-emitting firms. This action inhibits production by imposing an additional production cost per tonne, and shifts the industry supply curve leftward from S_1 to S_2. The objective is to change the equilibrium from E_1 to E_2 and eliminate the overuse of carbon-intensive outputs in this market. Under this approach, the government is in effect charging the producer for the external costs of CO_2 emissions imposed on third parties. Thus, if the tax is set at the correct level, production of carbon-intensive output will be reduced to the efficient level Q_2, and the revenue it raises could be used to fully compensate those affected parties. Furthermore, the tax would encourage producers to install pollution-control equipment or substitute low-carbon-emitting production processes (including so-called 'clean energy') to minimise their carbon tax liability. Can you explain why this is sure to happen when pollution taxes are high but pollution control costs are becoming lower due to technological advancement? (*Tip*: try using marginal analysis.)

Carbon tax
A tax levied per tonne of carbon emitted by producers.

You might have noticed that in the previous paragraph we say the efficient level of output Q_2 is achieved *if the carbon tax is set at the right level*. Most economists are not climate change deniers, but some are doubtful regarding the extent of damage caused by the greenhouse effect as estimated by the IPCC. In other words, in order to set the carbon tax at the right level, the government has to have knowledge of the external costs of carbon emission. This is difficult to achieve when the assessment of these costs is still contentious. One approach to overcome this difficulty is to let the market decide, and this can be done through an emissions trading scheme, which has similar effects to a carbon tax but is much more flexible. This is what we will discuss next.

Emissions trading

Australian governments have been considering the introduction of an **emissions trading scheme (ETS)** for many years. The election of the Rudd Labor government in 2007 saw the promotion of a comprehensive scheme, sometimes known as the Carbon Pollution Reduction Scheme (CPRS), authored by prominent economist Ross Garnaut. Although the government argued long and hard that addressing climate change was an important moral and economic issue, the inability of nations to agree on a global solution to the problem at the 2009 Copenhagen Climate Change Conference led to a decision to take such a scheme off the agenda in Australia. This was said to be one of the reasons that Kevin Rudd lost the leadership of the Australian Labor Party in 2010. Similarly, following the poor showing of the Democratic Party at the 2010 US mid-term elections, President Barack Obama also rejected a trading scheme. By contrast, in late 2017, China had developed comprehensive plans to introduce such a scheme.

Emissions trading scheme (ETS)
A government scheme requiring firms that produce emissions of pollutants to purchase tradeable permits giving them the right to do so.

So what does an ETS involve? Often called a **cap and trade scheme**, the idea is for government to first announce the total allowable amount of annual carbon emissions that it considers to be consistent with its objective of, say, contributing to the stabilisation of climate change. This upper limit to allowable emissions is called the 'cap'. The total amount of emissions specified by this cap is then broken up into a number of permits that may be bought and sold in the market – the 'trade'. Firms wanting to emit CO_2 must buy permits allowing them to do so. However, no matter how much trade in permits occurs or what prices are paid for these permits, the total annual emissions of carbon cannot exceed the designated cap. Typically, each permit allows its owner to emit a given fraction of the total emissions allowed under the cap. So, for example, if there were 1000 permits of which a particular firm owned 200, this firm could emit 20 per cent of the total emissions allowed by the cap.

Cap and trade scheme
An alternative name for an emissions trading scheme.

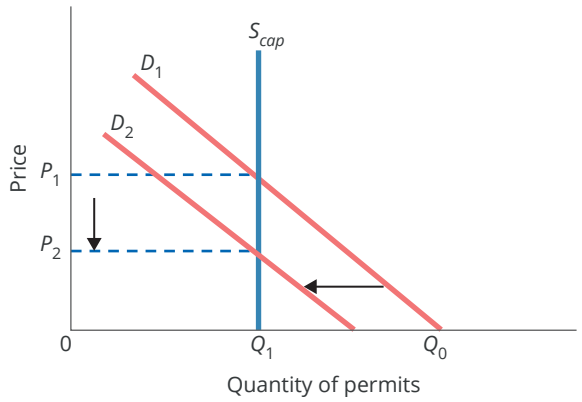
The market for permits in a particular country or region can be explained using a demand and supply diagram, such as **Exhibit 10.2**. The first thing to note in the exhibit is the supply of permits, which is indicated by the perfectly inelastic supply curve S_{cap}. This curve indicates that no matter how high or low the market price of permits may be, the quantity of permits supplied by government will remain unaltered at Q_1. The next important component of **Exhibit 10.2** is the demand curve D_1, which shows the quantity of permits that firms in the market would purchase at each possible price. Like any other market demand curve it shows that the quantity demanded of permits is inversely related to price. Also, note the quantity of permits, Q_0, that firms would demand if the price of a permit (i.e., the cost of pollution) were zero. A little reflection reveals that the total emissions that correspond to this quantity of permits is the total amount of emissions that would be produced if there was no permit system in place. This is the amount of emissions that firms will dump into the environment when there is no cost associated with doing so.

The market price P_1 of permits is given by the intersection of the demand curve D_1 and the supply curve S_{cap}. The need for firms to pay for permits to emit CO_2 results in increased production costs, which in turn shifts the industry supply curve in **Exhibit 10.1** leftward from S_1 towards S_2. The new supply curve will eventually settle in a position where the resulting equilibrium output produces just the amount of emissions that are allowed by the cap.

To help understand how a flexible permit price brings about automatic adjustment in the market, imagine that a slowdown in the economy occurs. In such a situation, this slowdown would cause a decrease in demand for, *inter alia*, carbon-intensive outputs resulting in reduced industry output, reduced emissions and a reduced demand for permits. This reduced demand for permits is shown by a shift in the demand curve for permits from D_1 to D_2 in **Exhibit 10.2**. This reduced demand for permits would lower their price to P_2, lower firms' costs and shift the industry supply curve for carbon-intensive outputs to the right. This would have the beneficial effect of increasing output, towards its original level before the onset of the slowdown in the economy.

To further see how useful a flexible cap and trade system can be, imagine a situation where a significant technological improvement in production processes reduces carbon emissions without increasing each firm's costs. This is happening right now as the cost of clean energy, such as solar and wind power, decreases significantly from year to year. In this situation each firm would need fewer permits to produce the same level of output, resulting in a shift to the left of the demand curve for permits. This, in turn, would lead to a fall in the

Exhibit 10.2 **The market for permits**



A decrease in demand for permits resulting from, say, a new emissions-reducing production technology, reduces the equilibrium price of permits.

price of permits. With the consequent lowering of costs for producers of carbon-intensive outputs, the industry supply curve would shift to the right and a higher level of output would be produced, but with no increase in total emissions. And this adjustment would occur without any need for government intervention whatsoever. On the other hand, if there were a carbon tax in place government would have to guess at the appropriate new level for the tax and undertake regulatory actions to change the tax level.

With neither the carbon tax nor an emissions trading scheme being favoured by the Abbott government, in 2013 it introduced its novel proposal for abatement subsidies, which were the centrepiece of its Direct Action Plan.

Abatement subsidies

The **abatement subsidies** proposed by the Abbott government in its Direct Action Plan were to be paid out of its Emissions Reduction Fund, details of which were revealed in a White Paper published in 2014.[2] This Emissions Reduction Fund was to provide over $2 billion to subsidise abatement of mostly carbon emissions by private firms and government bodies.

Abatement subsidies Government payments to polluters in exchange for a reduction in their emissions of pollutants.

This subsidisation takes the form of government payments to firms in exchange for a reduction in their emissions. On a regular basis the government calls for tenders, through which firms indicate the extent to which they will reduce emissions and the amount they wish to be paid for such reduction. Not all tenders are accepted by the government. Rather, it gives preference to those tenders that show the greatest possible pollution reduction per dollar spent. This scheme has the advantage that, just like a carbon tax or an ETS, firms have discretion as to how best to technically reduce their emissions. Successful tenderers enter into a contract with the government to reduce emissions by the agreed amount within seven years. As long as sufficient funds are made available, this scheme is theoretically capable of reducing carbon pollution by any specified amount (any cap). But, like existing subsidies for clean energy, there is a limit to the extent to which the federal government budget can subsidise pollution reduction – a problem that does not occur when firms pay for the right to pollute under an ETS or a carbon tax regime.

Applicable concept: emissions trading scheme

Global perspective

Why can an ETS give better outcomes than encouragement of clean energy?

Have you ever wondered how it is possible to make sense of the plethora of policies that are being used to solve environmental problems, particularly the problem of carbon emissions? Every day we receive a barrage of exhortations to switch to a greener, maybe hybrid, car; install solar panels on our roof; replace our white goods with energy-efficient models; switch to bike riding or public transport; install climate smart technology; use only off-peak power and so on. But what are we to make of this confusion of advice? Won't an ETS contain emissions in an all-encompassing way that saves the confusion brought about by so many government-inspired piecemeal solutions? And won't it bring an end to all the taxpayer-subsidised advertising that's designed to get us to think hard about these issues and change our ways?

As strange as it may seem, the answer to these last two questions is 'yes'. The one great advantage that economists see in an ETS is that it will put the right price on the production of carbon emissions, allowing economic agents to factor these prices into every decision that has implications for carbon pollution and ultimately climate change. Putting the right price on carbon means that when firms and individuals want to produce carbon pollution – in other words, despoil clean air – they must pay to do so; and there could be no more simple, sensible or straightforward idea than

[2] Australian Department of Industry, Science, Energy and Resources, 'Emissions Reduction Fund White Paper', April 2014, http://www.environment.gov.au/climate-change/government/emissions-reduction-fund/publications/white-paper, accessed 28 November 2017.

that. After all, if we want to use other scarce resources, such as land, labour, minerals or timber, we must pay for them or for the products that embody them, like fruit, home-cleaning services, motor cars or houses.

Alamy Stock Photo/Stephen Barnes

When we use so-called 'green' products or services, we know very little about their real effect on the environment. Thus, although consumers may have a good idea about the direct effect of solar panels in reducing carbon emissions from coal-fired power plants, they know almost nothing about the environmental costs of producing and installing solar panels. How much carbon, for example, goes into making the glass and metal that are used in solar panels? How much carbon is used in the processes of panel installation and continual monitoring of the system? What about hybrid cars? Owners and prospective purchasers will know about the fuel savings that are a proxy for carbon emissions, but what about the extra metal that goes into a hybrid car to give it the strength to carry its enormous battery load? Putting a price on carbon will mean that there will be large changes in prices for products and services brought about by differences in their total carbon input throughout the production chain. And there will be some large changes in the costs of operation of this equipment as the true cost of carbon embodied in the energy inputs, such as coal, oil, gas, wind and solar, are taken into account. In some cases the factoring in of carbon costs will further reinforce the messages we have been getting from governments to use particular green products or to behave in certain ways; however, in other cases the opposite will happen. Either way, the important thing for the environment is that our decisions are made on the basis of the true costs we impose on the environment.

Green products are the key to our future, but products are not really green if all they do is send misleading messages to governments and consumers that mask the true extent of environmental damage they impose. In the following extract from an article by John Breusch in the *Australian Financial Review*, the way in which a carbon reduction scheme can cut through these misleading messages is explained.

> [Will] the ETS actually encourage Australian businesses and households to reduce their emissions? To answer this, it's useful to remember why we're trying to use a price signal in the first place.
>
> Doing the right thing by the planet can be a headache.
>
> For instance, is it really true that the CO_2 emitted building a Toyota Prius – and particular [*sic*] mining and smelting the zinc for its battery – is greater than the savings generated when you drive it around town?
>
> Is there anything to the claim that the manufacture of a low energy light bulb produces more carbon pollution than the bulb saves once it's stuck in your ceiling?
>
> And is there any point riding a bike if you also eat meat produced from methane-belching cows?
>
> Implemented properly, a market-based mechanism like an ETS should provide an independent umpire in all those annoying – and generally unsolvable – debates.
>
> That's because it aims to build into the price of every product the cost of the greenhouse gases emitted during its creation, whether it be a car, the energy produced by combusted petrol, coal-fired electricity or those strange spiral light globes.
>
> In other words, the ETS works by changing relative prices.
>
> If an energy-saving light globe is cheaper to buy and run than an incandescent bulb, a rational consumer should opt for the former – and in doing so, help clean up the atmosphere.[1]

Elsewhere in his story Breusch laments the way that the relatively pure ETS originally proposed for Australia has been successively watered down with a dazzling array of exemptions and subsidies aimed at making political acceptance of the scheme palatable.

Even if a relatively pure ETS were to be introduced in Australia, does this guarantee that the problem of climate change will be solved? Unfortunately, the answer is no. While the ETS will make a significant contribution to solving our carbon emissions problems at home by changing the prices of products and services according to their contribution to carbon emissions, it cannot make much difference to the global problem of climate change. That will need agreement on the part of all of the world's major economies, including China and India. If the inability of governments to agree to a global plan at the 2009 Copenhagen Climate Change Conference is anything to go by, it will be a long time before that happens.

What do you think?

Explain your answers to the following questions on the basis of the research you've undertaken:

1 If Australia goes it alone and introduces an efficient ETS while the rest of the world refuses to do so, will this adversely affect Australian exporters' ability to compete in international markets?
2 Would imported carbon-intensive goods have a competitive advantage over similar goods produced in Australia?
3 Do these questions and your answers to them raise ethical issues about stewardship of the environment and the relationship between developed and developing countries?

1 J. Breusch, 'Everything you need to know about the ETS but were afraid to ask', *Weekend Australian Financial Review*, 28–29 November 2009, p. 22.

Which policy solution is best?

Of the three policy solutions discussed above, economists prefer the ETS over a carbon tax or abatement subsidies. The ETS has most support because it replicates what would happen in the real world if the owners of the clean air that is destroyed by emissions were able to exercise property rights over it. A trading scheme simply treats clean air as a limited economic resource that firms must pay to use. Clean air then becomes like any other scarce resource, such as land or labour – firms will economise on its use in response to price signals. Once the trading scheme is in place and permits have a price, it is entirely up to the firms themselves as to how they go about maximising profit in the light of the cost of the resource (clean air) they face. It is up to them, for example, whether to introduce new technology that reduces emissions (e.g., a better filtration system) or simply pay for permits. Over time, the price of permits will change as economic conditions change and as firms seek the best way to go forward. Thus, as we have seen above, if many firms introduce emissions-curbing technologies, demand for permits might decrease and their price fall with the result that firms' costs fall and the industry can expand without any increase in emissions beyond the level of the cap. And once a carbon trading scheme is in place, all of this adjustment is possible without the intervention of the heavy hand of the state.

In many respects, the carbon tax solution provides a similar outcome to that of a cap and trade scheme. Just like the cap and trade scheme, if polluters have to pay for the right to emit (this time as a result of having to pay a tax rather than buy a permit) they will economise on their use of clean air as a dump. In an endeavour to maximise their profits they can choose between paying the tax or installing pollution-control equipment. A disadvantage of a carbon tax is that there will be political pressure on government to keep the tax low and, in contrast, to cap and trade where the upper level of emissions (the cap) is determined directly, government can only guess at the rate of tax that will move the industry supply curve to the left to the extent required to bring emissions down to the appropriate level. Although government can bring about a change in the level of emissions by changing the tax rate, a criticism of carbon tax is that governments may be slow to react to changes in the economy, such as the introduction of new technologies or the onset of a period of low economic growth, which

require an altered tax rate to maintain a given level of emissions. By contrast, under a cap and trade scheme, emissions are maintained at the level of the cap.

While abatement subsidies can reduce emissions to the desired level, they have a number of shortcomings as compared to use of taxes or carbon trading. First, whereas taxes or emission permits raise polluting firms' costs, shifting the supply curves of carbon-intensive product to the left and reducing their output, subsidisation of pollution abatement will have no such effects. This is because the supply curves of firms being reimbursed for the cost of abatement will remain unchanged. In other words, the use of taxes or an ETS will reduce the size of the carbon-polluting sector of the economy whereas a subsidy scheme will not, although the abatement subsidies can incentivise investment in pollution curbing technologies.

Furthermore, in order to fund the abatement subsidies the government must raise taxes paid by other parties. This means that prices will rise in the non-polluting sector as a result of the additional tax costs they bear. The upshot of this rise in prices is that demand for these 'clean' outputs will fall. Overall, the resultant effect is for the non-polluting sector to become smaller relative to the polluting sector.

Finally, as comforting as moves to use subsidies to encourage wind or solar power may sound, the reduction of emissions by the use of these technologies may have enormous economic costs that may be more harmful to the economy and the environment than the emissions that are prevented.[3] For example, in a study by the National Generators Forum (a peak body for electricity generators) it was estimated that a New South Wales government scheme that subsidised installation of household solar roof-top generators was costing between $520 and $640 for each tonne of carbon reduced.[4] By contrast, estimates of prices for a permit to emit 1 tonne of carbon following the introduction of a cap and trade scheme have been less than $50 per tonne – or less than one-tenth of the cost of reduction through the use of solar panels. It would clearly be much cheaper to eliminate 1 tonne of carbon emissions by purchasing a permit and retiring it (taking it off the market) than to install roof-top solar panels. (Once an ETS is in place, purchase and retirement of permits could be undertaken by any interested party including the government and conservation groups.)

By and large the Direct Action Plan's abatement scheme received a lukewarm response from commentators and industry analysts when it was launched. In the years immediately following its introduction it attracted little attention and appeared to be of minor importance in the overall push to reduce pollution.

Analyse the issue

Applicable concepts: emissions trading scheme, market equilibrium, and costs of production

How would a reduction in the number of emission permits affect conventional and solar electricity generators?

In the above discussion of solutions to the problem of excessive carbon emissions, it was pointed out that the cost of reduction of carbon emissions through the use of solar rooftop panels is at least 10 times the cost of purchasing and retiring a permit to emit 1 tonne of carbon. As straightforward as this may seem, we should keep in mind that these calculations were carried out for the National Generators Forum, which represents electricity generators who mostly

[3] Subsidisation of clean energy production is often achieved by awarding government-created Renewable Energy Certificates (RECs) to persons and organisations that install equipment that produces clean energy. These certificates, which have a market value, can typically be either sold by the recipients or used to reduce the price of the clean energy equipment they purchase. Either way, these certificates can mean a significant reduction in the cost of installing clean energy equipment, such as solar panels or solar hot water systems. Read about RECs at https://www.rec-registry.gov.au.

[4] See A. Hepworth, 'Home solar costs 25 times more than ETS to cut gas', *The Australian*, 6 October 2010, p. 1.

use fossil fuel to generate their electricity. These calculations also took no account of how the market for permits might change over time. This is especially important given the inexorable decrease over time in the costs of renewable energy generation. It would thus be sensible to ask what would happen if, after the introduction of an emissions trading scheme, governments decided to make large-scale reductions in the level of emissions by purchasing and retiring emissions permits rather than encouraging installation of alternative energy-generating equipment like roof-top solar panels. Furthermore, how would such a change affect different sectors of the power generation industry?

iStock.com/chinasong

What do you think?

Using a diagram of the market for emissions permits like that in **Exhibit 10.2**, answer the following questions.

1 How would large-scale purchase and retirement of permits affect the equilibrium price and equilibrium quantity of permits?
2 How would this change in the equilibrium price of permits affect firms that generate electricity in the conventional manner using fossil fuels?
3 Would the change in the equilibrium price of permits help or hinder the solar power industry?

What does the future hold?

In 2017, following a number of serious problems with continuity of electricity supply around the country, the Turnbull government's policy emphasis moved away from the more general issue of pollution abatement enshrined in its Direct Action Plan, towards the issue of security of supply of electricity. While its new policy, described as the National Energy Guarantee (NEG), had provisions for the setting of emission reduction targets, its emphasis was squarely on requiring electricity suppliers to overcome the problem of security of supply (supply interruptions) brought about by solar and wind power's incapacity to provide backup power at times when solar power (generated only during daylight hours) and wind power (generated only when the wind blows) are insufficient to meet demand. Thus the emphasis shifted from pollution abatement to provision of backup capacity involving storage of clean energy through mechanisms such as pumped hydro and battery storage. This was, in part, prompted by South Australia's disastrous blackout in September 2016, which resulted in an undertaking by Elon Musk to supply South Australia with battery backup for its electricity grid within 100 days.[5] As clean energy progressively replaces energy from fossil fuels there will be an ongoing decline in carbon pollution accompanied by a growing need for backup storage to cover times when wind and solar cannot meet demand.

When all is said and done, there is more to the issue of reducing carbon emissions than has been discussed here. While economists are keen on the automatic adjustments that accompany the use of emissions trading schemes, they are aware that in a democracy there must be electoral support for the chosen policy approach. They are also aware that different approaches involve different implementation costs. One of the reasons for the rising popularity of carbon taxes as an alternative to an ETS is that a tax confined to large emitters, which are responsible for the vast bulk of emissions, may be a far simpler instrument than an ETS.

Finally, although we have been concentrating here on how best to go about reducing emissions, it is the view of many commentators that climate change has already become such a huge problem that transformation of the Australian national electricity market into one based on renewable energy is not a question of 'when' or 'if'. It

[5] See N. Harmsen, 'Elon Musk's giant lithium ion battery completed by Tesla in SA's Mid North', *ABC News online*, 24 November 2017. Available http://www.abc.net.au/news/2017-11-23/worlds-most-powerful-lithium-ion-battery-finished-in-sa/9183868, accessed 18 December 2017.

is about how the federal and state governments can come up with a cohesive and consistent policy framework to combine base load generation, renewables and storage capability to deliver lower price and stable supply of electricity for household and commercial users in Australia in the years to come.

Economics and ethics Politics of carbon abatement?

A major stumbling block to the introduction of policies designed to reduce carbon emissions is that they have differing effects on different members of the community. Although the net effect of these policies is to increase the overall welfare of the community, like many significant economic policy initiatives there will be winners and losers. Economists generally regard an ETS as the ideal policy to tackle CO_2 emissions. Yet they are fully aware that the outcome of the introduction of such a scheme will be much dissatisfaction on the part of those who are adversely affected. For example, most schemes have provision for low-income earners to be compensated for increases in prices of carbon-intensive essentials, such as public transport, electricity and heating fuels. Governments see this as an ethical response to the hardship that an ETS would cause low-income earners.

Unfortunately, every attempt to compensate the losers from the introduction of an ETS has the effect of moving the overall scheme away from its ideal form. Indeed it is possible that the raft of changes wrought by public opposition to the introduction of such a scheme could render it almost worthless. This is one reason why the idea of a simple carbon tax levied only on the largest producers of CO_2 has recently become more popular with policymakers. Nonetheless, the introduction of an ETS must weigh the efficiency gains, which result in lower carbon emissions against the adverse effects such a scheme can have on the least well-off members of society.

In summary

- **Climate change** refers to long-term changes in the world's climate, which are widely believed to involve global warming and which are also widely believed to be caused by human activity.
- A **carbon tax** is a tax levied per tonne of carbon emitted by producers.
- An **emissions trading scheme (ETS)** or **cap and trade scheme** is a government scheme requiring firms that produce emissions of pollutants to purchase tradeable permits giving them the right to do so.
- **Abatement subsidies** involve government payments to polluters in exchange for a reduction in their emissions of pollutants.
- From the economist's perspective, an ETS is the preferred instrument for tackling CO_2 emissions, which are strongly implicated in climate change.

2 Behavioural economics

In our analysis of human behaviour in dealing with the problem of scarcity, the focus is on material wellbeing. We assume that the economic person makes choices with a narrow goal of advancing their self-interest. Consumers aim to maximise their utility and the firms' goal is to maximise their profits, by comparing the cost and benefit of an action. In short, the **rational choice model** assumes human beings make rational decision to attain their goals with consistency and coherence.

Rational choice model A framework of studying behaviour by assuming human beings make consistent, rational choices to advance their self-interest.

Social norms, fairness and altruistic behaviour

We have applied this rational choice model with much success to explaining a wide range of decision-making – from the decision to emigrate, to attending university, to why businesses couldn't find workers (when unemployment rose sharply) during the coronavirus pandemic, to petrol price cycles, and tax incidence. However,

you must have observed that human beings are a lot more complex than the economic person we portrait, and material attainment is only a part of their existence, albeit an important part. While material considerations influence our action, our behaviours are often guided by our values, social norms, religious and other beliefs. We have a sense of fairness, honesty and a desire for doing the right thing. For example, we are taught not to cheat even though it will not be detected and punished by the law. Most of us will not trash public places even when there are no people around observing us. I bet you would return a wallet found on the side of a quiet street with a decent amount of cash in it to its owner, even when you could keep it without anyone knowing it. We feel good when we are doing the right thing since we have internalised these moral values.

We also frequently observe people care about others, and they do not always act in their self-interest only. We observe many instances of altruistic acts – passers-by risking their lives pulling a person out of a burning car, a stranger jumping into cold water to save a drowning child, and people donating to the charities, to name a few.

It appears the narrow rational choice model that we used in earlier chapters to study decision-making has ignored these human preferences. Is it because economists are unaware of these influencers of human behaviour? Adam Smith is regarded by many as the father of modern-day economic science. One of the most quoted texts from his book, *The Wealth of Nations* written in 1776, tells of the motivation – self-love – behind human cooperation in production and exchange in a commercial society:

> 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. We address ourselves, not to their humanity but to their self-love, and never talk to them of our necessities but of their advantages.[6]

However, it is important to note that while this notion of self-love or self-interest is central to the economic explanation of human behaviour, it is not the only factor the book talks about. In *The Wealth of Nations* and in his earlier book, *The Theory of Moral Sentiments* written in 1759, Adam Smith discusses at length the important roles of ethics and human temperaments in fostering cooperation, essential to the smooth running of a commercial society built on a high degree of specialisation and domestic and international trade. Kenneth Arrow, the youngest Nobel laureate (1972) observed that almost every commercial transaction contains an element of trust and the economic backwardness of many third world countries is due to the lack of trust.[7] In short, an efficient economic system cannot be attained solely based on self-interested motives in a 'jungle' environment where the strong prey on the weak, and where everyone has to be vigilant and fend for themselves to protect their own properties. To run smoothly the society needs trust and cooperation as a lubricant.

In her book, *The Secret Sins of Economics*, Deirdre McCloskey separates the factors influencing human behaviour into two groups, which she labelled as the P-variables (prudence, price, profits) and the S-variables (solidarity, shame, sacred). While these two sets of variables have their unique influence on behaviour, they also work together or in the language of econometrics interact to explain behaviour more fully. The close relationship between the P- and S-variables can be illustrated in the thinking of social corporate responsibility. In a capitalist society, the corporation has a responsibility to its shareholders, which is the maximisation of company profits and its share values. Corporate success is achieved by creating value for consumers, not by exploiting the environment or using immoral practices. A company operates within a wider (national and global) community, to which it has accountability to foster social cohesion and environmental sustainability. Therefore, the profit motive is modified by our concerns for social solidarity (fairness, humanity and sustainability). These humanistic and environmental concerns are shared by consumers; they would cast their votes by choosing not to purchase from companies that fail to uphold these social values.

Another group of economists take a slightly different approach to studying these human virtues and social norms by incorporating social values in the rational choice model. Not unlike consuming a gelato cone or buying a gym membership, these human virtues and social norms can be treated as goods. By being cooperative

---

6 A. Smith, *The Wealth Of Nations*, Book IV, Chapter II, p. 456.

7 K.J. Arrow, 'Gifts and Exchanges', *Philosophy & Public Affairs*, 1972, Vol. 1, No. 4.

(not littering in public), fair (not cheating) or altruistic (saving a drowning child), the actor feels good about themselves. This good feeling creates utility in the same way eating a gelato cone does. Human beings are said to have a taste or preference (in economic jargons) for these noble human qualities. Of course, there is a cost of obtaining this good feeling for being cooperative or altruistic in terms of inconvenience or putting your life at risk. However, as long as the expected utility (good feeling) derived from these acts outweighs their costs, human beings will display these behaviours voluntarily. Therefore, these social behaviours are consistent with, and can be explained by, the rational choice model in the same way as other goods and services. In the 'Analyse the issue' section that follows we will examine one of these altruistic behaviours – donation to the charity. The empirical data will show that our donation behaviour, like consuming a gelato cone, follows the law of demand as predicted by the rational choice model.

In this book we apply the rational choice model to examine the operation of the market: how our producers and consumers interact to generate and respond to prices and profit signals, with the goal of maximising own welfare. In **Chapter 5**, one of these P-variables we looked at is family size (or number of children). In developed countries when per capita GDP rises, we observe falling family size. As you might recall, we argued that it is incorrect to draw a conclusion that children are an inferior good from the observed association of rising income and falling family size. Instead, we should look for a third factor and we found it in the rising price (cost) of raising children in developed or rapidly developing countries. When the price of raising children goes up, the parent will demand fewer children to maximise their utility; this is what the law of demand predicts. You might be irritated when economists treat children as a good or an asset (see Gary Becker) whose existence is to give their parents utility (be it present happiness or future security). However, it is important to point out that this analysis should not be taken to imply that economists are obsessed with prudence, or in McCloskey's words 'have forgotten love and courage, justice and temperance, faith and hope'. As a social scientist, the economist aims to establish the most parsimonious model to explain behaviour by abstracting away irrelevant details. The validity and value of an economic model lies in its simplicity and testability (or falsifiability). So far, economics has passed this test. Economists are not ignorant of the love and faith in a family that gives meanings to human existence.

The juxtaposition of the prudence and sacred variables in the rational choice model as proposed by McCloskey and the treatment of social norms and values as goods prove to be a fruitful way to study their influence on behaviour. And the validity of this approach can be tested in the same way as other economic factors.

Analyse the issue

Applicable concepts: utility, opportunity cost, demand, income effect

The economics of giving

Economics has been enriched by not only cognitive psychology but also neuroscience, giving us a deeper understanding about the biological basis when making economic decision. If people derive happiness (or utility) from giving to a charitable donation, will the pleasure centres of our brain be activated in the same way as when we eat a lobster? Harbaugh and his colleagues used functional magnetic resonance imaging (fMRI) technique to investigate how our brain responds to donations. In the study, participants were given $100 and part of it was donated to a food bank.[1] Of all donations, roughly half was voluntary. That is, the participants can accept or reject those parts of this money to be transferred to the food bank. For this part of donations, the decision to give is entirely voluntary. Therefore, giving this part to the charity incurs a cost to the participant. The other part of the donation was compulsory. Using fMRI scanners it was observed that participants' pleasure-related parts of their brain were activated, indicating they receive joy from making the donations. When the donation was voluntary, neural activation was stronger, suggesting that altruistic motive received more reward from donating to the public good. The neural connection between giving and joy was confirmed in a more recent fMRI study by Kahnt and his associates.[2]


Shutterstock.com/mark higgins

To further investigate the economics of donation, if donation is treated as a good, giving the donor pleasure, we ask: Does the law of demand apply to the consumption of this good, which is donation to the charity? In other words, will the amount of donation be negatively related to its cost? In most countries, charitable donations are tax deductible. Changes in taxation laws provide us with naturalistic experiments to observe taxpayers' response to changes in cost of donation brought about by changes in tax rates. Suppose the top marginal tax rate in a country is 30 per cent, meaning that each $100 of donation will reduce the top earner's taxable income by $100, and their tax payment will be reduced by $30. Thus, the cost of $100 donation is only $70. Lowering the top marginal tax rate to 20 per cent will increase the cost (or price) of a $100 donation to $80. Therefore, it is predicted that top income earners will donate less if the top marginal tax rate is lowered. In a US study using time series data, Auten and his colleagues examined the effects of a lowered and broadened tax structure on charitable donations. They estimated that for a US taxpayer facing a 30 per cent marginal tax rate, as a result of the tax reform the amount of donation would be reduced by between 25 and 36 per cent.[3] This shows that donation is just like other goods and services, and follows the law of demand.

What do you think?

1 In the neural experiment mentioned above, the researchers distinguished two groups of participants according to whether they accepted (altruist) or rejected (egoist) the voluntary parts of donation. It was observed that when making a donation, fMRI scanning revealed different levels of brain activities of these groups, with the altruists showing more glow warm activities than egoists.

a Draw two separate demand curves of donation, one for altruists and one for egoists in the same diagram. Who have a higher demand for donation – altruists or egoists? (You can assume they have the same income.)

b Suppose the egoist and altruist face the same cost of donation, who will donate more to charity? Illustrate with a diagram.

2 Suppose the tax rates are unchanged and donation is a normal good. Consider a person who has a promotion and her income now moves to a higher tax bracket. With the help of a diagram, discuss if she will donate more or less to charity. (Hint: Consider both income and price effects.)

1 W. T. Harbaugh, U. Mayr, and D. R. Burghart, 'Neural responses to taxation and voluntary giving reveal motives for charitable donations', *Science*, 15 June 2007, Vol. 316, No. 5831, pp.1622–1625. doi: 10.1126/science.1140738

2 S. Q. Park, T. Kahnt, A. Dogan, S. Strang, E. Fehr and P. N. Tobler, 'A neural link between generosity and happiness', *Nature Communications*, Vol. 8, No.1, pp. 1-10. DOI: 10.1038/ncomms15964

3 G. E. Auten, H. Sieg and C. T. Clotfelter, '*Charitable giving, income, and taxes: An analysis of panel data, American Economic Review*, February 2002.

In summary

- Human beings' pursuit of self-interest is the underlying assumption of the rational choice model. But we frequently observe many instances of other-regarding behaviour. Our concerns of others' welfare, social norms and values (sense of justice, fairness, reciprocity) have powerful influence on our behaviour.
- Economics does not ignore human virtues. The rational choice model on which classical economic analysis is based can adjust to incorporate these norms and values to explain behaviour.

3 Behaviours violating the rational choice model

So far, the other-regarding behaviours that we have examined (e.g., giving to the charity, acting honestly and fairly, caring for the community) have all been treated as having a self-interest basis, and are consistent with, and can be explained by, the rational choice model. These behaviours are sanctioned by social norms and ethical values. And through internalisation and enculturation, they create intrinsic value (happiness) because we feel we are doing the right thing. Moreover, these behaviours follow the law of demand. There is nothing irrational about them. In this section, we will look at behaviours that violate the assumptions of the rational choice model.

Inconsistency of human behaviour

Economists do not always agree regarding the definition of rationality. However, it is generally accepted that to be rational, one must be, at the minimum, consistent with their decisions. Let us look at a few examples of inconsistency observed in experiments Kahneman and Tversky conducted.[8] These experiments also highlight the limited capacity of human beings in discerning outcomes involving uncertainty. In the first study, the participants are given choice problems, each of which contains two options.

Problem 1: Option A (4000, 0.80) or Option B (3000, 1)

Let us first look at the content of each option for Problem 1. Option A gives an 80 per cent chance of winning $4000 and Option B gives $3000 guaranteed (100%). The experimental result shows that of all participants 80 per cent choose B and 20 per cent choose A. This is not surprising, as most people are risk averse. Although Option A gives a higher expected monetary value of $3200 (= $4000 × 0.80), the certainty of a lesser amount ($3000) is preferred by most participants. Psychologists call this *certainty effect* – or as in a popular saying, a bird in hand is worth two in the bush.

Problem 2: Option C (4000, 0.20) or Option D: (3000, 0.25)

Now, consider the second problem. Option C gives a 20 per cent chance of winning $4000, and Option D gives a 25 per cent chance of winning $3000. The data shows that 65 per cent of the same group of participants prefer C to D. At first sight, the results from the two problems do not seem to be out of extraordinary. You would probably make the same choices as the majority – a lesser sum with certainty (B) in Problem 1 and a higher amount with a slightly lower probability (C) in Problem 2.

But if we examine the two problems closer, we will notice that Options C and D in Problem 2 are respectively equivalent to Options A and B in Problem 1. Here is the explanation. We multiply both Options A and B with a probability value (p) of 25 per cent. We can see that 25 per cent chance of Option A (4000, 0.8) occurring is the same as winning $4000 with a probability of 20 per cent (= 80% × 25%). Therefore, pA is equivalent to Option C (4000, 0.2). Similarly, 25 per cent chance of Option B occurring (winning $3000 with certainty) implies pB is equivalent to Option D. If you prefer B to A, then to be consistent you would prefer 25 per cent chance of B to 25 per cent chance of A. In short, if B is preferred to A, then pB (or D) is preferred to pA (or C). Did you notice the inconsistency of decision-making in Problem 2? Similar results have been reported in other similar experimental studies.

Multiplying both A and B by a probability value (25%) turns the certain event B (winning $3000 with certainty) into an uncertain event. Apparently, this change has a greater adverse effect on B than A, with the

[8] D. Kahneman and A. Tversky, 'Prospect Theory: An analysis of decision under risk', *Econometrica*, 1979, Vol. 47, No. 2, pp. 263–292.

result of reversing many participants' choices. The rational choice model fails to account for this inconsistency. To further understand how human beings make choice under uncertainty, we consider a second experimental study reported in the same paper. This study consists of the below two problems.

Problem 3: Option E (6000, 0.45) or Option F (3000, 0.90)

In Problem 3, Option E gives a 45 per cent chance of winning $6000 and Option F gives a 90 per cent chance of winning $3000. Both Options E and F have substantial probabilities (45% and 90%), and the experimental result shows that when the events are probable, most participants prefer Option F (86%) to Option E (14%). That is, they prefer a lesser amount with a higher probability (F) to a larger amount with a lower probability (E).

Problem 4: Option G (6000, 0.001) or Option H (3000, 0.002)

In Problem 4, with Option G having a 0.1 per cent to win $6000 and Option H a 0.2 per cent chance to win $3000, it is probable but unlikely to win any money in either case. Most participants preferred Option G (73%) to F (27%); that is, for events with a very small chance of winning, the participants prefer the option with the greater amount. Using similar calculations, as in the first experimental study involving Problems 1 and 2, we can show that Options E and F in Problem 3 are equivalent to Options G and H in Problem 4, respectively. In this study, the same inconsistency can be observed as in the first experimental study.

Kahneman and Tversky proposes a new approach – the Prospect Theory – to explaining the choices made by majority of participants in these experiments, which violate the consistency assumption in rational choice theory (see their paper for more detailed explanations). It suffices to refer to their conclusion here. In the first study involving events with substantial probabilities, the result shows a tendency of human beings to over-emphasise certainty of gain. On the other hand, the second experiment involves events with very small probabilities; the data shows we are prone to over-estimating very low probabilities. This over-estimation of minuscule probability is commonly observed in gamblers' behaviour. Have you bought the Lotto tickets before? If you have, you must have placed a much greater hope on winning a large amount of money than the objective fact (objective probability).

Framing and loss aversion

Our inconsistency is not restricted to situations dealing with uncertainties. A problem can be presented in many ways and the rational choice model predicts us to make the same decision under different framing of the same situation. However, Tversky and Kahneman found that our decision is often influenced by how the problem is presented or framed, resulting in preference reversal.[9] Study the following scenario in their experiment to investigate the effect of framing. Some details are altered to avoid any possible misinterpretation.

Imagine that Australia is preparing for the outbreak of an unusual African disease, which is expected to kill 600 people if no intervention takes place. Two possible programs to combat the disease have been proposed.

The subjects are randomly divided into two groups of roughly equal size. For Group 1, the below information about the outcomes of the two proposed alternative programs are presented. The subjects are asked to choose which program they prefer (results shown in brackets following each option):

- If Program A is adopted, 200 people will be saved. (72%)
- If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved. (28%)

[9] A. Tversky and D. Kahneman, 'The framing of decisions and the psychology of choice', *Science*, 1981, Vol. 211.

For Group 2, the same scenario is presented but followed by a different description of the outcomes of the alternative programs. Again, the subjects are asked to choose which program they prefer (results shown in brackets following each option):

- If Program C is adopted, 400 people will die. (22%)
- If Program D is adopted, there is a one-third probability that nobody will die and a two-thirds probability that 600 people will die. (78%)

The proportion of participants who choose each option is given in parentheses. Careful examination of the descriptions shows the outcomes of programs presented to both groups are identical. However, the preferences are reversed. For Group 1 where the outcome is framed as lives *saved*, the majority of participants (72%) prefer the certainty of saving 200 people to a risky outcome with the same expected value. These participants are risk adverse when the outcomes are presented as gains. However, when the outcome is presented as lives *lost* (Group 2), the certainty of 400 deaths becomes less desirable than two-thirds probability of 600 people will die. Most participants (78%) prefer taking risk when the situation is described in terms of losses. This experimental result clearly violates the assumption of rational choice model that we make the same decision given the same situation. It shows that framing of the situation influences our perception, leading us to reverse our preferences.

How can we explain this preference reversal? Tversky and Kahneman proposed the asymmetric value function to account for this observation. Based on experimental evidence, they argue that people's response to losses is more intense than gains of the same amount. In another study in the same paper, the subjects were told two events (A and B) occur to them. In event A, they get an unexpected gift of $100 and in event B they have a broken water pipe, which costs them $80 to fix. Are they better off or worse off as a result of the occurrence of these two events? Tversky and Kahneman found that many people when presented with the events actually *refuse* to accept this pair of events, which give them a net increase in wealth. Instead of combining them, they view the two events separately and the displeasure due to the loss (broken pipe) is much greater than the pleasure from the gift.[10] It is this loss aversion that explains the preference reversal observed in the above experiments.

Nudging

Nudging
A choice architecture to influence optimal decision-making without changing the incentives and reducing individual freedom.

If framing can influence our decision-making, we may be able to change behaviour by changing the presentation of a public policy without altering its content. This idea is known as **nudging**. Sunstein and Thaler were the first to populate the idea of nudging in their 2008 book *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Nudging aims to influence people's decision-making by changing the environment or framing the situation where choices are made, which Sunstein and Thaler call 'choice architecture'. The technique does not involve altering incentives or curtailing individual freedom. Here is a common example: by placing junk foods at top or bottom of the shelves, and healthy foods at eye levels in the school canteen, school students are more likely to eat more healthy foods.

As another example, we look at organ donation. Organ donation presents a huge challenge to most countries. When surveyed about their attitudes towards organ donation, the majority of us are in support of it, but very few of us actually sign the organ donation form. For example, in Canada it was found less than a quarter of the population have signed the organ donation cards[11] and in Australia the number is one in three.[12] To encourage organ donation, it is common in many countries that when registering for their driver's license people are also

[10] A. Tversky and D. Kahneman, 'The framing of decisions and the psychology of choice', *Science*, 1981, Vol. 211.
[11] N. Robitaille, 'A little nudge goes a long way in increasing organ donor registrations', *The Conversation*, 2 May 2019, https://theconversation.com/a-little-nudge-goes-a-long-way-in-increasing-organ-donor-registrations-115051
[12] Healthdirect, 'Organ and tissue donation', https://www.healthdirect.gov.au/organ-and-tissue-donation
[13] In Australia, the option to register as an organ donor on driver's license renewal is no longer available. If you have registered on your driver's license renewal form. you still need to register with Australian Organ Donor Register for your consent to be valid.

asked to register as an organ donor.[13] In Ontario, Canada, with a view to improving the registration rates, printed on the top of the consent-to-donate form in bold are these suggestive questions:

- 'If you needed a transplant, would you have one?'
- 'How would you feel if you or someone you loved needed a transplant and couldn't get one?'

These reflective statements aim to create an empathetic environment and caused in some part of the state over 100 per cent increase in donor registrations[14] – a huge success at very little cost. Some governments have also used nudging to increase donor registration by taking advantage of one of our cognitive biases, which we will discuss in the next section.

Biases and erroneous decision-making

Loss aversion and over-estimation of very small probabilities of gains lead to inconsistent decisions. Another reason why we make erroneous judgement or inconsistent decision is our failure to act on objective information due to various types of biases. Here, three such cognitive biases will be discussed.

Status quo bias

Because of myopia and the uncertainty of gains/losses, we have an inertia of continuing with the current activity (status quo) rather than making a change. Economists call this the **status quo bias**. One example of status quo bias involves pension saving. Retirement saving is in the long-term interest of an employee. However, the cost of saving now is certain whereas the benefit in the future is uncertain and discounted heavily for some individuals. Employees tend to stay with their current spending habits and will not actively opt into a pension saving plan offered by the company or the government for their benefit. When asked to opt in, not to change is often the chosen option. Behavioural economists use our status quo bias to 'nudge' people to make better decision to improve their personal and social welfare by altering the default settings. Instead of requiring the employee to actively choose to join the pension plan, they are automatically enrolled in it, but have the choice to opt out if they wish. This slight change in the design of choice environment (or 'choice architecture') greatly increases pension saving in the UK and New Zealand.[15] This provides not only future security for these employees but reduce the tax burden on future generation to support the aging population in most western countries.

Status quo bias
The preference to continue with the current activity (status quo) rather than making a change because of myopia or future uncertainty.

Another big success of altering defaults to change people's behaviour for the better is in organ donation. As noted earlier, due to the status quo bias, despite support of organ donation few people take the active step to opt in as donors. Assuming implied consent, many countries change the legislation to make opt-in the default option. This means you have to actively sign an opt-out form if you do not want to be a donor. The new default setting results in a large increase in organ donation, saving thousands of lives.

Sunk cost bias

In **Chapter 6**, we pointed out that if the cost of an input incurred in the past is not recoverable, rationality requires that this cost should not be taken into account when making production decision. This type of cost is called sunk cost. However, in the real-world people often show an emotional attachment to investment made in the past and continue with the activity, feeling that 'the money spent in the past will otherwise be wasted'. The **sunk cost bias** occurs when we continue with an activity because of investment in the past even though the invested amount is not recoverable.

Sunk cost bias
The choice to continue with an activity because of investment in the past even though the invested amount is not recoverable.

Let's take look at an example of the sunk cost bias in action. You regularly visited friends living in a nearby suburb. On one visit you noticed that in their son's room there was now a second, almost identical bunk bed,

[14] N. Robitaille, 'A little nudge goes a long way in increasing organ donor registrations'.
[15] Behavioural Economics Team of the Australian Government, 'Harnessing the power of defaults', https://behaviouraleconomics.pmc.gov.au/sites/default/files/resources/harnessing-power-defaults.pdf (accessed 31 December 2020)

only in much better condition than the one that they have had for a while. Surprised, you asked the hostess why there were two sets of bunk beds when they had only one son. You had also mentioned to the hosts in the past you were looking for a bunk bed for your growing family. The hostess explained that it was given to them by their Asian neighbour who were immigrants in Australia and had recently returned to their home country. Your friend was so kind that she offered to give you their neighbour's bunk bed for free, but added that she couldn't give you their own bunk bed (which was in fact not in as good condition). The reason, she said, was that they paid $500 for it. Do you think your friend was making a rational decision?

To find out if sunk cost influences decisions, Arkes and Blumer conducted an experiment at Ohio University Theatre. A season ticket for 10 plays was sold to three groups to which participants were randomly assigned. One group paid the normal price of $15; the second group paid $13 ($2 discount); and the third $8 ($7 discount). The buyers were told the discounts were part of a promotion by the theatre. Since the season ticket price is a sunk cost, you would expect the three groups used the season ticket to attend a similar numbers of plays. Using the data in the first half of season, the researchers found the normal-price group attend about 25 per cent more of the plays than the two discount price groups. To these groups, sunk cost matters. But why? When participants paid a high price, they felt a greater obligation to attend so that they could get their money's worth from the ticket. It is the same emotional attachment that explains why your friend was not willing to give away her own bunk bed.

To our dismay, Arkes and Blumer also found that formal economic training would not reduce the sunk cost effect! Therefore, it is not surprising that businessmen are also vulnerable to this bias. The sunk cost bias is also known as the Concorde fallacy. The Concorde Supersonic jet was a joint project started in 1960s by the British and French governments. The project suffered high-cost overrun – six time higher than planned – and was a financial disaster. Despite being not economically viable, the governments forged ahead, claiming that there was too much invested (in the project) to quit.[16]

Over-confidence bias

Over-confidence bias
People exhibiting greater subjective confidence in themselves than the objective accuracy of their judgement, resulting in over-estimation of positive abilities and under-estimation of negative ones.

Let us now look at the last cognitive bias. It has been well documented that people tend to have much greater subjective confidence in themselves than the objective accuracy of their judgement, resulting in over-estimation of positive abilities and under-estimation of negative ones. Economists call this **over-confidence bias**. The most well-established finding of over-confidence is people's over-estimation of driving skill. In numerous studies, the majority of those surveyed rated themselves as better than average. In one such study, the proportion is 93 per cent![17]

Effective ways to reduce the transmission of COVID-19 include washing hands, social distancing and wearing masks. Non-adherence to these practice is frequent because many people are over-confident that the virus will affect others rather than themselves. The same belief is likely to be one of the reasons why many people still attended anti-lockdown protests even during the height of the pandemic and many did so without wearing masks, despite the health authority's advice regarding the danger of contracting/spreading the virus in mass gatherings.

Decision heuristics

In the real world, many tasks are complex involving many relevant variables, and we often have to handle these tasks without full information. In other situations, the task is obscure or we have to make a decision or judgement in a short time. Therefore, instead of obtaining the necessary information for decision-making, which can be very costly, we often estimate a value of the variable of interest or make a judgement by relying on a short cut

16 A. I. Teger, 'Too much to quit', 1980, Pergamon Press, New York.

17 O. Svenson, 'Are we all less risky and more skillful than our fellow drivers?' *Acta Psychologica*, 1981, Vol. 47, No. 2, 143–148. doi:10.1016/0001-6918(81)90005-6

or hints from the environment. A **decision heuristic** is a mental short cut or rule of thumb to make a quick decision or form a judgement when dealing with a complex or obscure task. A decision heuristic gives us roughly the correct answers in most situations (otherwise we as a species wouldn't have survived long). However, these short cuts can also lead us to biased decisions in many others. In the following we will discuss three heuristics that we use to simplify decision-making for complex or obscure tasks.

Decision heuristic
Mental short cut or rule of thumb to make a quick decision or form a judgement when dealing with a complex or obscure task.

Anchoring heuristic

When making an estimation of some obscure value there is a tendency to rely on any additional information that appears to be relevant, when it fact it is not. In one study Strack and Mussweiler asked their experimental subjects to guess Mahatma Gandhi's age at the time of his death.[18] The subjects were randomly divided into two groups. Before guessing, the experimenter asked the first group whether Gandhi was older or younger than 9 years when he died, and asked the second group if he was older or younger than 104. The number (9 or 104) served as an anchor. Although it is clearly not plausible Gandhi died at 9 or 104, the data show that the implausible anchors influenced estimation. The low implausible anchor produced an average guess of 50 years, and higher anchor a much greater average of 67.[19] This result illustrates the **anchoring heuristic**.

Anchoring heuristic
The tendency to use the first or most recognisable piece of information to make decision or judgement when faced with uncertainty or lack of any information.

To make better sense of this finding we compare the research interest of cognitive psychology and economics. Cognitive psychologists have done much research on our intuitive or instinctive response to our environment, while economists focus mainly on human beings' well-considered, calculated approach to choice making. In his book *Thinking, Fast and Slow*, Kahneman distinguishes two systems of thinking: System 1 and System 2. System 1 is fast, intuitive and tends to be automated and unconscious, involving less mental effort. System 2 is the opposite, to which most of decision-making economists study belongs. In order to conduct our daily activities effectively, which may not always lend itself to slow and deliberate considerations, we have to rely on System 1 thinking, which provides us with a short-cut to fast decisions and putting less strain on our mental capacity.

One familiar example of System 1 thinking involves our instinctive response to potential danger. In the middle of the night on your way home, you saw a guy walking towards you. Your intuition prevails: He is tall and strong, and you decided to walk (run?) in the opposite direction. The clue (outlook of a person) allows you to make a quick decision with the minimum effort. In the experiment involving Gandhi's age, even though the anchor number is implausible, it allows participants to make a comparison without having to mentally search for other information (which requires mental effort). Our System 1 thinking is at work in both situations!

Availability heuristic

When assessing the frequencies of an event, familiarity and salience of instances of the event will influence our memory and judgement. **Availability heuristic** is the tendency to overestimate the frequency of an event or phenomenon if the examples of the event are familiar and to underestimate the frequency of those that are unfamiliar. For example, some people fear travelling by air even though the chance of an air crash is much lower than a car crash. This is because when an airplane accidence occurs, it is likely to be widely reported with images of the crash scene shown on TV and in the newspapers, making air crash instances much more salient and easier to recall, and therefore perceived as more likely to occur than it actually is.

Availability heuristic
The tendency to overestimate (underestimate) the frequency of an event or phenomenon if the examples of the event are familiar (unfamiliar).

In one experiment, Tversky and Kahneman prepared four audio-recorded lists of names: two lists of entertainers and two lists of public figures. Two lists (one of entertainers and one of public figures) contained 19 names of famous females and 20 names of less famous males. The other two lists (one of entertainers and one of public figures) contained 19 famous males and 20 less famous females. A group of 99 participants were instructed to listen attentively to the recordings. After listening, they were asked to determine whether each

[18] T. Mussweiler and F. Strack, 'Considering the impossible: Explaining The effects of implausible anchors', *Social Cognition*, 2001, Vo. 19, No. 2,145–160. doi: 10.1521/soco.19.2.145.20705

[19] For your information, Gandhi died at age of 78.

of the four lists contained more men or women. The data show 80 of the participants judged, erroneously, that the gender of the *famous* minority was more frequent.[20]

This familiarity bias worked to the Australian government's advantage during the first year of the COVID-19 pandemic. At the early stage of the pandemic (around March and April 2020), many people were sceptical of the severity and susceptibility of the virus. When Tom Hank's positive testing for COVID-19 was announced when he was working on location in Queensland, we observed a big change in people's perceived risk of the virus. This celebrity effect illustrates the influence of the availability heuristic on our judgement about our vulnerability against the virus; it raises our awareness about the damage of COVID-19. This availability heuristic, unfortunately, could also work against the health department's effort around the world to contain the virus when top government officials and political leaders attended political rallies without wearing masks.

In 2020, anti-lockdown protests occurred in many western countries, erupting into riots in some. While these protestors' motivation was based on their conviction of liberty and human rights, believing that the government exaggerated the danger of the virus, the availability bias can reinforce this belief. To understand why, let us examine some statistic. Even in countries like the US, with COVID-19 deaths exceeding 500 000 (as of March 2021), this number represents less than 0.5 per cent of the US population. To many American protestors, particularly at early stage of the pandemic, they would say: 'No one that I know of caught and died of this virus', showing availability bias was at work again!

During 2020 governments all over the world rushed to get the COVID-19 vaccines developed as fast as possible. This created a fair bit of concerns about their efficacy and side effects. In order to alley people's fears, some people suggested using movie and sport celebrities to endorse the vaccines. On 10 January 2021, the BBC reported that the Queen (94) and Prince Philip (99) received their COVID-19 vaccination at Windsor Castle. Vaccinations of members of the Royal Family are usually private matters and seldom get reported in the mass media. The news was released at a time when the first phase of administering national-wide vaccination in the UK took place, which aimed to vaccinate 15 million people by mid-February. It is likely that the British Health Authority relied on the availability heuristic to persuade people that the vaccine was safe.

Herding heuristic

Herding heuristic
The tendency to make a choice by following others on the belief that the decision by the group should be right.

Let's now look at the **herding heuristic**. If you have visited Osaka, you might have tried a delicious street food, Takoyaki (Japanese squid balls). There are two Takoyaki stalls near Namba Station in Dotonbori opposite each other. Not knowing which one makes better squid balls, how are you to decide which stall to buy from? Yes, as a tourist, you would most likely follow the crowd and choose the stall with the longest queue. But while you are waiting at the queue, you observe that the other stall does not have one single customer, and that the owner of the other stall looks bemused, but *not* worried. Why is she not worried? It is because it all happens by chance. Without any information the first customer will make a random choice; the second will join her thinking that she knows which stall makes better Takoyaki. And the queue will get longer because of our herding mentality. The owner of the other stall knows this herding bias, it is not because of the quality of her squid balls, and later in the day luck will be kinder to her.

During the COVID-19 pandemic, we saw ordinary people rush to the supermarket to purchase large quantity of toilet paper in Australia (and in many other countries). Stocking up toilet paper illustrates the use of herding as a decision-making heuristic. Herding occurs when people follow others' behaviour believing (rightly or wrongly) that the decision of the majority should be correct.

While herding may have no or little rational basis, the herd mentality can be self-fulfilling. When people follow others to purchase large quantity of toilet paper, fearing its future shortage, this expectation (future toilet paper shortage) and the associated herding behaviour actually causes the emptying of toilet paper

[20] A. Tversky and D. Kahneman., 'Availability: A heuristic for judging frequency and probability', *Cognitive Psychology*, 1973, Vol. 5, No. 2, pp. 207–232.

shelves 'proving' that the fear of shortage is well justified. Herding is a very common phenomenon in finance, often leading to overshooting of share prices and exchange rate of a foreign currency, propagating financial instabilities. It is also interesting to note choosing toilet paper to hoard makes some economic sense. Have you wondered why people stocked up on toilet paper but not potatoes or meats during COVID-19, at least not in Australia? It is because the cost of stocking toilet paper is much lower than that of food. And the perceived benefit outweighs the small storage cost to safeguard against there being no toilet paper available if borders closures prevent toilet paper importation. There is also empirical evidence that as a heuristic device to save cognitive effort, herding is more common among people with lower cognitive ability. (Did you stock up on toilet paper during the COVID-19 pandemic?)

Understanding of herding behaviour can inform public policy formation and implementation. One fiscal problem facing many countries is tax non-compliance. Tax is the single most important source of revenue for the government to fund social services and capital infrastructure. Tax non-compliance can badly reduce this revenue base as seen in some south European countries. Beside the standard methods of increasing tax-compliance, such as process simplification and heavier penalties for non-compliance, studies show that sending behavioural based messages to taxpayers is cost effective to increase tax compliance.[21] For example, tax compliance rate was found to increase when people were told at the start of filling a tax return that a great proportion of their citizens in their state paid their taxes.[22]

When coronavirus started to spread to Europe and the US, many people were reluctant to wear masks, as they felt awkward and stigmatised when others did not wear them. A study utilising data collected in April and May 2020 in Germany found that compliance was not sufficient when mask wearing was voluntary, whereas a compulsory mask policy was more effective in gaining public acceptance and compliance.[23] This is because when others are all wearing masks, you will not be the odd one out. While there are other interesting findings from this study relating to the perceptions that compliant and non-compliant people have of one another, the increased acceptance of mask wearing under the compulsory mask policy can be explained by the herding mentality altering people's attitude and behaviour.

We now summarise the contributions of behavioural economics to understanding human behaviour. We observe that people possess various biases that influence them to make inconsistent decisions and at times erroneous judgements. These non-rational behaviours may have social and evolutionary purpose. However, since they violate the rationality principle, their prevalence poses serious challenge to the rational choice model. Another common critique towards the rational choice model is its assumption of unbounded rationality. According to one argument, the economic person simply does not possess the mental capacity to obtain large amounts of information and performs economic calculus before making a rational choice. For obscure problems, we just do not have a clue. To compensate, we develop heuristics to automate decision–making as the situation requires. Regardless of whether these heuristics will lead us to correct decisions, economists supplement their analytic toolbox with insights from cognitive and neural psychological research to better explain and predict behaviour. In the past decade, we have seen the establishment of behavioural insight teams in many government departments, testimony of the practical application of behavioural economics. By harnessing findings obtained from behavioural economic studies, governments are in a much stronger position to design and implement cost effective strategies and policies in public health, tax compliance, national saving and many other areas for the betterment of all.

---

21 World Bank Group, 'Behavioral insights for tax compliance', 2021, http://documents1.worldbank.org/curated/en/472181576511865338/pdf/Behavioral-Insights-for-Tax-Compliance.pdf (accessed 2 January 2021)

22 S. Coleman, 'The Minnesota Income Tax Compliance Experiment – State Tax Results', April 1996, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4419

23 C. Betsch, L. Korn, P. Sprengholz, L. Felgendreff, S. Eitze, P. Schmid, and R. Böhm, 'Social and behavioral consequences of mask policies during the COVID-19 pandemic'. *Proceedings of the National Academy of Sciences*, 2020, Vol. 117 No. 36, pp. 21851–21853, https://www.pnas.org/content/117/36/21851

You're the economist — Are our judges biased?

In presenting the behavioural biases and heuristics, you must have noticed that we referred to a lot of experimental studies. How pervasive are they in our daily life, as opposed to controlled experimental conditions? You might also wonder if when important decisions are made – for example, in our criminal justice system – whether our law enforcers would also be influenced by these biases and heuristics. We all care a lot about justice and hate to see unjust decisions being made in the courtroom. Take, for example, decisions that involves length of sentencing. It is not uncommon that when assessing an identical case, different judges would recommend very different prison sentences. For ordinary people to have trust in the judicial system, it is important that our judges are rational and not influenced by biases when making sentencing decisions and that sentences are proportionate to the severity of the crimes. If judges can make vastly different prison sentences, where are the sources of these differences? Some of these factors involve personal characteristics, such as personal experiences, and religious and political affiliations. Would these judges be also influenced by some of the irrationalities discussed in the chapter? This is what Enough and Mussweiler attempted to discover.

You may know that following a conviction, the prosecutor or the probation officer can demand a certain length of prison sentence; however, the judge is entirely free to make their sentence decision. In Enough and Mussweiler's 2001 experimental study,[1] the participants played the part of the criminal trial judges. They were presented with a hypothetical criminal case plus a sentence demand made by the prosecutor. For one group the sentence demand was two months and for the other it was 34 months. All participants were subsequently asked to decide on the length of sentence if they were presiding over the case. Do you think their decisions would be influenced by the sentence demands?

1 B. Enough and T. Mussweiler, 'Sentencing under uncertainty: Anchoring effects in the courtroom 1', *Journal of Applied Social Psychology*, 2001, Vol. 31, No. 7, pp. 1535–1551.

In summary

- Human behaviour is not always consistent and many cannot be adequately explained by the rational choice model.
- In dealing with uncertainty, human beings are observed to over-estimate miniscule probability for gain. They also display certainty-bias, loss-aversion, status quo and other biases.
- Due to lack of information and our limited mental capacity, human beings develop heuristics to provide a quick response to obscure problems, and to reduce complex problems to make them manageable.

Key concepts

Climate change
Carbon tax
Emissions trading scheme (ETS)
Cap and trade scheme
Abatement subsidies
Rational choice model
Nudging
Status quo bias
Sunk cost bias
Over-confidence bias
Decision heuristics
Anchoring heuristic
Availability heuristic
Herding heuristic

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1 Evaluate how successive Australian governments have proposed very different policy approaches to the task of tackling climate change and carbon emissions

- **Climate change** refers to long-term changes in the world's climate, which involves global warming and which is widely believed to be caused by human activity.
- A **carbon tax** is a tax levied per tonne of carbon emitted by producers. An **emissions trading scheme (ETS)** or **cap and trade scheme** is a government scheme requiring firms that produce emissions of pollutants to purchase tradeable permits giving them the right to do so. **Abatement subsidies** involve government payments to polluters in exchange for a reduction in their emissions of pollutants.
- From the economist's perspective, an ETS is the preferred instrument for tackling CO_2 emissions, which are strongly implicated in climate change.

2 Understand how economics incorporate social preferences in the rational choice model to explain behaviour

- Our concerns for others' welfare, social norms, sense of justice, fairness and reciprocity influence our decision-making. These social virtues facilitate transactions enhancing economic efficiency. Economists incorporate these social norms and values into rational choice theories to explain a wider range of behaviour.

3 Explain non-rational behaviour by combining psychology and economics

- People are observed to make inconsistent decision and at times erroneous judgement, exhibiting various behavioural biases (e.g., loss aversion bias, status quo bias and over-confidence bias).
- Due to their limited mental capacity and complexity of many economic problems, human beings develop heuristics to enable themselves to make quick decision and to deal with complex problems.
- By utilising insights obtained from behavioural economic research, governments are in a much stronger position to design cost effective public policies and implementation strategies.

Answers to 'You're the economist'

Are our judges biased?

A criminal trial is characterised by judgemental uncertainty. Faced with uncertainty, people would rely on simplifying heuristics for decision-making. Not unlike the experiment in which the participants were asked to guess the age of Gandhi at the time of his death, our criminal trial judges were influenced by the sentence demand, which acted as an anchor. Enough and Mussweiler found that for the group given a higher sentence demand, the average sentence was 28.7 months compared to 18.8 months for the other group. This is the anchoring heuristic at work in the courtroom. If you think that the participants' sentence decisions could be influenced by the anchor, YOU ARE THINKING LIKE THE ECONOMIST.

Study questions and problems

1 Using suitable diagrams, explain what happens to emission permit prices and the prices of carbon-intensive outputs when growth in the economy increases demand for carbon-intensive outputs but no technological change occurs.

2 Explain the significance of a situation where falling costs of producing clean energy cause the demand curve for permits in **Exhibit 10.2** to move so far to the left that it does not intersect the supply curve at all. What circumstances might lead to a subsequent shift of the supply curve?

3 An ETS is sometimes called a 'cap and trade' scheme. Explain what constitutes the 'cap' and what is meant by the 'trade'.

4 Why is it that economists dislike large-scale abatement subsidies as a way of controlling carbon emissions?

5 In Australia, charitable donations are tax deductible. The government found that a notable rise in donations to charities when the income tax rates were increased across all income brackets. Can the rational choice model explain this behaviour?

6 John is faced with the below gambles:
Gamble 1: ($100, 0.5; -$80, 0.5)
Gamble 2: ($12m, 0.5; -$10m, 0.5)
   a When asked if he will accept the gambles, John calculates the expected payoff. What are the expected payoffs for each gamble?
   b John decides to choose Gamble 1 but not 2, even though Gamble 2 has a much higher expected payoff. Explain John's decision.

7 If you own a car, you must buy compulsory third party (CTP) motor vehicle insurance. Many insurance companies sell the CTP insurance cover in Australia (e.g., Suncorp, RACQ, NRMA, to name a few). A few weeks before your CTP is up for renewal, you receive a promotion letter from NRMA offering you a $10 discount if you transfer your CTP insurance to NRMA. Suppose the cost of changing insurance company is very minimal, but you decide not to. Why not? What behavioural economic concept can be used to explain your decision?

8 In a study conducted in the US, the participants were asked to decide if there were more suicides or murders in New York. Most answered there are more murders when the fact is there are always more suicides. What bias does this research finding reflect?

Multiple-choice questions

1 A problem with an individual country's unilateral introduction of an ETS is:
   a it will make little difference to global emissions unless the country is a major contributor to emissions.
   b it will raise the cost of goods the country produces resulting in a loss of competitiveness in world markets.
   c it will close down all of the country's carbon-emitting industries.
   d a, b and c are correct.
   e a and b are correct.

2 An advantage of an ETS is:
   a it means that traditional fuels like coal and oil can continue to be used in the same quantities as before.
   b it encourages producers to install equipment that reduces carbon emissions.
   c like a carbon tax, it prevents the capping of emissions at a given level.
   d a, b and c are correct.
   e b and c are correct.

3 The supply curve of permits in an emissions trading market is:
   a perfectly elastic.
   b perfectly inelastic.
   c unitary elastic.
   d none of the above.

4 Which of the following types of carbon-reduction policies is generally preferred by economists?
- **a** Mandated use of green energy.
- **b** Subsidies for solar and wind power.
- **c** A tax on petrol and diesel fuel.
- **d** A carbon trading scheme.

5 A major shortcoming of abatement subsidies, like those included in the Abbott government's Direct Action Plan, is that:
- **a** they can't be used to achieve a particular level of pollution reduction.
- **b** the business community would rather see taxes used to reduce pollution.
- **c** they increase the costs of polluting firms' outputs.
- **d** large-scale pollution reduction would require government to vastly increase its revenue raising so that it could pay the required subsidies.

6 You paid $15 for a movie ticket. To your disappointment, after watching it for 15 minutes you realised the movie was boring. But you decided to sit through the movie because you didn't want to waste the money. What type of bias did you commit?
- **a** Herding bias.
- **b** Anchoring bias.
- **c** Sunk cost bias.
- **d** Status quo bias.

7 You live in Rockhampton, central Queensland. After the once-in-a-lifetime cyclone destroyed hundreds of houses in northern Queensland, Mr Smith your next-door neighbour decided to increase the insured amount of his current home and content insurance. What type of bias does this behaviour illustrate?
- **a** Framing bias.
- **b** Sunk cost bias.
- **c** Anchoring bias.
- **d** Availability bias.

8 In an experiment, Kahneman and Tversky asked two groups of high school students to estimate an arithmetic problem within five seconds.
Group 1 estimate the product of 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1.
Group 2 estimate the product of 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8.
It is found the median estimate for group 2 is 512 and 2250 for group 1. The correct answer is 40 320! The result shows that students made much greater under-estimation for ascending sequence (Group 2) than descending sequence (Group 1). What decision heuristic did the students use in their estimation, which explains these systemic errors?
- **a** Herding heuristic.
- **b** Anchoring heuristic.
- **c** Availability heuristic.
- **d** None of the above.

PART 4

MACROECONOMIC FUNDAMENTALS

11 Measuring the size of the economy

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand the concept of gross domestic product.
2 Understand how gross domestic product is measured.
3 Appreciate the shortcomings in using gross domestic product as a measure of the economic welfare of a country.
4 Learn the difference between nominal and real gross domestic product.
5 Develop an understanding of the nature of the national saving and investment relationship in an open economy.

Accurately measuring the size and performance of the economy is a vital and critical task. Prior to 2020 it had been three decades since Australia's last recession in 1990–91. But how do we measure the size of an entire economy and then determine whether it's growing or contracting from one period to the next? And what role does the government have to play in the nation's macroeconomic performance?

These days, governments are – rightly or wrongly – considered by the community to have an important influence on macroeconomic outcomes, such as the rate of economic growth, the level of unemployment, inflation, interest rates, the value of the exchange rate, and so on. The outcomes of elections are often significantly determined by the electorate's judgement of the current and likely future state of the economy and the macroeconomic management abilities of those running for office. Indeed, in the run-up to the 2019 federal election, a major electoral issue was again (as it was in 2013, 2016 and most other federal elections since the 1980s) the relative macroeconomic management credentials of the two major parties.

Getty Images Plus/iStock/roberthyrons

How then should we assess the macroeconomic performance of the economy? How should one correctly measure the size of the national pie? What data would one seek to determine how well the economy is doing? The answer requires a basic understanding of *national income accounting*, which is the system used to measure the aggregate income, production and expenditures of a nation. It provides a valuable indicator of an economy's performance. As an exercise, you can visit the website

of the Australian Bureau of Statistics (ABS) to find out the size of the Australian economy in 2019–20 and its growth from 2018–19 to 2019–20.

Before the Great Depression in 1929–33, there were no national accounting procedures for estimating the data required to assess an economy's performance. To provide accounting methodologies for macroeconomic data, Simon Kuznets published a small report in 1934 titled *National Income, 1929–32*. For his pioneering work, Kuznets has since been dubbed 'the father of GDP'. Also deserving of mention in terms of developing the system of national accounts is Colin Clark (who lived most of his adult life in Australia).

National income accounting serves the nation in a way similar to how accounting serves a business or household. In each case, the use of proper accounting methodology is vital to identify economic problems and formulate plans for achieving goals.

1 The concept of gross domestic product: the basics

The most widely reported measure throughout the world of a nation's economic performance is **gross domestic product (GDP)**. Gross domestic product is the market value of all final goods and services produced within a nation's geographic borders during a period of time, usually a quarter or a year. GDP, therefore, excludes production abroad by the nation's businesses.

Gross domestic product (GDP) The market value of all final goods and services produced within a nation's geographic borders during a period of time, usually a quarter or a year.

Another measure whose use is gaining increasing international traction is that of **gross national income (GNI)** (conceptually identical to – but estimated differently from – what is also called *gross national product*). Gross national income is the income accruing to a country's residents from the production of all final goods and services during a period of time, no matter where the goods and services are produced in the world. For example, Australia's GNI includes BHP's profits on its foreign operations, but GDP does not. On the other hand, until it ceased operations in October 2017, Australia's GNI excluded foreign-owned Toyota Australia's repatriated profits from its Australian car manufacturing operations, while GDP included it.

Gross national income (GNI) The income accruing to a country's residents from the production of all final goods and services during a period of time, no matter where in the world the goods and services are produced.

The main reason for the emphasis on GDP is that it is the measure more relevant to the level of domestic economic activity and therefore, to the generation of domestic jobs. A country's GDP and GNI may differ only slightly or very significantly, depending on the importance of international trade and other international transactions to the country. In the case of many countries that are very open to trade, GNI can be quite different from GDP. For instance, in the third quarter of 2019, the GDP of the Chinese territory of Hong Kong was HKD$722.8 billion and its GNI was around HKD$764.1 billion; that is, GNI was larger by HKD$41.3 billion, which was equal to 5.7 per cent of the GDP of that quarter. On the other hand, Ireland's GDP is higher than GNI. Even though Ireland produces a lot of income per person in the country, but GNI shows that less of that income stays in the country. Ireland employs a lot of foreign investment and, the income from these investments repatriated to those foreign investors needs to be deducted.

One advantage of GDP is that it avoids the 'apples and oranges' measurement problem. If an economy produces nothing but 1000 apples one year and nothing but 1000 oranges the next, can we say that the quantity and value of output have changed in any way? To answer this question, we need to be able to evaluate the relative value of apples and oranges to society. To overcome this issue, GDP uses dollars to measure value, rather than the quantities, such as the number of cars, computers, aluminium sheets, heart transplants, legal cases, haircuts, toothbrushes, rainwater tanks and so on produced. The market-determined dollar value of each good and service produced in the economy is used, thus establishing their relative value, and also enabling production to be aggregated and compared across time.

GDP also requires that we give the following two points special attention.

Point 1: GDP counts only new domestic production

National income accountants calculating GDP exclude transactions in two significant areas: *second-hand transactions* and *'non-productive' financial transactions*.

Second-hand transactions

GDP does not include the sale of a used car or the sale of a house constructed some years ago. Such transactions are merely exchanges of ownership of previously produced goods and are not *current* production of new goods that add to the existing stock of cars and homes. However, the sales commission paid to the salesperson for selling a used car or a home produced in an earlier period counts in current GDP. This is because the salesperson performed a service during the current period, and it is considered newly produced.

'Non-productive' financial transactions

GDP does not include purely financial transactions, such as trading in financial securities (like stocks and bonds); although, again, any brokerage charged would be included, as this represents a current period service provision. GDP also does not include the dollar value of **transfer payments**. A transfer payment is a government payment to individuals that does not represent a payment in exchange for currently produced goods or services.

Transfer payment
A government payment to individuals that does not represent a payment in exchange for currently produced goods or services.

Social welfare payments, veterans' benefits, other types of pensions and unemployment benefits are transfer payments. These transactions are not included in GDP because they do not represent the production of any new or *current* output. Similarly, stock market transactions represent only the exchange of certificates of ownership (stocks) or indebtedness (bonds) and not actual new production.

Point 2: GDP counts only final goods

GDP counts *only final* goods and services because including *all* goods and services produced would inflate GDP by *double-counting*; that is, many items would be counted more than once. **Final products** are finished goods and services produced for the ultimate user. National income accountants must take care not to include **intermediate products**, which are goods and services used as inputs into the production of final goods and services but that are not produced for consumption by the ultimate user.

Final products
Finished goods and services produced for the ultimate user.

Intermediate products
Goods and services used as inputs for the production of final products.

Goods can be considered as final goods or intermediate goods depending on the use or user. Suppose a wholesale distributor sells glass to a professional house builder who uses it for the windows of a house that he is building. In this situation, the glass is an intermediate good used in the production of the house. The value of the glass is included as part of the final selling price of the house, and it is the value of the house that is counted in GDP.

As another example, suppose the wholesale distributor sells glass to a hardware store that subsequently sells the glass to a customer who uses it for her house renovation or window repairs. GDP does not include the first transaction because the hardware store is not the final user. When the customer buys the glass from the hardware store to renovate their home or repair their broken window, the final purchase price of the glass to the ultimate user is added to GDP as consumption expenditure.

The concept of value-added

The idea of economic *value-added* is particularly important from a national accounting perspective. For example, suppose a retail store sells a woollen suit for $420 ($20 of which is paid to the government in sales tax), which it buys from the manufacturer for $300. At earlier stages of the production chain, the manufacturer purchased the raw material from its fabric supplier for $150 and the fabric supplier, in turn, bought the raw wool for $40 refer to **Exhibit 11.1**). The sum of expenditures at the various stages of production is $910 ($420 + $300 + $150 + $40), but the value of final output (i.e., its price to the final consumer), which embodies the value-added of each stage of production plus the sales tax, is only $420. As you will note, this is simply another example of the principle of only counting final goods and services when measuring GDP.

However, to further illustrate the idea of value-added, consider another perspective on the above example. The retailer paid the manufacturer $300 for the suit and received $400 for it (after the sales tax had been deducted). This $100 represents the value added by the retailer and is made up of wages, salaries and supplements of employees (e.g., delivery and salespeople) plus the retailer's gross operating surplus, which covers an

Exhibit 11.1 Concept of value-added

Firm	Value of the product	Value-added	Contribution to GDP
Farmer	Value of raw wool = $40	= $40	
Fabric supplier	Value of fabric = $150	= ($150 – 40) = $110	
Manufacturer	Value of jumper = $300	= ($300 – 150) = $150	
Retail stores	Value of jumper from retailer = $420	= ($420 – 300) = $120	GDP at market prices = ($40+110+150+120) =$420
Tax	=$20		GDP at factor cost = $420 – 20 = $400

allowance for interest payments on borrowings, plus profit. The profit is the income of the owners of the retail store, and the interest is the income of any lenders of funds to the store. Thus, all if the value added at this stage in the production process accrues to somebody as income.

This is also true of all other stages and, by adding the value added for all the production stages (here $100 + $150 + $110 + $40), we get what is called **GDP at factor cost**. In this example, it is equal to $400. However, looking at it from the expenditure side, **GDP at market prices** is $420 since this is how much the end buyer paid for the suit. The difference between the two measurements is accounted for in the national accounts by **indirect taxes less subsidies**, here $20.

GDP at factor cost
A measure of GDP arrived at by adding all the incomes of all factors of production (including labour, land and capital).

GDP at market prices
A measure of GDP arrived at by valuing GDP produced at the prices at which the goods and services sell. The difference between this and GDP at factor cost is the existence of government indirect taxes less subsidies.

In summary

- **GDP** is the market value of all final goods and services produced within a nation's geographic borders.
- **GNI** (or GNP) is the income accruing to a country's residents from the production of all final goods and services, no matter where the goods and services are produced in the world.
- GDP only includes newly-produced final goods and services.
- **GDP at factor cost** is arrived at by adding together all the 'value-added' at each stage of the production chain; to this are added government 'indirect taxes less subsidies' to arrive at **GDP at market prices**.

Indirect taxes less subsidies
Government taxes levied on the production and/or sale of goods and services sold, less any subsidies paid to business from government. Examples include general sales taxes (like Australia's GST), excise taxes and customs duties.

2 Measuring GDP

The production of goods and services in an economy consists of many interrelated activities all fitting together. These activities include consumers earning and spending money and businesses earning and spending money, and such activities occur in markets for products and markets for resources. All these activities are a little like a jigsaw puzzle where all the various pieces need to fit together.

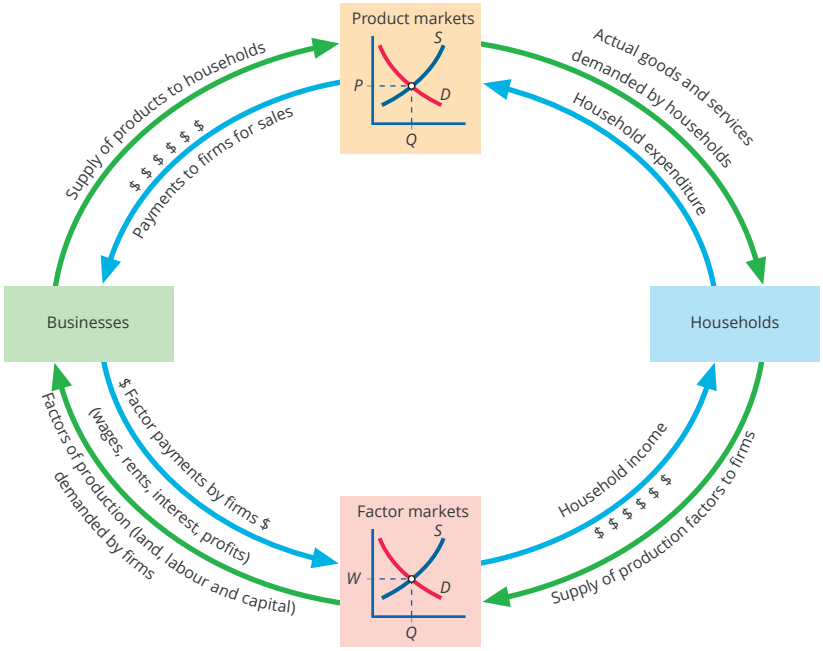
One way to understand how it all fits together is to use a simple macroeconomic model called the *circular flow model*. The **circular flow model** shows the flow of products from businesses to households and the flow of productive resources from households to businesses or firms. In exchange for these products and resources, money payments flow between businesses and households. **Exhibit 11.2** shows the circular flow in a simple hypothetical economy with no government, no financial markets and no foreign trade. In this ultra-simple pure market economy, only households and businesses make decisions.

Circular flow model
A diagram showing the flow of products from businesses to households and the flow of resources from households to businesses. In exchange for these products and resources, money payments flow between businesses and households.

The circular flow model

The upper half of the diagram in **Exhibit 11.2** represents *product markets* where households exchange money for goods and services they want and which are produced by firms. The *supply of products to households* arrow represents all finished goods and services produced, sold and delivered by firms to consumers. The *goods and services demanded by households* arrow represents household demand for the products produced by the firms. Expenditures by households on these products translate into sales revenue payments to the firms in return for producing the goods and services demanded by the households. Notice that the box labelled *product markets* contains a supply and demand graph. This represents the idea that the forces of supply and demand in individual markets determine the price and quantity of each product exchanged.

Exhibit 11.2 **The basic circular flow model**



In this simple economy, households spend all their income in the upper loop and demand consumer goods and services from businesses. Businesses seek profits by supplying goods and services to households through the product markets. The market supply and demand model determine prices and quantities in individual markets. In the factor markets in the lower loop, resources (land, labour and capital) are owned by households and supplied to businesses that demand these factors in return for money payments, such payments constituting the income of the household sector. The forces of supply and demand determine the quantity of factors provided and their prices – for example, the quantity of labour supplied and associated wages and salaries. Overall, goods and services and factors of production flow clockwise and the corresponding payments flow anti-clockwise.

The bottom half of the circular flow diagram consists of factor markets in which firms demand the materials, land, labour, capital and entrepreneurship – the factors of production – needed to produce the goods and services sold in the product markets. In our hypothetical economy, it is assumed for simplicity that households own all the factors of production so firms, therefore, must purchase their needs directly from households. The factors of production demanded by firms and the supply of production factors to firms arrows represent this flow of production factors from households to firms. In return for these production factors, firms provide factor payments of wages, rents, interest and profits to households, which translate into the income of the household sector. As in the product markets, market supply and demand determine the prices of factors and the factor quantities supplied.

Our simple model also assumes that all households live from hand to mouth, in that households spend all the income they earn in factor markets on products. Households, therefore, do not save. Likewise, all firms spend all their income earned in the product markets on purchasing resources from the factor markets. In other words, they don't invest in plant and equipment.

The simple circular flow model fails to mirror the real world. But it does aid in understanding the fundamental relationships among product markets, factor markets, the flow of money payments and the basic concepts behind GDP measurement. In a moment, we will turn our attention to a more realistic model.

Flow versus stock

The arrows in **Exhibit 11.2** are flows, rather than stocks. A **flow** is a rate of change in a quantity during a given time period. The amount of steel produced per month, the number of computer games purchased per day, the amount of income earned per year and the number of litres of water pouring into a bathtub per minute are examples of flows. Flows are always measured in units per time period, such as tonnes per month, billions of dollars per year or litres of water per hour.

Flow
A rate of change in a quantity during a given time period, measured in units per time period, such as dollars per year. For example, income and consumption are flows that occur per week, per month or per year.

A **stock** is a quantity measured at one point in time. A company's assets, the amount of money in a cheque account and the amount of water in a bathtub are examples of stocks. A stock is measured in things like tonnes, dollars, litres and so on at a given point in time, while a flow is defined as something (e.g., tonnes, litres, kilograms or dollars) per unit of time (e.g., per day, hour or month).

Stock
A quantity measured at one point in time, such as a company's assets or the amount of money in a cheque account.

The essential difference between a stock and a flow is that, at a moment in time, a stock will exist, but the flow can be zero. As the unit of time over which flow is being measured gets smaller and smaller, the measured flow gets smaller, and at a particular instant in time, the flow goes to zero.

An important point here is this: all measurements in the circular flow model are flows. As such, they tell us what has been added during some given period to the total stocks of goods and services available in the economy, but they do not represent the total stocks themselves. Consumption expenditures, business production, wages, rents, interest payments and profits are flows of money arising from newly-produced goods and services in the period in question (e.g., a month, a quarter, a year). These newly produced goods and services add to the total level of already available stocks of products in the economy.

A four-sector circular flow model

Exhibit 11.3 puts all the puzzle pieces together. It presents a more complex circular flow model by adding two additional sectors to the model: government and foreign markets. Also added to the model are financial markets, which play the vital role of facilitating the flow of saving throughout the economy. Other than households spending their incomes on the output of firms, these additions add three *leakages* from the amount of income paid to households that reduce their subsequent spending on the production of domestically based firms.

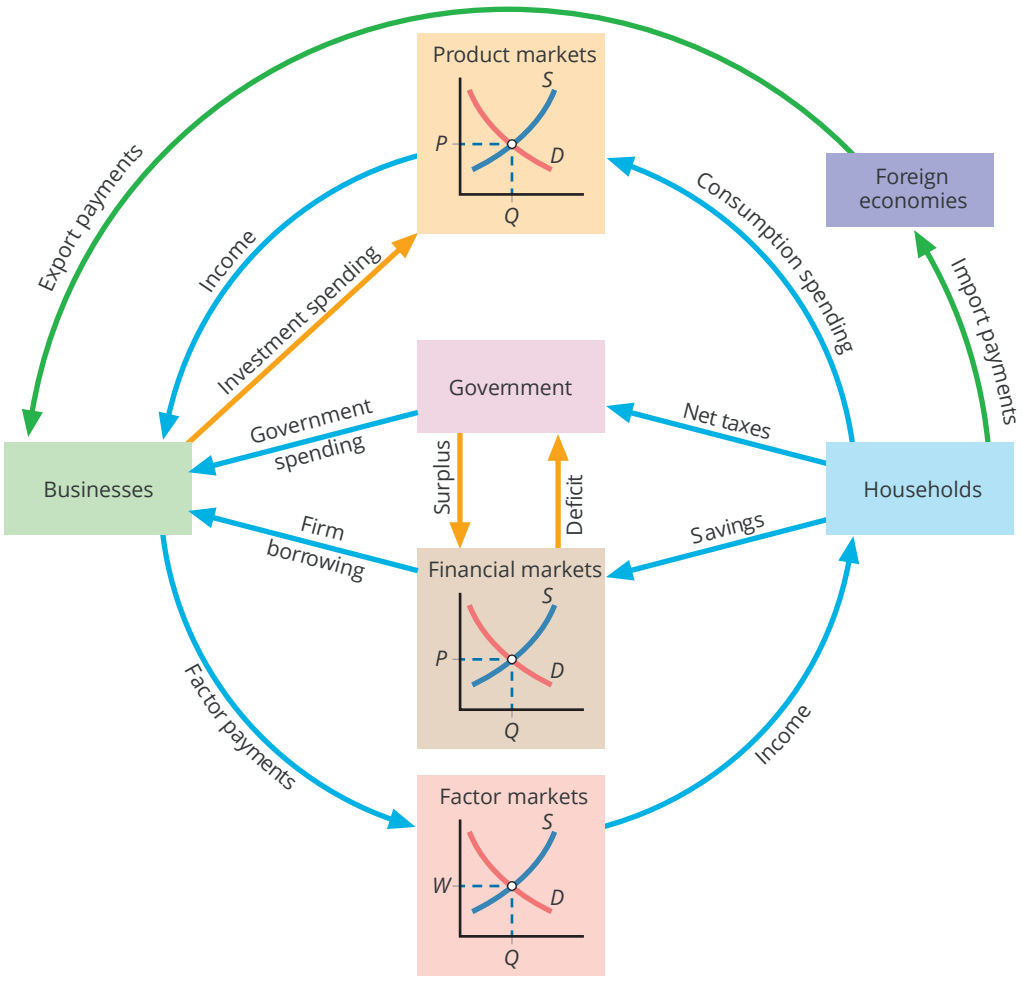
From the income of the households, first, part of a households' income is saved. Second, part of it is taxed by the government. Third, part of the income is spent on imports (foreign production). On the other hand, there are three sources of spending injections for firms' output other than from households. First, firms invest in new plant and equipment purchased from other firms, plus they incur voluntary or involuntary spending on inventories. Second, the government uses taxes raised for expenditures on goods and services produced by firms. Third, foreigners purchase exports from firms.

The expenditure approach to measuring GDP

How does the government actually calculate GDP? One way is to use the expenditure approach to measure total spending flowing through product markets in the circular flow diagram. The **expenditure approach** measures GDP by adding all the expenditure on final products during a period of time. An alternative method is to use the

Expenditure approach
The national income accounting method that measures GDP by adding all the spending on final products during a period of time.

Exhibit 11.3 The circular flow model of an open economy



This exhibit presents a circular flow model for a real-world economy (e.g., Australia's) that has a financial sector and a government sector and engages in international trade. The model includes links between the product and factor markets in the domestic economy and the financial markets, government and foreign economies. For simplicity, only dollar payment flows are shown.

Income approach The national income accounting method that measures GDP by adding all incomes, including compensation of employees, rents, net interest and profits.

income approach. This approach measures GDP by combining all incomes of all factors of production, including wages, rents, interest and profits – that is, all the income generated in the production process. Conceptually, both approaches should yield the same result. However, in practice, measurement issues result in small differences between the two estimates.

A little over two months after the end of each quarter, the ABS publishes calculations of GDP on both the expenditure and income bases. In practice, GDP estimates based on a third approach that involves valuing the quantities of products produced (the output approach, also known as the production-based approach) come a little later. Although conceptually each should give the same figure, all three measures – expenditure, income and output estimates – are calculated independently and in practice yield somewhat different estimates due to the different measurement approaches.

Exhibit 11.4 shows 2018–19 GDP using the expenditure approach, which breaks down expenditures into four components. The data in this table show that all production in the Australian economy is ultimately purchased by spending from households, businesses, government or foreigners. Let's discuss each of these expenditure categories.

Exhibit 11.4 **Australian gross domestic product using the expenditure approach, 2018–19**

Item	Amount ($ billion)	% of GDP
Household consumption (*C*)	1064.1	56.3
Gross private domestic investment (*I*)[1]	350.5	18.4
Government consumption	361.1	19.1
Government investment[1]	95.0	5.0
Total government expenditure (*G*)	456.1	24.1
Gross national expenditure (GNE = *C* + *I* + *G*)	1870.7	98.8
Net exports (*X* – *M*)	22.4	1.2
Exports	419.0	22.1
Imports	396.6	20.9
Gross domestic product (GDP = *C* + *I* + *G* + *X* – *M*)[2]	1893.1	100

[1] In the table, the total change in inventories in 2018–19 listed in Table 36 of ABS Catalogue 5206 was simply allocated to government investment and gross private domestic investment (GPDI) on a pro-rata basis.
[2] In practice, there is always a statistical discrepancy among the estimates of GDP, GNE and Net Exports. For our purposes, the statistical discrepancy calculated by the ABS for 2018–19 was distributed proportionally across the expenditure categories in GNE to ensure consistency of GNE with GDP and Net Exports.

Source: Australian Bureau of Statistics, Cat. No. 5206, Table 36.

Household consumption expenditures (*C*)

The most significant component of GDP in 2018–19 was $1064.1 billion for *household consumption expenditures*, represented by the letter *C*. Household consumption expenditures comprise total spending by households for durable goods, non-durable goods, and services. Durable goods include such items as cars, appliances and furniture because they last longer than one year and can be used more than once. Food, clothing, soap and petrol are examples of non-durables because they are considered used up or consumed in less than a year. There is an element of arbitrariness involved here. For instance, the authors would each admit to having shirts that are older than some cars on the road. Nevertheless, by convention, shirts are non-durable, and cars are durable goods.

The services category, which these days are the largest category, includes recreation expenditures (e.g., movies, football games, the theatre), legal advice, medical treatment, education, personal services (e.g., haircuts, beauty visits and the like), car servicing and so on – basically, any transaction that involves the buyer receiving a benefit that is not in the form of a tangible object (i.e., one that you pay for but cannot drop on your foot!).

The total consumption of goods and services by households (*C*) in a modern developed economy is by far the largest component of GDP, and in Australia comprised around 56 per cent of the total production of all goods and services in 2018–19.

Gross private domestic investment (*I*)

In 2018–19, $350.5 billion was spent for what is officially called *gross private domestic investment* (*I*). This figure includes 'gross' (i.e., all spending, regardless of whether the spending was to replace equipment or buildings that were wearing out), 'private' (not government) and 'domestic' (not foreign) spending by businesses for investment. At this point, it is worth emphasising the distinction between *gross investment* and *net investment*.

Gross investment Any spending that maintains or increases the capital stock in the country.

Net investment Only that spending that actually increases the nation's capital stock.

Gross investment is any spending that maintains or increases the capital stock in the country. **Net investment** is only the spending that actually increases the nation's capital stock. As an example, suppose the owner of a trucking company spends $10 million on the company's fleet of trucks but $8 million of that is for repairs and other maintenance to keep the current stock of trucks in good shape. The other $2 million is for actual additions to the fleet. In this case, gross investment would be $10 million, while net investment would be $2 million. It is gross investment that is counted in GDP.

Gross private domestic investment is the sum of two components: (1) fixed investment expenditures for newly produced capital goods, such as commercial and residential structures, machinery, equipment and tools; and (2) changes in private-sector inventories, which is the net change in the value of unsold finished products, unfinished products and raw materials purchased by firms but as yet unused in production.

Why include changes in inventories in a particular period as part of measured GDP in that period? When businesses have more on their shelves at the end of the current year than the previous year, more new production has taken place than is consumed during the current year. However, these products are still part of the production for the current year and are therefore included in GDP. Conversely, a decline in inventories from one period to the next means that measured GDP would be less than total dollar sales in the period in question because households would have spent more on products than firms actually produced during that period.

It is perhaps easy to see why a run-up in inventories is not included in the consumption spending component of GDP (after all, the production has not yet been purchased and consumed), but why include it as part of investment spending? The reason is that this is simply by convention. The increase in unsold inventories represents production during the period in question. This production is considered to have been purchased by the firms concerned, and it is in this way that the total value of production must always be equal to the total value of spending in any period. The production is either purchased by somebody (either a domestic resident or a foreign buyer) or assumed to have been implicitly purchased by the firm in question (at least until it is sold in some future period).

Note that gross private domestic investment is simply the national income accounting category for 'investment' defined in **Chapter 2**. The only difference is that investment in **Exhibit 2.4** in **Chapter 2** was in physical capital rather than the dollar value of capital.

At this point, we caution you concerning the use of the terms 'investment' and 'capital'. In everyday usage, any time a person or a company uses some of their accumulated savings – or borrows money – to buy shares, fixed-interest securities, a piece of existing real estate or the like, they will usually say that they are 'investing'. Similarly, you will often hear of people 'needing capital' for some venture or other. Here they mean they need some financing to fund their venture.

However, in economics, we use the term 'investment' to refer to spending on physical products that are intended to be used as inputs in production processes. The reason for this is that, in everyday parlance, much of what is commonly referred to as investment is merely the transfer of existing assets from one person to another – as occurs, for instance, when someone buys some shares from someone else on the share market. To avoid confusion, from now on, whenever we use the word 'investment', it will be meant in a sense used in this discussion of the national accounts. Alternatively, we will use the term 'financial investment' when we are referring to the acquisition of financial assets.

Similarly, the term 'capital' – or 'capital stock' – is used in economics and in this book to refer to the existing stock of buildings, factories and machinery available to be used in production processes. In terms of the above discussion on flows and stocks, investment is a flow while capital is a stock. Private and public investments will increase the capital stock in the country.

Now we will take a closer look at gross private domestic investment. Note that the value of newly constructed residential housing is included in the $350.5 billion spent for private-sector investment in 2018–19. A new factory, warehouse or manufacturing robot is undoubtedly a form of investment, but why include residential housing as business investment, rather than consumption by households?

The explanation is that a new home is considered investment because it provides services over time into the future, which the owner can choose to rent out for the financial return. For this reason, all newly produced housing is an investment, whether the owner rents it out or occupies the property. In addition, while a motor car purchased by a household is regarded as durable consumption when it is purchased by a firm to use in its business, it is viewed as an investment. It is part of fixed investment – 'fixed' not in the sense that it is stationary but in the sense that it continues to exist from one period to another, whereas inventory investment may be sold in the next period.

Also note that, again, the important issue is the use to which the purchase is being put. A motor car purchased by a company to assist it in producing further output (even if it is just to carry the CEO around) is an investment. On the other hand, the same motor car purchased by a household to take the children to netball and school is considered consumption expenditure because it is not being used to assist in the further production of output.

Government consumption and gross public investment expenditures (*G*)

This category includes the value of all goods and services that the government (including at federal, state and local levels) purchases in a particular period. For example, spending on the salaries of police, the defence forces and other government employees, as well as spending on such things as paper, stationery, garbage disposal and New Year's Eve fireworks displays enters the GDP accounts as government consumption at the prices the government pays for them.

Also, the government spends money on investment-type goods, such as highways, bridges and government buildings. In 2018–19, federal, state and local government consumption expenditure and gross investment (*G*) was $456.1 billion (361.1 + 95.0). You should also note that consumption expenditures and gross investment of state and local governments far exceed those of the federal government. This is because much of the federal government's 'spending' is transfer payments either to individuals or to other tiers of government.

It is vital to understand that consumption and gross investment expenditures by the government (what we mean by *G*) exclude transfer payments because, as defined at the beginning of the chapter, they do not represent newly-produced goods and services. Instead, transfer payments, such as pensions, unemployment benefits, benefits from other social programs and federal government grants to state and local government, are simply a transfer of purchasing power from one section of the community to another via government and not in exchange for a newly produced good or service.

Net exports (*X* – *M*)

The last GDP expenditure item is *net exports*, expressed in the formula (*X* – *M*). *Exports* (*X*) are expenditures by foreigners for domestically produced goods and services. *Imports* (*M*) are the dollar amount of a nation's purchases from producers in other countries. Because we are using expenditures for domestically produced output to measure GDP, one might ask why imports are subtracted from exports. The answer concerns how the government collects data from which GDP is computed. Spending on imports is not subtracted when spending data for consumption (*C*), investment (*I*) and government spending (*G*) is reported.

When these three components of GDP are added, the result is what is referred to as **gross national expenditure (GNE)** = *C* + *I* + *G*. This is the total spending in the Australian economy on goods and services, irrespective of whether that spending was on domestic production or imported products. Thus, GNE will overstate the value of expenditures on domestically produced products by the extent of imports (M). In 2018–19, Australian GNE was $1870.7 billion.

Gross national expenditure (GNE)
The sum of consumption, investment and government spending in the economy in a quarter or a year (= C + I + G), irrespective of whether the spending was on domestic production or imports.

On the other hand, GNE will not include the expenditure by foreigners on Australian production. Thus, GNE will understate expenditure on Australian production by the extent of exports (X). To get GDP, we, therefore, need to add net exports ($X - M$) to GNE.

The 2018–19 GDP outcome is obtained by subtracting $396.6 billion in imports from GNE ($C + I + G$) and adding in the $419.0 billion in exports to yield GDP of $1893.1 billion.

A formula for GDP

Given the above discussion, using the expenditure approach, GDP is expressed mathematically as:

$$GDP = C + I + G + (X - M)$$

For 2018–19, in billions of dollars (see **Exhibit 11.4**):

$$1893.1 = 1064.1 + 350.5 + 456.1 + (419.0 - 396.6)$$

This simple equation plays a central role in macroeconomics. It is the basis for analysing many macroeconomic problems and formulating macroeconomic policy. When economists study the macroeconomy, they can apply this equation to predict the behaviour of the major sectors of the economy: consumption (C) is spending by households; investment (I) is gross investment spending by firms and households; government consumption and government gross investment expenditures (G) is spending by the government; and net exports (X – M) is net spending by foreigners.

You're the economist How much does Serena add to GDP?

Serena works part-time at a small toy manufacturer where they make and sell toys. She earns an annual wage of $40 000, but she was unemployed for three months during the year. Serena received $2500 as unemployment benefits from the government. And also, she bought her mum's old car and paid $3000 to her mom. She made and sold 1000 small toys at $25 per toy during the year. Using the expenditure approach, how much has Serena contributed to GDP during the year?

GDP in other countries

Exhibit 11.5 provides GDP comparisons among selected countries for 2016. For example, the US has the highest GDP, which is 65 per cent larger than China, the world's next-biggest economy, and about 92 times that of New Zealand. As the emergent economic superpower of the 21st century, China has now overtaken Japan as the world's second-largest economy, dropping Japan to third. Of course, when adjusted for relative population sizes, US and Japanese GDP per capita are each much greater than that of China. At current relative rates of GDP growth, the Chinese economy may well overtake the US economy in absolute size within the next 10–15 years, but it is probably pretty unlikely that it will overtake the US on a per capita basis during this century. Australia's GDP is about 7 per cent of that of the US.

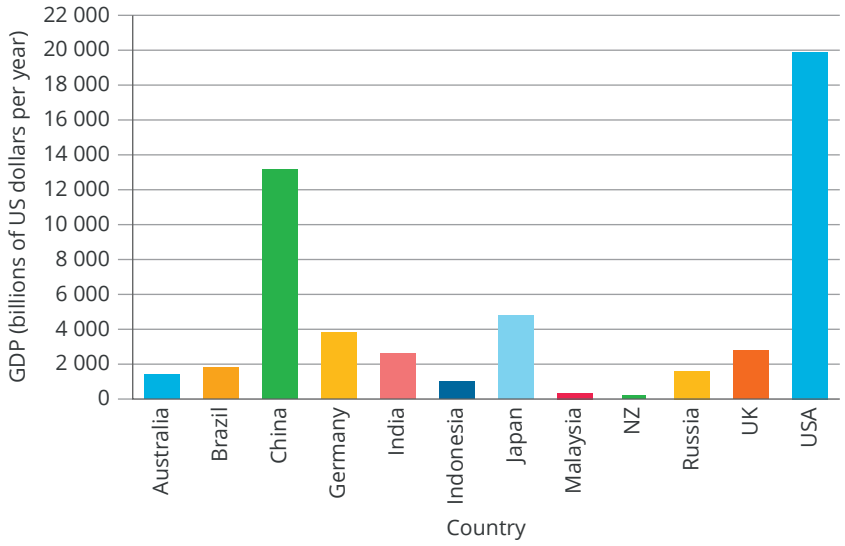
Other national accounts

In addition to GDP, several other national accounts are useful for studying the economy as a whole. We will now have a brief look at some of these.

Net domestic product (NDP)

It can be argued that depreciation should be subtracted from GDP. Recall that GDP is not entirely a measure of newly produced additional output because it includes the value of the investment in plant and equipment required to replace the capital that has worn out over time in the production process. The output measure that adjusts

Exhibit 11.5 An international comparison of GDP, 2018



Source: World Bank, https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?view=char. Data are price-adjusted and expressed in USD.

for this depreciation in plant and equipment is **net domestic product (NDP)** and is defined as gross domestic product minus depreciation of the capital worn out in producing output. Stated as a formula:

NDP = GDP – depreciation (also known as consumption of fixed capital)

Net domestic product (NDP) Gross domestic product minus depreciation of the capital worn out in producing output.

Consumption of fixed capital is the official government term for an estimate of the depreciation of plant, equipment and structures. Because it is impossible to measure depreciation accurately, an estimate is entered. In 2018–19, $343.9 billion was the estimated amount of GDP attributable to depreciation during the year, leaving NDP of $1549.2 billion.

Gross – and net – national income

Earlier, gross national income (GNI) – or gross national product – was defined as the income accruing to a country's residents from the production of all final goods and services during a period of time, no matter where the goods and services are produced in the world. To obtain GNI, the total income generated in the production of GDP is adjusted for the net primary income (interest and dividends) paid to or received from foreigners.

The total income generated from the production of GDP should conceptually equal GDP measured using the expenditure approach, but, since the two are measured differently, there is always a statistical discrepancy. For our purposes here, we have ignored this statistical discrepancy and simply used GDP measured on the expenditure basis and as reported in **Exhibit 11.4**.[1] In 2018–19, Australia's net primary income payable to foreigners was $61.6 billion, so an estimate of Australia's GNI in 2016–17 was $1831.5 billion.

Net national income Also known as net national product; measures the income accruing to a nation's residents from the production of final goods and services net of an allowance for capital depreciation.

GNI can be further adjusted to allow for the consumption of fixed capital to yield **net national income** (or net national product), which would represent the income accruing to a nation's residents net of depreciation – subtracting the $343.9 billion depreciation yields a figure of $1487.6 billion for net national income.

[1] In 2018–19, Total Factor Income was $1698.2 billion and Indirect Taxes less Subsidies was $190.8 billion to yield GDP measured on an income basis of $1889 billion. Since GDP measured on an expenditure basis was $1893.1 billion, the ABS-determined statistical discrepancy in 2018–19 was, therefore, $4.1 billion.

Gross and net disposable income

From gross national income, net transfers overseas (mainly foreign aid and pensions paid to recipients living abroad) can be subtracted (in 2018–19 this amounted to $0.8 billion) to obtain what the ABS refers to as **gross disposable income**, which in 2018–19 was $1830.7 billion. Again, depreciation can be subtracted from this to yield 2018–19 **net disposable income** for Australia of $1486.8 billion. **Exhibit 11.6** presents and derives the values for some of these aggregates for 2018–19.

Gross disposable income Obtained after subtracting net transfers overseas (mainly foreign aid and pensions paid to recipients living abroad) from gross national income.

Net disposable income Gross disposable income less an allowance for capital depreciation.

Exhibit 11.6 **Calculations of various national macroeconomic aggregates for 2018–19**

		Amount (billions of dollars)
1	Gross domestic product (expenditure basis)	= 1893.1
2	Depreciation (also known as consumption of fixed capital)	= 343.9
3	Net domestic product	= 1549.2 [(1) − (2)]
4	Net income payable to foreigners	= 61.6
5	Gross national income (also known as gross national product)[1]	= 1831.5 [(1) − (4)]
6	Net transfers to foreigners	= 0.8
7	Gross disposable income	= 1830.7 [(5) − (6)]

[1] As noted in the text, for this table, the 2018–19 statistical discrepancy of $4.1 billion between GDP measured on an expenditure basis ($1893.1 billion) and the total income generated from GDP ($1889.0 billion) was disregarded, and GDP measured on an expenditure basis was used. The above figures for gross national income and gross disposable income are therefore $4.1 larger than those reported in Cat. 5206 for 2018–19.

Based on Australian Bureau of Statistics data, Cat. No. 5206, various tables.

In summary

- The four-sector circular flow model shows how products and resources – and payments to that – flow among the household, government, business and foreign sectors.
- GDP is also conceptualised as the production of new goods and services flowing to these four sectors.
- GDP is measured using the expenditure approach by adding the spending of households (*C*), spending by firms for investment (*I*), spending by government (*G*), and the spending of foreigners less any spending on imports during the period (*X* – *M*).
- Other important national income measures are **net domestic product**, **gross national income** and **gross disposable income**.

3 GDP shortcomings as a measure of economic welfare

Due to the very complex methods used in estimating GDP – and the sheer enormity of the task – GDP is necessarily a less-than-perfect measure of a nation's economic pulse. Nonetheless, measured GDP certainly plays a very critical role in influencing government economic and social policies. Therefore, there is, quite appropriately,

some degree of concern that a false impression of a nation's material wellbeing may result from an imperfectly measured GDP. In this section, we discuss some of the measurement issues that are worth remembering.

Non-market transactions

Because GDP counts only market transactions, it excludes certain unpaid activities, such as household production, child-rearing and do-it-yourself home repairs and services. For example, if you take your dirty clothes to the cleaners, GDP increases by the amount of the cleaning bill paid. But GDP ignores the value of cleaning these same clothes if you clean them yourself at home. And yet, in both cases, the clothes end up being cleaned! Similarly, non-market activities, such as house-and-yard and charity work, are not included in the estimates, nor are the not-insubstantial activities of students in adding to their (and the community's) stock of human capital.

There are two reasons for excluding non-market activities from GDP. First, it would be extremely imprecise to attempt to collect data and assign a dollar value to services people provide for themselves or others without compensation. Second, it is difficult to decide which non-market activities to exclude and which ones to include. Perhaps mending your roof, painting your own house and repairing your car should be included. But what about washing your car, mowing your lawn and pruning your shrubs?

Distribution, kind and quality of products

GDP data provides no information as to whether a small fraction of the population consumes most of a country's GDP or whether consumption is evenly divided. Furthermore, GDP data alone provide no information in respect of the quality and kinds of goods and services that comprise a nation's GDP.

Consider the fictional economies of Kuga and Agor. Kuga has a GDP of $2000 billion, and Agor has a GDP of $1000 billion. At first glance, Kuga appears to possess superior economic wellbeing. However, suppose a substantial proportion of Kuga's GDP consists of military goods. At the same time, Agor's products include some military goods but comprise mostly computers, entertainment goods and services, clothing, tractors, wheat, milk, houses and other consumer items.

Moreover, imagine that the majority of the people of Kuga couldn't care less about the output of military goods and would be more pleased with the production of consumer goods. It would be reasonable to conclude, therefore, that as far as the people of Kuga are concerned, not many of them would feel as though they were better off than the people of Agor.

Furthermore, suppose that the population of Kuga is twice that of Agor. Therefore, GDP per capita in the two countries is actually the same. Moreover, in Kuga, 20 per cent of its people are actually the beneficiaries of 80 per cent of its GDP production. In comparison, in Agor, 60 per cent of its people are the beneficiaries of 80 per cent of its output. In other words, the GDP is shared more widely in Agor than in Kuga. All these additional issues will contribute to the overall economic wellbeing felt by the two communities in question.

Another significant practical issue involved in measuring GDP is how to allow appropriately for changes in output quality over time. For example, if from one year to the next, the number of bananas produced increases from 100 to 105, we would be reasonably safe in saying that banana output has increased by 5 per cent. This is because it is reasonable to assume that the basic quality of the bananas has remained similar.

However, now suppose the good we are considering is a laptop, and there is the same percentage increase in the number produced from one year to the next (i.e., 100 to 105). This is a much more complex good with many different quality characteristics. If there are enhancements to any of these characteristics, how are they to be measured and compared?

For example, suppose each of the 105 laptops in year two has 20 per cent more memory and extra colour monitor features, require less CPU time for tasks and have slightly better monitor resolution than the 100 from the year before. Is it appropriate still to say GDP increased by only 5 per cent? After all, in the banana case, if each of the 105 bananas in year two was also 20 per cent bigger, should we still simply count them? The vexed question of quality change is one that is a continual challenge for the ABS.

Neglect of leisure time

The wealthier a nation becomes, in general, the more leisure time its citizens can afford. Rather than continuing to work long hours, people often choose to increase their time for recreation, their families and travel. Over the last century, the length of the typical workweek in most developed countries has declined steadily from about 50 hours in 1900 to about 35 hours in 2017. As another example, suppose a community voluntarily chooses at some point to spend less time working to produce output and more time meditating on the mysteries of life. Its GDP will undoubtedly reduce, but has its welfare reduced, increased or stayed the same?

The underground economy

Illegal gambling, prostitution, loan-sharking and illegal drugs are goods and services that meet all the requirements for GDP. They are final products with a value determined in markets (although the markets are illegally conducted!), but GDP does not include unreported criminal activities. Some argue that even if some way could be found to estimate the quantum of these outputs accurately, they should be excluded from GDP anyway because they are 'bads' rather than 'goods'.

In the language we developed in **Chapter 1** when discussing economics and ethics, these activities do not represent the pursuit of 'enlightened self-interest'. Others, however, argue (although somewhat controversially) that if people derive satisfaction from such activities and are willing to pay for them, they should be included in the same way that the purchase of alcohol and certain magazines and DVDs of questionable social value are currently included.

The 'underground' economy also includes activities not reported because of tax evasion. One way to evade paying taxes on a legal activity is to trade by barter, rather than by buying or selling goods and services. One person fixes a neighbour's car in return for babysitting services, and the value of the exchange is unreported. Other legal sales are made for cash by some individuals and businesses who then fail to report the income earned to the Australian Taxation Office.

Estimates of the size of this hidden economy vary from time to time and from country to country. Some studies by economists estimate the size of the underground sector in Australia at between 5 and 10 per cent of GDP. This range of estimates is comparable to the estimated size of the underground economy in most developed countries.

Economic 'bads'

More production means a larger GDP, regardless of the level of pollution created in the process. Recall from **Chapter 4** the discussion of negative externalities, such as pollution caused by steel mills, chemical plants and cigarettes. Air, water and noise pollution are examples of economic 'bads' (which are not illegal) that impose costs on society that are often not reflected in private-market prices, and quantities bought and sold.

When a polluting company sells its product, this transaction increases GDP. However, critics of GDP argue that it fails to account for the diminished quality of life from the 'bads' not reported in GDP. Another 'bad' that many people are very concerned about is ecological degradation. For this reason, several economists prefer to use a measure known as the 'genuine progress indicator' (GPI).[2]

The GPI attempts to adjust measured GDP for the creation of such 'bads' as mentioned above, as well as to allow for any reduction or degradation of the nation's ecological systems and resources. Researchers at the Australia Institute have compared changes in their GPI measure with changes in GDP per capita and claim that while GDP per capita trebled from 1950 to 2000, their GPI measurement increased by only 73 per cent.

---

[2] Please note, there are also alternative measures that attempt to measure changes in welfare or wellbeing and which have been attracting some attention, including the Happy Planet Index (HPI), the Human Development Index (or HDI, particularly relevant to measuring economic, social, educational and health improvements in developing countries), and the Index of Sustainable Economic Welfare (ISEW).

Can GDP measure human welfare?

Economics and ethics

Some commentators' criticisms of GDP go far further than the listed shortcomings described above, and their criticisms also apply to other measures like the Genuine Progress Indicator and the Human Development Index, which attempt to address some of those shortcomings. Their criticism is on an even more fundamental philosophical basis that, to them, it is quite spurious and misleading to attempt to measure human welfare or to try to compare it over time meaningfully.[1] Particularly pertinent here is part of the University of Kansas Address by US Senator Robert F. Kennedy delivered on 18 March 1968, less than three months before his assassination.

> Our Gross National Product, now, is over 800 billion dollars a year, but that Gross National Product – if we judge the US by that – that Gross National Product counts air pollution and cigarette advertising, and ambulances to clear our highways of carnage. It counts special locks for our doors and the jails for the people who break them. It counts the destruction of the redwood and the loss of our natural wonder in chaotic sprawl. It counts napalm and counts nuclear warheads and armoured cars for the police to fight the riots in our cities. It counts Whitman's rifle and Speck's knife, and the television programs that glorify violence in order to sell toys to our children. Yet the Gross National Product does not allow for the health of our children, the quality of their education or the joy of their play. It does not include the beauty of our poetry or the strength of our marriages, the intelligence of our public debate or the integrity of our public officials. It measures neither our wit nor our courage, neither our wisdom nor our learning, neither our compassion nor our devotion to our country. It measures everything, in short, except what makes life worthwhile. And it can tell us everything about America except why we are proud that we are Americans.

1 See, for example, 'The spurious precision of GDP', by Edward Hadas, appearing in *Breaking views*, 22 February 2012, writing on the topic, 'Ethical economy: Don't obsess about GDP measures', https://www.breakingviews.com/columns/ethical-economy-dont-obsess-about-gdp-measures/

Finally, in some perverse ways, reductions in community wellbeing may manifest as increases in measured GDP. For instance, suppose that due to a decrease in public safety resulting from, say, increased street violence, more police are employed. Their salaries will show up as an increase in GDP, and yet it is difficult to see how the use of more scarce human resources to try to counteract increased community violence is an increase in welfare.

In summary, while GDP is undoubtedly a very useful summary measure of the size of the economy, for all of the above reasons, growth in GDP should never be confused with improvements in welfare (economic or otherwise).[3]

Applicable concept: national income accounting 'goods' and 'bads'

Analyse the issue

Is the use of GDP for measuring economic performance and welfare helping or hindering the sustainable development of our world?

Suppose a factory in your community has been dumping hazardous wastes in the local water supply and people have contracted cancer and other illnesses from drinking polluted water. The Department of the Environment discovers this pollution and, orders a clean-up and a fine for the damages.

3 It is worth mentioning that even Simon Kuznets, mentioned earlier, who pioneered the measurement of GDP, was careful to warn against using it as a proxy for economic welfare.

In terms of the GDP, an amazing 'good' result occurs: the primary measure of national economic output, GDP, increases due to the increase in cleaning the environment, health-care costs and government spending. GDP counts the millions of dollars spent to clean up the water supply. GDP even includes the health-care expenses of anyone who develops cancer or other illnesses caused by drinking polluted water. Also, GDP consists of the money spent by the government to regulate the polluting company.

Now consider what happens when raw materials like oil and minerals are used to produce houses, cars and other goods. The value of the oil and minerals is an intermediate good implicitly computed in GDP because the value of the final goods is explicitly calculated in GDP. Using scarce resources to produce goods and services, therefore, raises GDP and is considered a 'good' result. On the other hand, don't we lose the value of the oil and minerals in the production process and isn't this a 'bad' result?

Shutterstock.com/ngaga

Critics of current methods used to estimate GDP have called for a new measure designed to adjust GDP in recognition of these types of issues. These new accounts would adjust for changes in air and water quality and depletion of crucial non-renewable raw materials like oil. These accounts would also adjust for changes in the stock of renewable natural resources, such as forests and fish stocks. Also, it is argued by some that accounts should be created to measure global warming and destruction of the ozone layer.

As explained in this chapter, a dollar estimate of physical capital depreciation is subtracted from GDP to compute NDP. The argument here is that a dollar estimate of the damage to the environment should also be deducted. Critics argue that ignoring measuring such environmental problems threatens future generations.

Critics of so-called 'green national accounts', however, argue that assigning a dollar value to environmental damage and resource depletion requires a methodology that is extremely subjective and complex. Notwithstanding these difficulties, central statistical agencies around the world have not ignored the criticisms and continue to make efforts to refine their methods, but in prudent, considered and appropriate ways.

What do you think?

Consider the collision of the Sanchi oil tanker carrying natural-gas condensate that occurred in January 2018, 300 kilometres off Shanghai, China. It caught fire shortly after the collision and burnt for over a week. From this incident, a large slick was formed on the surface of the sea. The environmental impact of this was different compared with crude oil spills and this spill created immediate toxic effect on the marine inhabitants. The environmental damage, including possible contamination of beaches, as well as the damage to the fishing industry, had been significant and the impacts were felt by Japan, Korea and China directly. How might measured GDP be a misleading indicator of the impact of this miserable environmental disaster?

In summary

- A country's GDP can be quite an imprecise and misleading indicator of economic wellbeing.
- It does not reflect the distribution, kind and quality of products, it neglects **non-market activities** even though they certainly add value to our welfare.
- It will also understate the volume of economic activity on account of the existence of the underground economy.
- Finally, GDP neglects to adjust for the pollution or resource depletion, which may result from the production of goods and services.

4 Changing nominal GDP to real GDP

So far, the GDP we have been discussing is more precisely referred to as **nominal GDP**. Nominal GDP is the value of all final goods and services based on the prices existing during the period of production. Nominal GDP is also alternatively referred to as the *current dollar* or *money GDP*. Nominal GDP grows over time not only as a result of increases in output but also as a result of general price rises.

To get a clearer picture of how much an economy is growing over time, it is necessary to adjust nominal GDP so that it reflects only changes in output and not changes in prices. This adjusted GDP allows meaningful comparisons of aggregate economic activity over time when prices are changing and is referred to as *real GDP*. Therefore, the percentage change in reals GDP is used in calculating economic growth rates. **Real GDP** is the value of all final goods and services produced during a given period based on the prices existing in a selected reference year. At the time of writing, the ABS was using 2017–18 as the reference year. Real GDP is also referred to as constant-dollar GDP.

Nominal GDP
The value of all final goods and services produced in a period based on the prices existing during the time period of production.

Real GDP
The value of all final goods and services produced during a given time period based on the prices existing in a selected reference year.

The chain volume measure of GDP

The method used by the ABS to take the changes-in-the-price-level 'air' out of the nominal GDP 'balloon' and compute real GDP is to calculate the so-called **chain volume measure of GDP**. The chain volume measure of GDP is a measure of GDP that provides a measure of all the final goods and services produced during a given period using a set of relevant prices prevailing in some selected reference year.

The **GDP implicit price deflator** is a price index derived implicitly from the ratio of a set of nominal or current price GDP data to the corresponding chain volume measures of GDP for the same time periods. The GDP implicit price deflator is a very broad 'deflator' index, which implicitly – or indirectly – measures not only price changes of consumer goods but also price changes of business investment, government expenditures and net exports. Percentage change in GDP implicit price deflator provides a broad-based measure of a country's inflation. Do not confuse the GDP implicit price deflator with the consumer price index (CPI), which is widely reported in the news media. The CPI is a different index, measuring only consumer prices, which we will discuss in the chapter on inflation and unemployment.

Chain volume measure of GDP
The chain volume measure of GDP is a measure of GDP that provides a measure of all the final goods and services produced during a given period using a set of relevant prices prevailing in some selected reference year.

GDP implicit price deflator
A price index derived implicitly from the ratio of a set of nominal or current price GDP data to the corresponding chain volume measures of GDP for the same time periods.

For now, let's see how it works. We begin with the following conversion equation:

$$\text{Real GDP} = \frac{\text{nominal GDP} \times 100}{\text{GDP implicit price deflator}}$$

Now, using 2017–18 as the reference year, suppose you are given the 2018–19 nominal GDP of $1893.1 billion and the 2018–19 GDP implicit price deflator of 103.2. To calculate 2018-19 real GDP, use the above formula as follows:

$$\text{2018–19 real GDP} = \frac{(1893.1 \times 100)}{103.2} = \text{1834.4 billion}$$

Exhibit 11.7 shows actual Australian nominal GDP, real GDP and GDP implicit price deflator index computations for selected years. Column 1 reports nominal GDP, column 2 gives real GDP figures for these years and column 3 lists corresponding GDP implicit price deflator index numbers. Notice that, in the years before 2017–18, the GDP price deflator is less than 100, implying that, on average, prices in those years were less than those in 2017–18. At the reference year of 2017–18, nominal and real GDP are identical, and the GDP, implicit price deflator index, equals 100.

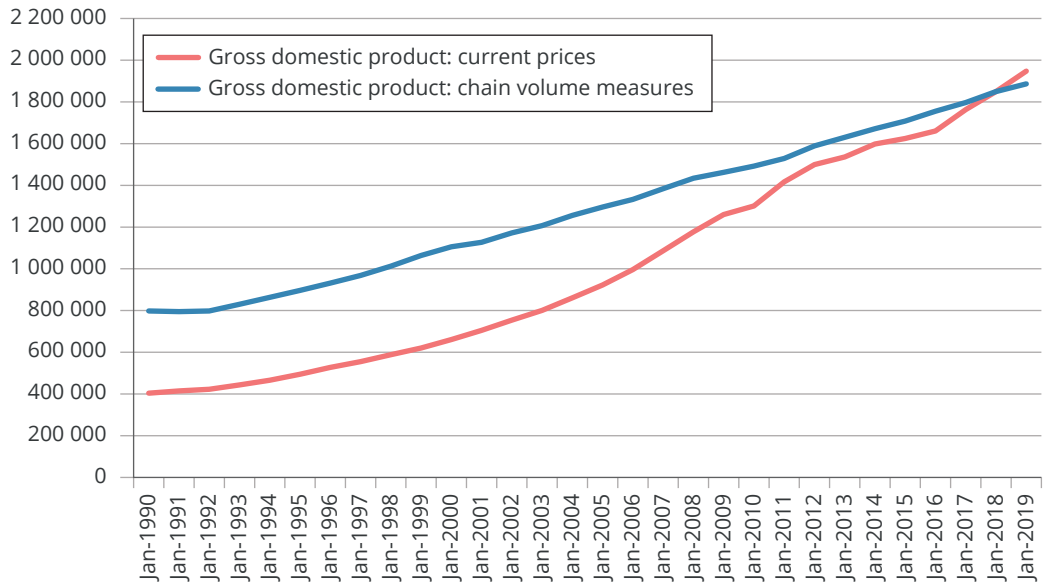
Exhibit 11.8 graphs real GDP and nominal GDP for the economy over the same period as in **Exhibit 11.6**. Note that nominal GDP usually grows faster than real GDP. For example, if we calculate the economy's growth rate in nominal GDP between 2015–16 and 2016–17, the economy's annual growth rate was 6 per cent. If instead

Exhibit 11.7 **Australian nominal and real GDP**

Year	(1) GDP: Current prices (millions of dollars)	(2) GDP: Chain volume measures (millions of dollars)	(3) GDP implicit price deflator
1989/90	40 3855	79 8023	50.6
1990/91	41 4478	79 4853	52.1
1991/92	42 2440	79 8132	52.9
1992/93	44 3390	83 0292	53.4
1993/94	46 5757	86 3360	53.9
1994/95	49 4567	89 6479	55.2
1995/96	52 7599	93 1250	56.7
1996/97	55 5291	96 8188	57.4
1997/98	58 8097	101 2499	58.1
1998/99	62 0141	106 3869	58.3
1999/00	66 1159	110 5712	59.8
2000/01	70 5066	112 7062	62.6
2001/02	75 4253	117 2160	64.3
2002/03	80 0960	120 7160	66.4
2003/04	86 1039	125 6117	68.5
2004/05	92 2371	129 6365	71.2
2005/06	99 6430	133 2593	74.8
2006/07	108 6593	138 3822	78.5
2007/08	117 7313	143 4441	82.1
2008/09	126 0145	146 2225	86.2
2009/10	130 1211	149 2456	87.2
2010/11	141 6622	152 9212	92.6
2011/12	149 9458	158 9129	94.4
2012/13	153 6307	163 0206	94.2
2013/14	159 8530	167 1501	95.6
2014/15	162 4601	170 8151	95.1
2015/16	166 0714	175 5478	94.6
2016/17	176 3627	179 7041	98.1
2017/18	184 9880	184 9880	100.0
2018/19	194 7669	188 6576	103.2

Source: Australian Bureau of Statistics, Cat. No. 5206, Table 36.

Exhibit 11.8 **Real GDP and nominal GDP, 1989–90 to 2018–19**



Source: ABS, http://www.abs.gov.au, Cat. No. 5206, Table 36.

Each year's real GDP reflects output valued at 2017–18 reference-year prices, but nominal GDP is annual output valued at prices prevailing during the current year. The intersection of real and nominal GDP occurs in 2017–18 because in the reference year both nominal GDP and real GDP measure the same output at 2017–18 prices. Note that the nominal GDP curve has risen more sharply than the real GDP curve since 1989–90 as a result of inflation impacting on the nominal figures.

the calculation is made for real GDP growth between the same years, the growth rate was just 1.9 per cent. You must, therefore, pay attention to which measure of GDP is being used in an analysis.

Is the economy up or down?

You're the economist

Declan claims, 'GDP rose this year by 3 per cent'. But Monica replies, 'No it didn't, Declan, GDP fell by 0.3 per cent'. Can both Declan and Monica be right?

In summary

- In measuring the growth of an economy over time, **real GDP** should be used rather than nominal GDP.
- **Nominal GDP** is the value of all final goods and services based on the prices existing during the period of production. Nominal GDP grows over time not only as a result of increases in output but also as a result of general price rises.
- **Real GDP** is the value of all final goods and services produced during a given period based on the prices existing in a selected reference year.
- The **GDP implicit price deflator (IPD)** is the ratio of nominal GDP to real GDP and is a very broad measure of average price changes of the goods and services produced by a country.

5 Relationship between national saving and investment in an open economy

One last important relationship is linking a community's saving and its investment. Investment has already been defined above, but we need to define what is meant by saving. An individual saves whenever their income in a period is greater than their consumption. This is also true for a nation. National saving is simply that portion of the aggregate income generated by a nation's residents, in a given period that is not consumed by them in that period.

In very simple economies, a community's saving is, by definition, its investment ($S = I$). For example, imagine a desert island economy where the sole resident, Lilly-Anne Crusoe, is both the household sector and firm sector. Each period she survives quite happily by catching fish and gathering coconuts just sufficient for her needs. In this case, the island's GDP (the fish and coconuts) is exactly equal to the consumption (C) of its sole resident, so GDP = C, with no saving or investment undertaken.

One day Lilly-Anne realises that by giving up some of her usual fish catching and coconut gathering time to make a fishing net, she would be able to catch even more fish in the future (or have more leisure time). The island's national accounts for that period would be as follows. Total production or output (GDP) would now equal a smaller-than-usual number of fish and coconuts, which Lilly-Anne readily consumes (C), plus an inedible fishing net, which is her investment for that period (I) – so for the island economy in this period, GDP = $C + I$.

Furthermore, by definition, the difference between the island's total production and consumption (GDP – C) is the economy's saving (S) by its (rather hungry) inhabitant. This saving is identical to the island's investment during the period ($S = I$ = the net). In the following periods, Lilly-Anne enjoys the fruits of her earlier thrift by being able to have greater consumption than was possible before. This simple example illustrates that in closed economies (no trade with the rest of the world), like the simple island economy just described, a nation's saving will always, by definition, be equal to its investment (S will always = I).

Of course, in more complex real-world economies, the saving and investment decisions are not made by the same person (like Lilly-Anne) but are undertaken by different groups of individuals. Households decide how much to save, while firms decide how much to invest in capital (plant and equipment). How can it be the case, then, that these two sets of decisions will always be consistent with each other so that S will always equal I in every period?

The answer is that the desired saving by households will not always be equal to the desired investment by firms. However, actual measured saving and measured investment will always be equal. Whenever desired saving by households is different from the desired investment by firms, there will be an unintended inventory accumulation or decumulation by firms. For example, if planned household saving is greater than expected by firms when they made their production and investment plans, they will experience an unexpected (and unwanted) increase in their levels of inventories. As noted earlier, however, these inventories are nonetheless considered part of investment spending in the period in question, and so measured saving (S) equals measured investment (I).

National saving
A country's total income accruing to its residents less its consumption.

National investment
A country's total spending on capital goods by the private sector and government. If national investment is greater than national saving, then the country will have had an inflow of foreign investment.

Now let's have a quick look at how it all works for an open economy (one that trades with the rest of the world) with a government sector, like modern real-world economies. Here, the private sector (households and firms) saves whenever its income is greater than its consumption. Similarly, the government saves whenever its tax revenue is greater than what it spends on government consumption. The total of both these sources of saving is a country's national saving.

This **national saving** – which is just total income accruing to a nation's residents arising from production that is not consumed – must, by definition, be part of the country's aggregate investment (like Lilly-Anne's net). But this is not the only source of saving for aggregate national investment.

In an open economy, the rest of the world can also contribute some of its saving to increase a country's aggregate **national investment** further. This foreign saving that flows into a country is known as foreign investment, and it occurs if foreigners believe they can get a good financial return on their saving by investing

in other countries. Thus, it is crucial to understand that, for an open economy, a country's aggregate national investment can be greater (or less) than its national saving. For example, Australia's aggregate national investment is usually greater than its national saving, while Japan's is typically less.

Whenever a country's aggregate national investment spending is greater than its national saving, the additional investment spending will be financed from **net foreign investment** flowing into the country. This net foreign investment will equal the increase in the total of a country's net foreign liabilities (i.e., its net foreign debt plus the net foreign ownership of its domestic factors of production). Thus, we have:

Net foreign investment The increase (or decrease) in the total of a country's net foreign liabilities (i.e., its net foreign debt plus the net foreign ownership of its domestic factors of production).

national investment − national saving = net foreign investment

In 2018–19 Australia's national saving, defined as gross disposable income less total consumption, was $405.5 billion (= 1830.7 billion − 1064.1 billion − 361.1 billion; refer to **Exhibits 11.3 and 11.5**). Gross investment by both the private and government sector was a total of $445.5 billion (=350.5 billion + 95.0 billion; refer to **Exhibit 11.3**), implying Australia had a net inflow of foreign investment in 2018–19 of $40.0 billion.

In **Chapter 18**, we will refer to this issue of net foreign investment again when we discuss the balance of payments.

Applicable concepts: terms of trade considerations in measuring real GDP

Global perspective

In the 21st-century globalised world, most countries are heavily reliant on international trade for their economic welfare. An important consideration then is a country's terms of trade (TOT), which is easily understood as the average price of a country's exports divided by the average price of its imports. If a country's export prices improve while the prices of the products its residents' import stay the same, then the country's TOT improves and its residents are economically better off.

For example, suppose a country's export volumes do not increase from one period to another, but the prices it receives for those exports in international markets increases by 20 per cent. In this situation, measured real GDP would not increase since the volumes of exports remain the same, but the country's residents are better off! The 20 per cent extra revenue its exports delivers allows its residents, for example, to be able to afford more imported products they may like or to take more overseas vacations.

To account for this, it's possible to compute what is called real gross domestic income (RGDI). Many countries routinely calculate RGDI, but it does not garner very much attention by commentators or the media. Notwithstanding this, when a country's TOT are changing significantly and rapidly, changes in real GDP (RGDP) can give quite a misleading picture of how the economic welfare of the country's residents is changing.

For instance, in times when international commodity prices are increasing rapidly, those countries around the world that benefit most from such a commodity price boom are countries that are commodity exporters; for example, Norway, New Zealand, Canada and Australia. For these countries, RGDI grows more quickly than RGDP. On the other hand, those countries that are commodity importers, for example, Japan, Korea and the US, RGDI will grow more slowly than does RGDP.

What do you think?

Now, consider **Exhibit 11.9**, which provides relevant Australian data over a recent period.

Exhibit 11.9 **Australian RGDP versus RGDI: 2012/13–2018/19**

Year	RGDP	RGDI
2012/13	1630206	1663065
2013/14	1671501	1691869
2014/15	1708151	1689391
2015/16	1755478	1700066
2016/17	1797041	1789650
2017/18	1849880	1849880
2018/19	1886576	1910527

Source: Australian Bureau of Statistics, ABS Cat. No. 5206, Australia's National Accounts, Table 34: Key aggregates and analytical series.

1 Based on the data, in what year were Australian residents benefitting most strongly from improvements in the TOT? What is the quantification of the benefit in absolute terms and as a percentage of RGDP?
2 Looking at the years 2012/13 and 2013/14, what is the growth in RGDP compared with the growth in RGDI? Now compute the growth rates between 2017/18 and 2018/19 for each of RGDP and RGDI. What do you conclude must have been happening to Australia's TOT in the two different periods?

In summary

- **National saving** = total income accruing to a nation's residents from production during a period (its gross disposable income) less its total consumption (both private and government).
- **National investment** = total spending by the private and government sectors on capital goods.
- **National investment** = National saving + **net foreign investment**.

Key concepts

Gross domestic product (GDP)
Gross national income (GNI)
Transfer payment
Final products
Intermediate products
GDP at factor cost
GDP at market prices
Indirect taxes less subsidies
Circular flow model
Flow
Stock
Expenditure approach
Income approach
Gross investment
Net investment
Gross national expenditure (GNE)
Net domestic product (NDP)
Net national income
Gross disposable income
Net disposable income
Nominal GDP
Real GDP
Chain volume measure of GDP
GDP implicit price deflator
National saving
National investment
Net foreign investment

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1 Understand the concept of gross domestic product

- **Gross domestic product (GDP)** is the market value of all final goods and services produced within a nation's geographic borders for a given time while **Gross national income (GNI)** (or GNP) is the income accruing to a country's residents from the production of all final goods and services for a given time, no matter where the goods and services are produced in the world.
- GDP (and GNI) only includes new production of final goods and services.
- **GDP at factor cost** is arrived at by adding together all the 'value added' at each stage of the production chain, and to this is added 'indirect taxes less subsidies' paid to the government to arrive at **GDP at market prices**.

2 Understand how gross domestic product is measured

- The *four-sectors circular flow model* is a simple macroeconomic model that conceptually shows how products and resources – and payments thereto – flow among the household, government, business and foreign sectors. GDP is also conceptualised as the production of new final goods and services flowing to these four sectors.
- GDP can be measured by either the **expenditure approach**, the **income approach**, or by the **output approach**.
- GDP is measured by adding the consumption spending of households (*C*), spending by firms for investment (*I*), spending by government (*G*), and the spending of foreigners less any spending on imports during the period ($X - M$).
- Other important national income measures are **net domestic product**, **gross national income** and **gross disposable income**.

3 Appreciate the shortcomings in using gross domestic product as a measure of the economic welfare of a country

- A country's GDP is quite a narrowly defined quantitative, rather than a qualitative, measure of a country's production of goods and services.
- GDP can be quite an imprecise and misleading indicator of the economic wellbeing of the inhabitants of a country. It does not reflect the distribution, kind and quality of products; it neglects non-market activities, even though they certainly add value to our welfare; and it ignores the value of leisure time.
- GDP will also understate the volume of economic activity on account of the existence of the underground economy.
- GDP neglects to adjust for the pollution or resource depletion, which may result from the production of goods and services.

4 Learn the difference between nominal and real gross domestic product

- In measuring the growth of an economy over time it is common to use changes in **real GDP** rather than changes in so-called **nominal GDP**, which is the value of all final goods and services based on the prices existing during the period of production.
- Nominal GDP grows over time not only as a result of increases in output but also as a result of general price rises.
- **Real GDP** is the value of all final goods and services produced during a given period based on the prices existing in a selected reference year. The **GDP implicit price deflator (IPD)** is the ratio of *nominal GDP* to *real GDP* and is a very broad 'deflator' index, which implicitly – or indirectly – measures not only price changes of consumer goods but also price changes of business investment, government expenditures and net exports.

5 Develop an understanding of the nature of the national saving and investment relationship in an open economy

- A country's total income accruing to its residents from production during a period (its **gross disposable income**) less its total consumption (by both the private and the government sector) is the nation's **national saving**.
- The total spending on capital goods by the private sector and government is its aggregate **national investment**.
- The national saving must, by definition, be part of the country's aggregate investment. However, in an open economy, the rest of the world can also contribute some of its saving to increase a country's national investment further.
- If the national investment is greater than national saving, then the country will have had a net inflow of foreign investment.

Study questions and problems

1 Using the basic or simple circular flow model, explain why the value of businesses' output of goods and services equals the income of households.

2 Which of the following are final goods or services and which are intermediate goods or services?

a Timber sold to a family to be used in repairing their backyard decking.

b A tonne of fertiliser sold to a farmer to use for his sugar cane crop to ready it for the market.

c Household cleaning purchased by a family from a domestic cleaning services company.

d Glass sold to a windscreen maker to be used to make truck windscreens.

3 A small economy produced the following final goods and services during a given month: 0.5 million kg of food, 40 000 shirts, 16 houses, 45 000 hours of medical services and 12 speedboats. Calculate the value of this output at the following market prices:

- $3 per kg of food, $15 per shirt, $250 000 per house, $60 per hour of medical services, $140 000 per speedboat.

4 An economy produces final goods and services with a market value of $1850 billion in a given year, but only $1810 billion worth of goods and services are sold to domestic or foreign buyers. What is the nation's GDP? Explain your answer.

5 A new aeroplane is produced in Australia and sold to Qantas to provide freight and passenger services to its customers. Is the aeroplane a final good or an intermediate good? Should the value of the aeroplane be included in GDP? Should the freight and passenger services by Qantas to its customers be included in GDP? Can you find the issue of double-counting?

6 Suppose the following are national accounting data for a given year for some particular country. Calculate the country's gross national expenditure, GDP and NDP using the expenditure approach.

	Amount (billions $)
Consumption of fixed capital	320
Gross private domestic investment	350
Government consumption expenditures	330
Government investment expenditures	95
Imports	355
Exports	370
Household consumption expenditures	1000
Net foreign income paid overseas	45
Net transfers to foreigners	2

7 Again using the data from question (6), assuming a zero statistical discrepancy between GDP valued on an expenditure basis and the total income generated from GDP, derive the country's gross national income (GNI) and its gross disposable income (GDI) for the year.

8 Following on from question (7), what was the country's total gross national investment during the year? What is the country's gross national saving? What, therefore, must have been the value of net foreign investment flowing into the country during the period?

9 Which of the following are counted in this year's expenditure-based estimates of GDP? Explain your answer in each case.
- **a** KuickCopy bought a new copy machine during the year for its copying business.
- **b** The Ng family bought a foreign-made boat from a second-hand boat dealer.
- **c** Sally Chau purchased 50 shares of BHP Billiton stock as part of a new share offering by the company.
- **d** Tom Smith received his unemployment benefits.
- **e** RelaxRealEstate agents sold a second-hand house and charged a sales commission on the deal.
- **f** Due to increased air and water pollution, an expensive clean-up operation is contracted out by the government to a private company.
- **g** Some previously illegal forms of gambling are legalised in all states.

10 Suppose a country's nominal GDP in 2018 is $1650 billion and nominal GDP in 2019 is $1715 billion. Consider 2018 as the reference year, and implicit price deflator for 2019 is 103. Calculate the real GDP for 2018, real GDP for 2019, implicit price deflator for 2018, and economic growth rate between 2018 and 2019. Also, calculate a broad-based measure of inflation in the country between 2018 and 2019.

Answers to 'You're the economist'

How much does Serena add to GDP?

Measuring GDP by the expenditure approach, Serena's output production is worth $25 000 because consumers purchased 1000 small toys at $25 each. Transfer payments and goods that are produced in the previous years are not included in the GDP calculation. Therefore, $2500 unemployment benefits Serena received and $3000 she spent to buy her mom's old car are excluded from GDP. Serena's wages of $40 000 is also not counted in the expenditure approach of calculating GDP. If you said that, using the expenditure approach to measure GDP, Serena contributes $25 000 to GDP, YOU ARE THINKING LIKE AN ECONOMIST.

Is the economy up or down?

Both Declan and Monica did not mention whether they are referring to real GDP or nominal GDP. This situation mentioned is possible when nominal GDP rises but real GDP falls. It is rare for this to occur in annual data but happens more frequently with quarterly data. An annual example happened in the Australian economy in 1982–83 whereby nominal GDP rose by 7.6 per cent. During the same period, real GDP fell by 2.3 per cent. Another example occurred between 1989–90 and 1990–91 when nominal GDP rose by 2.5 per cent and real GDP fell by 0.1 per cent. If you said both Declan and Monica could be correct because of the important difference between nominal and real GDP, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

(Note: For some of the following more than one answer may be correct)

1 Which of the following best describes gross domestic product (GDP)?
- **a** The market value of all inputs used to produce final goods and services in a country during a period of time.
- **b** The market value of all final products produced by citizens of a country during a period of time.
- **c** The income accruing to a country's residents from the production of final goods and services during a period of time regardless of where the goods and services were produced.
- **d** The total spending in a period of time on final consumption and investment products by households, firms, all levels of government and the rest of the world.

2 If Monica buys a thick shake from a local shop, which of the following best relates this transaction to the circular flow model?
- **a** Goods flow from firms, and money flows from households through product markets.
- **b** Factors of production flow from households and money flows from firms in product markets.
- **c** Income flows from factor markets, and goods flow into factor markets.
- **d** Factors of production flow into households, and back into the product markets.

3 A country's gross national saving can be defined as:
- **a** the total volume of all the bank deposits of all residents in a country.
- **b** its total national investment less any net foreign investment flowing into a country during a period of time.
- **c** its gross disposable income minus its total consumption during a period of time.
- **d** its gross domestic product less total spending by the private and government sectors on consumption products during a period of time.

4 As a measure of the nation's economic pulse, GDP is often criticised because it:
- **a** uses expenditure estimates based on sample surveys.
- **b** fails to completely allow for changes in the quality of goods and services.
- **c** only comes out quarterly instead of monthly.
- **d** includes the earnings of residents from their investments overseas.

5 Gross national expenditure:
- **a** includes spending on exports by foreigners.
- **b** includes all spending in the economy, including that on second-hand goods.
- **c** includes spending on imports by residents.
- **d** includes transfer payments from government paid to residents abroad.

6 Subtracting an allowance for depreciation of fixed capital from gross domestic product yields:
- **a** real GDP.
- **b** gross disposable income.
- **c** nominal GDP.
- **d** net domestic product.

7 Tanya is studying a business course and her teacher has asked her to calculate the economic growth rate for Australia. Which of the following measures should Tanya use to calculate the economic growth?
- **a** Nominal GDP.
- **b** Gross national income.
- **c** Real GDP.
- **d** Real gross domestic income.

8 Which of the following is included in the calculation of GDP?
- **a** The purchase of 100 newly released shares in BHP Billiton.
- **b** The sales commission of the real estate agent who sold a family an existing home.
- **c** The value of a second-hand truck by a firm to deliver its newly produced products to customers.
- **d** An imputed value of the time spent by unpaid volunteers in a homeless shelter.

9 The GDP implicit price deflator is:
- **a** a measure designed to represent percentage changes in real output.
- **b** a measure designed to adjust for changes in the quality of goods and services over time.
- **c** a measure of changes in consumer prices.
- **d** a broadly based measure of price changes in the goods and services in GDP.

10 Australia had national saving is less than national investment in 2018–19. What does this imply?
- **a** Australia had no foreign investments in 2018–19.
- **b** Australia had a net outflow of foreign investments in 2018–19.
- **c** Australia had a net inflow of foreign investments in 2018–19.
- **d** Australia had no inflow of foreign investments in 2018–19.

Business cycles and economic growth

12

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand the concept of the business cycle 'rollercoaster' and its two phases.
2 Develop an understanding of business cycle indicators, including how they are used to determine the phase of the business cycle and to forecast where it may be headed in the near future.
3 Appreciate the importance of the relationship between changes in total spending and the business cycle.
4 Understand the long-term determinants of economic growth and what governments might do to influence a country's long-term rate of economic growth.
5 Appreciate the goals of macroeconomic policy in relation to the business cycle and economic growth.

Getty Images/fotosipsak

The headline in the morning newspaper reads 'Economy's growth rate slowing'. Later in the day, the lead story on drive-home radio is, 'The unemployment rate increases for the first time in 12 months'. On television, the evening news broadcasts an interview with several economists who predict that a downturn in the business cycle is likely before the end of the year. Next, the Leader of the Opposition appears on the screen and says, 'The government has lost its ability to manage the economy and it's time for a change'. These events are meant to show that the growth rate of the economy and the state of the business cycle are headline-catching news as these measures of macroeconomic instability affect everyone's future.

When real gross domestic product (GDP) is on the rise and the economy 'is booming', jobs become plentiful. A continuing fall in real GDP – or a continuing substantial slowing in its rate of growth – means the economy is in a 'slump' or a 'recession'; a state of the economy where many firms may be forced into bankruptcy and cause people to lose their jobs. Not being able to find a job when you want one is a painful experience not easily forgotten!

This chapter looks behind the macro economy at a story that touches each of us. It begins by discussing the business cycle. How are the expansions and contractions of business cycles measured? And what creates the business cycle rollercoaster? The business cycle – although sometimes very painful – may be considered

short-term in nature. The other crucial thing to understand about modern economies is what factors cause them to grow over the longer term and, in particular, why some countries grow more rapidly than others.

1 The business cycle rollercoaster

Business cycle
Alternating periods of economic growth and contraction, which can be dated by changes in output, income, sales and employment measures.

One central concern of macroeconomics is the upswings and downswings in the level of real output that make up the **business cycle**. The business cycle is alternating periods of economic growth and contraction. Business cycles are inherent in market economies. A key indication of cycles is the rise and fall in real GDP. These rises and falls are mirrored in other key measures of the economy, such as employment growth, the rate of unemployment, retail sales and industrial production growth.

For instance, in a downturn, we would see a drop in real GDP, industrial production, sales and employment, with unemployment rising. For the sake of simplicity of exposition, here we will use real GDP as a proxy for the business cycle. However, in reality, downturns and upturns in the business cycle are more than just fluctuations in real GDP. Also recall from the previous chapter that changes in real GDP measure changes in the value of national output, ignoring changes in the price level.

The two phases of the business cycle

Although it is an over-simplification, it is useful to represent the long-term growth in a modern economy as following a fairly smooth growth path (let us say, purely for expositional purposes, a straight line). What we call the business cycle is the succession of short-run fluctuations around this long-term trend. Part (a) of **Exhibit 12.1** illustrates a highly stylised theoretical business cycle. Although real-world business cycles vary in duration and intensity, each cycle can be divided into two distinct and distinctly different phases: **recession** and **expansion**.

Recession
A downturn in the business cycle in which output, sales and employment decline.

Expansion
An upturn in the business cycle during which real GDP, employment and other measures of aggregate economic activity rise.

Peak
The point in the business cycle at which the economy reaches its maximum after rising during an expansion.

Trough
The point in the business cycle at which the economy reaches its minimum after falling during a recession.

The business cycle looks like a rollercoaster. It begins at a **peak**, drops to a bottom, or **trough**, then climbs steeply, begins to level out and reaches another peak. The phase of the business cycle between the peak and the trough is known as the *recession phase*.

Once the trough is reached, the upswing starts again and continues until the next peak is reached. The phase of the business cycle between the trough and the peak is known as the *expansion phase*. Although forecasters cannot precisely predict the timing and duration of the phases of a cycle, the economy is always operating in one of these phases. Over time, there has also been a reasonably smooth long-term upward trend, despite short-term cyclical fluctuations around the long-run trend.

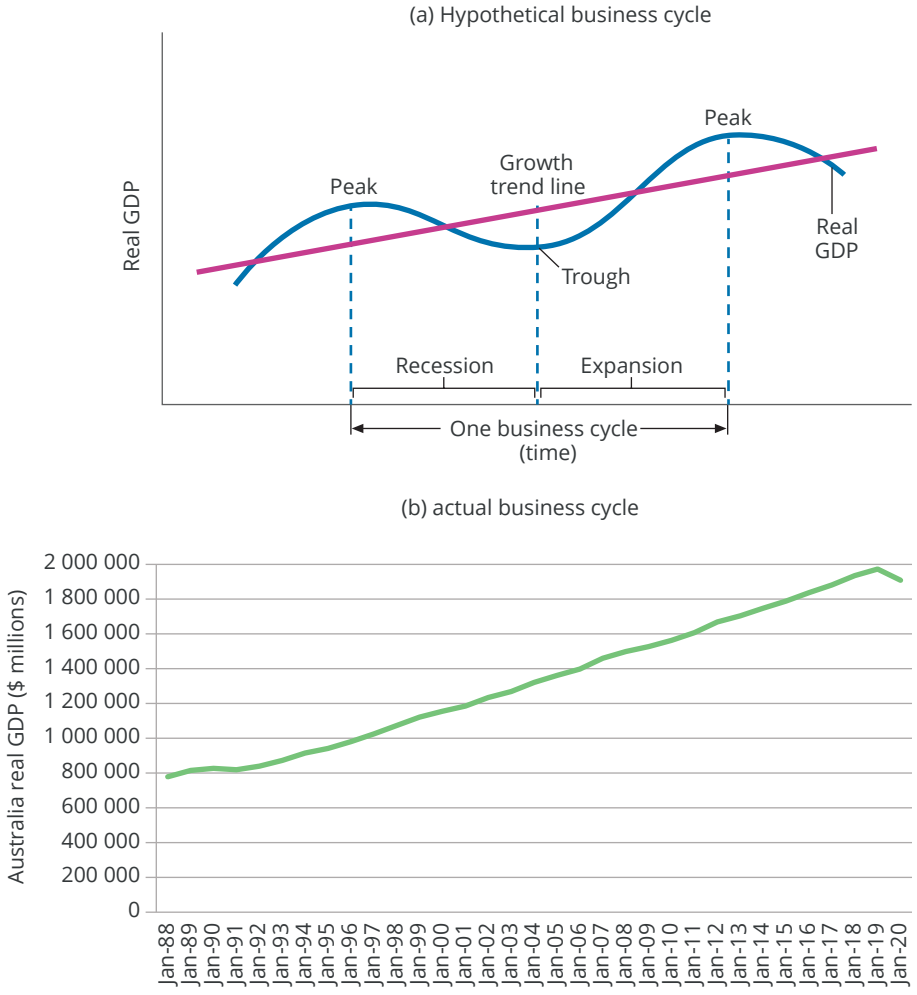
Two peaks are illustrated in part (a) of **Exhibit 12.1**. At each of these peaks, the economy would usually be close to or at full employment (the concept of full employment is discussed in more detail in **Chapter 13**); that is, as explained in **Chapter 2**, the economy would normally be operating near its production possibilities frontier, and real GDP is at its highest level relative to recent years.

The peak represents the end of the expansion phase, and the business cycle then enters a recession phase ('the economy enters recession'). Recessions are also sometimes called *contractions*. A recession is a downturn in the business cycle during which real GDP declines, as do other measures of aggregate economic activity. During a recession, the economy is functioning inside – sometimes well inside – its production possibilities frontier.

The ABS and the Reserve Bank of Australia usually consider a recession to have commenced when there are at least two consecutive quarters (six months) in which there are declines in real GDP, and the peak would be identified as occurring in the quarter immediately before real GDP began falling. In general, during a prolonged recession consumer demand drops significantly, business profits fall, the percentage of the workforce without jobs rises (unemployment rate), and production capacity in the economy is underutilised.

The trough is the point where the level of real GDP 'bottoms out'. At the trough, unemployment and idle productive capacity are at their highest levels relative to recent years. The length of time between the peak and the trough is the duration of the recession. In practice, an expansion is regarded as having commenced after a previous recession when real GDP expands for two consecutive quarters, and the trough would be identified as occurring in the quarter immediately before real GDP began rising again.

Exhibit 12.1 **Hypothetical and actual business cycles**



Part (a) illustrates a hypothetical business cycle consisting of two phases: recession and expansion. These fluctuations of real GDP occur around a growth trend line, which represents the fact that, over time, real GDP tends to trend upward. In reality, the fluctuations are not so clearly defined as those in this graph.

Part (b) illustrates the actual ups and downs of Australia's actual business cycle. The Australian economy was experiencing a downswing in 1990-91 after which followed a reasonably long period of expansion until 1999-2000. After a short recession, Australia then experienced another long expansion until 2007-08. At this time it was impacted by the Global Financial Economic Crisis (GFEC) and went into another recession again lasting about a year. Then the country was again in an upswing and continued the expansionary phase through until the COVID-19 pandemic began in 2019. Due to the impacts of the COVID-19 during 2020, the Australian economy moved to a recession phase.

Source for (b): Australia Bureau of Statistics, 5206.0 Table 34.

In **Exhibit 12.1**, part (b), the actual business cycle of the Australian economy is depicted. After a downswing in 1990, Australia experienced an expansionary period throughout the late 1990s, after which it experienced another downturn in 2000. Then, after a long expansion, the economy went into another GFEC-induced recession lasting about a year. After that, it commenced an expansionary phase through until the onset of the

COVID-19 pandemic in 2019. Due to the impacts of the COVID-19, during 2020 the Australian economy moved to a recession phase.

Since the end of the Second World War, recessions in Australia have averaged about 15 months. Notwithstanding this, as shown in **Exhibit 12.2**, the 1990s recession lasted over two years – from April 1990 to June 1992. The percentage decline in real GDP was 1.9 per cent, and the national unemployment rate hit a peak rate of 11.2 per cent in 1992. However, unlike a great many countries around the world, Australia escaped recession during the GFEC of 2008–09. The full impact of COVID-19 on the Australian economy is too early to predict at the time of the writing of this book; however, Australia experienced a recession where real GDP fell and unemployment increased during the year 2020. Due to the policy measures taken by the Australian government and the Reserve Bank of Australia, the severity of the recession was reduced. More will be said about why this may have occurred in **Chapter 19**.

Exhibit 12.2 **Severity of post-war recessions in Australia**

Recession dates (peak–trough)	Duration (months)	Percentage change in GDP (from peak to trough)	Peak unemployment
Apr. 1951–Sep. 1952	17	n/a	n/a
Dec. 1955–Aug. 1956	8	n/a	n/a
Dec. 1960–Sep. 1961	9	–3.3	n/a
Jul. 1974–Mar. 1975	8	–3.5	n/a
Aug. 1976–Nov. 1977	15	–0.8	6.7
Sep. 1981–May 1983	20	–3.8	10.4
Apr. 1990–Jun. 1992	26	–1.9	11.2
Mar. 2020 - ??	??	??	??
Average	15		

Source: Allan P. Layton, 1997, 'A new approach to dating and predicting Australian business cycle phase changes', *Applied Economics*, Vol. 29, pp. 861–8; and Reserve Bank of Australia, http://www.rba.gov.au.

What is the difference between *recession* and *depression*? There is an old saying: 'A recession is when your neighbour loses his or her job, and a depression is when you lose your job!' This one-liner is close to the real distinction between these two concepts. As explained earlier, economists use 'recession' to refer to any decline in the business cycle lasting at least six months, so why not use the term 'Great Recession' for the contraction of 1929–33? Because no subsequent recession – including the GFEC-induced international recession of 2008–09 – has approached the prolonged severity of the Great Depression, the term 'depression' has primarily now become a historical reference to the extremely deep and prolonged recession of the early 1930s.

It is believed that the recession in Australia in the early 1990s was undoubtedly the most severe experienced in this country since the Second World War. However, the economic recession due to the COVID-19 pandemic may be the most severe recession that Australia has experienced since the Second World War. Furthermore, many countries (but not Australia) experienced extremely debilitating recessions in 2008–09 as they were swept into the vortex of the GFEC. In the US, unemployment peaked at just on 10 per cent – very high by US post-war standards – and many commentators there refer to the 2008–09 episode as that country's *Great Recession*. The economic recession due to COVID-19 seems more severe that the Great Recession, however, it is too soon to analyse the impact (at least at the time of writing the book).

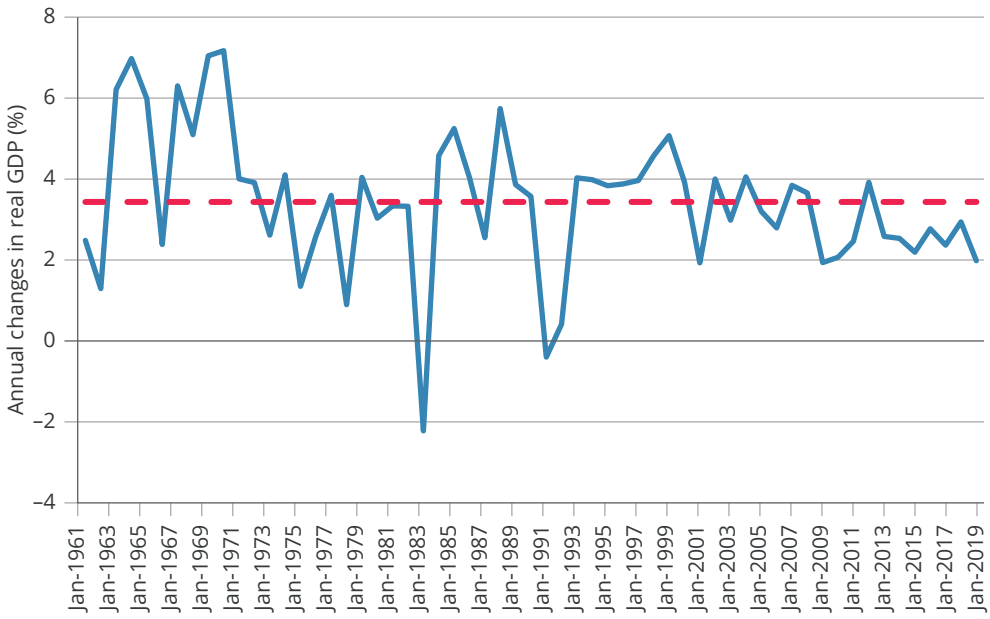
The trough is bad news and good news. It is simultaneously the bottom of the 'valley' of the downturn and the foot of the 'hill' of improving economic conditions called an *expansion*. Expansion is an upturn in the business cycle

during which real GDP rises, and sometimes quite rapidly, especially in the early quarters immediately following a trough. During the expansion phase of the cycle, profits generally improve, real GDP increases, employment begins to grow again, and over time the economy moves back towards full employment. However, employment increase will not be immediate since the firms will not hire or rehire workers at the onset of the expansion.

Finally, we will expand upon the definition of economic growth given in **Chapter 2**. Economic growth is an expansion in national output measured by the quarterly or annual percentage increase in a nation's real GDP. The growth trend line in the hypothetical model in part (a) of **Exhibit 12.1** is a stylised representation of the fact that over time our real GDP tends to rise.

This general, long-term upward trend in real GDP persists despite the peaks, recessions, troughs, and expansions. As shown by the dashed line in **Exhibit 12.3**, since 1960, real GDP in Australia has grown at an average annual rate of about 3.5 per cent. This may seem to be a small yearly change, but about 3.5 per cent annual growth will keep unemployment at low levels. It will lead to a doubling of real GDP (i.e., a doubling in the economy's productive capacity) in just 20 years – in other words, within a generation. One of our challenging ongoing economic policy goals is to maintain that long-term growth rate at around 3 to 4 per cent per annum.

Exhibit 12.3 **A historical record of the Australian Economic growth: 1960–2019**



Real GDP has increased at an average annual growth rate of 3.5 per cent since 1960–61. Across the business cycle, above-average annual growth rates have alternated with below-average annual growth rates. During a recession year, such as 1982–83 and 1990–91, the annual growth rate is negative and therefore below the zero-growth line. The economy entered the recovery phase in 1992 and, in the fiscal year 1992–93, the growth rate had rebounded to 4.1 per cent. In the years of the GFEC, while most other developed countries in the world went into recession with negative economic growth, Australia's lowest annual growth rate registered 1.9 per cent in the fiscal year 2008–09. In 2018–19, while the annual growth rate was a very low 1.9 per cent, the Australian economy nonetheless completed its 28th year of expansion, the longest in the post–Second World War era. However, the Australian economy ended its long span of expansions with the onset of COVID-19. The Australian economy went into a recession in 2020.

Source: Data calculated from Australian Bureau of Statistics, Cat. No. 5206, Table 36.

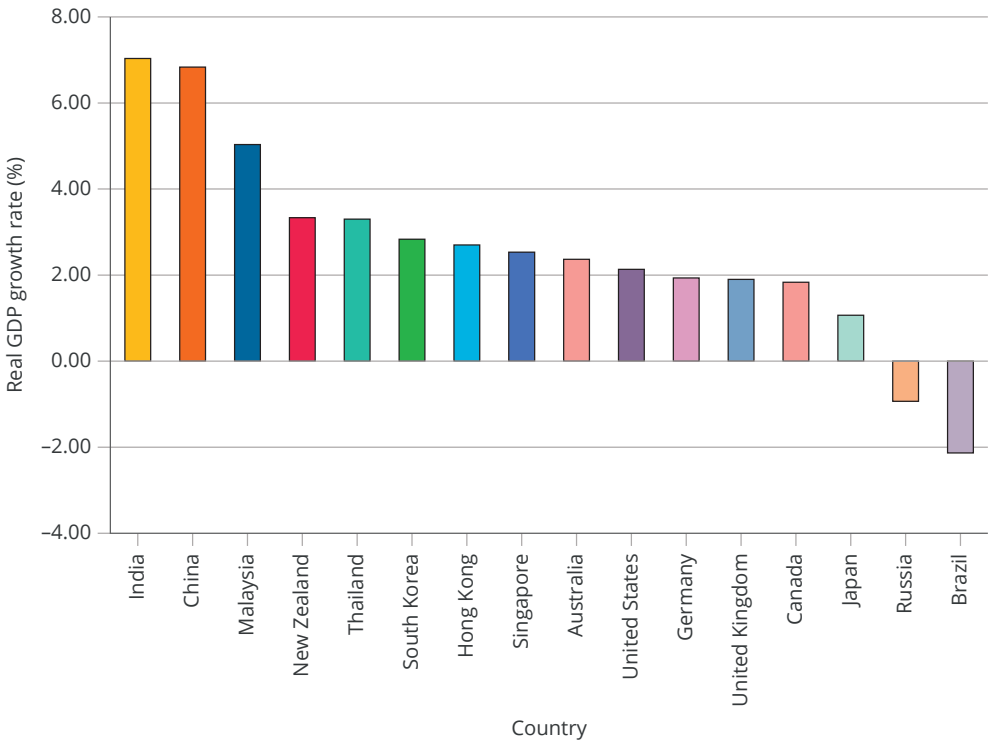
An examination of **Exhibit 12.3** reveals that the growth path of the Australian economy over time is not smooth, but consists instead of a series of year-to-year variations in real GDP growth. In some years, such as 1998 and 1999, the economy experienced above-average growth. In other years, such as 1983, the economy slipped below the zero-growth line. Note that the annual growth for the 1991–92 recession year was also negative and therefore dipped below the zero-growth line.

Subsequently, the economy entered the recovery phase in the second half of 1992, and the annual growth rate had rebounded to 4.1 per cent in the fiscal year 1992–93. In the years of the GFEC, while most other developed countries in the world went into recession with negative economic growth, Australia's lowest annual growth rate registered +1.9 per cent – in the fiscal years 2008–09. Australia had escaped the GFEC! In 2018-19, while the growth rate was a very low 1.9 per cent, the economy nonetheless completed its 28th year of expansion, the longest in the post–Second World War era. However, the Australian economy ended its long run of expansions with the onset of COVID-19. Australian economy went into a recession in 2020 and there will be long-term effects of the pandemic in the economy for years to come. Even though the Australian government and the Reserve Bank of Australia actively intervened in the economy to reduce the severity of the recession, these actions could not eliminate the recession completely.

Recent real GDP growth rates in a selection of countries

Exhibit 12.4 presents annual real GDP growth rates, averaged over the three years, 2015–17, for selected countries that effectively cover most of the globe – North and South America, Eastern and Western Europe, and Asia, and the Pacific. As the exhibit shows, economic growth rates vary significantly from country to country.

Exhibit 12.4 **Annual real GDP growth rates, averaged over 2015–17, selected countries**



Source: Central Intelligence Agency, *World Factbook*: https://www.cia.gov/library/publications/the-world-factbook/

At 7.03 per cent and 6.83 per cent, respectively, India and China registered the highest average economic growth within the selected countries. These are two of the four so-called BRIC (Brazil, Russia, India and China) countries, which are deemed to be at a similar stage of newly advanced and emerging economic development. The other two BRIC countries, Russia and Brazil, unfortunately, experienced the worst economic growth (−0.93 per cent and −2.13 per cent, respectively) among the selected countries over the period in question.

Where are we on the business cycle rollercoaster?

You're the economist

Suppose the economy has been in a recession and everyone is asking when it will recover. To find an answer to the state of the economy's health, a television reporter interviews Steven, a local car dealer. Steven says, 'I do not see any recovery in the economy. This year our first-quarter sales were down, and it was lower than the sales in the final quarter of the last year. Lower sales continued in the second quarter of this year again. We sold a lot more cars in the third quarter and the first and second quarters of this year, but the sales figures were not as high as it used to be in the third and fourth quarters last year'. Is Steven correct? Are his observations consistent in any way with either the recession or expansion phases of the business cycle or with the idea of a peak or trough?

In summary

- The **business cycle** refers to alternating periods of economic growth and contraction, and each cycle can be divided into two distinct phases: recession and expansion.
- A **recession** is a downturn in the business cycle in which output, sales, and employment decline, and an **expansion** is an upturn in the business cycle during which real GDP, employment, and other measures of aggregate economic activity rise.
- A business cycle **peak** is the endpoint of an expansion, and the **trough** is the endpoint of a recession.

2 Business cycle indicators

In addition to changes in real GDP, the media often report several other macroeconomic variables published by the ABS and other agencies that measure business activity. These economic *indicator* variables are usually classified into three categories based on the timing: leading indicators, coincident indicators, and lagging indicators. While the third set is mentioned in the discussion below, it is not often referred to these days, and so a detailed list is not provided here. **Exhibit 12.5** lists the variables corresponding to each of the other sets of indicators in the case of Australia.

One highly monitored forecasting gauge for business cycles in Australia is the Westpac–Melbourne Institute *Index of leading indicators* published monthly. **Leading indicators** are variables that change direction before the economy shifts from one phase into another (say, from an expansion into a recession). The Westpac–Melbourne Institute's index captures the headlines when there is concern over swings in the economy. For instance, it was reported that the Leading Index growth rate had weakened sharply over late 2019 to mid-2020, dropping from −0.28% in December 2019 to −4.79% in May 2020. pointing to a possible slowdown in economic growth in the latter part of 2019 and into mid 2020.[1]

Leading indicators Variables that change direction before the economy shifts from one business cycle phase into another.

The first set of variables in **Exhibit 12.5** is used to forecast the business cycle for about six months in advance. For example, a downturn ahead is signalled when declines exceed advances in the components of the leading

[1] See Westpac, 'Leading Index remains at recession levels', 17 June 2020, https://www.westpac.com.au/content/dam/public/wbc/documents/pdf/aw/economics-research/er20200617BullLeadingIndex.pdf (viewed in July 2020).

Exhibit 12.5 Business cycle indicators for Australia

Leading indicators
Average work week
New businesses formed as approximated by net demand for new telephone installations
New building approvals
Material prices
Stock prices
Money supply, M3
Gross operating surplus of companies
A macroeconomic measure of price mark-up: the ratio of the GDP deflator to unit labour costs
Coincident indicators
Unemployment rate
Total civilian employment
Household income
Industrial production
Retail sales
Gross domestic non-farm product

indicator data series. But beware! The leading indicators may rise for a couple of consecutive months and then fall for several months. Economists are therefore cautious and wait for the leading indicators to move in a new direction for several months before forecasting a change in the cycle.

Coincident indicators Variables that change at about the same time as the economy shifts from one business cycle phase into another.

The second group of variables listed in **Exhibit 12.5** is referred to as **coincident indicators**. Coincident indicators are variables that change at the same time as the economy shifts from one phase into another. For example, as real GDP rises, economists expect employment, personal income, industrial production and sales to rise, while the unemployment rate usually falls. The Melbourne Institute uses these coincident data series to determine the peak and trough dates in the Australian business cycle. The Economic Cycle Research Institute (ECRI) in New York City compiles business cycle peak and trough dates for many other countries around the world, and the US National Bureau of Economic Research (NBER) does this for the US.

Lagging indicators Variables that change direction after a phase change in the economy has occurred, thereby confirming its occurrence.

The third group of variables sometimes used in analyses of the business cycle are the so-called **lagging indicators**. Lagging indicators are variables that change direction after real GDP changes. For example, the duration of unemployment is a lagging indicator. As real GDP increases, the average time that workers remain unemployed does not fall until some months after the beginning of the recovery. The prime interest rate (the rate charged by banks on business loans to their best corporate customers) is another example. As noted previously, in practice lagging indicators are not used these days extensively, and so a comprehensive list is not supplied in **Exhibit 12.5**.

In summary

- Economic indicator variables are used to help better understand and predict the business cycle and are usually classified into three categories: leading, coincident and lagging indicators.
- **Leading indicators** change direction before the economy shifts from one phase into another.
- **Coincident indicators** change at the same time as the economy shifts from one phase into another.
- **Lagging indicators** change direction after the phase change in the business cycle.

3 Total spending and the business cycle

The uneven historical pattern of economic growth for modern industrial economies gives rise to the question: what causes business cycles? The theory generally accepted by economists today is that changes in total or aggregate expenditures and aggregate demand – are the principal cause of business cycle variations in real GDP.

However, as will be explained later in this text, business cycles can also be caused by shocks to so-called aggregate supply. For now, think of aggregate supply as a kind of amalgam of all the nation's microeconomic production and cost functions. A major shock to these – for example, as occurred in 1973–74 when the world price of oil quadrupled in less than a year – can throw the economy into disequilibrium and cause a recession. Notwithstanding supply-side shocks on occasion being an important driver of business cycle fluctuations, for the present purposes, we will focus our attention on the importance of fluctuations in aggregate expenditures.

Recall from the previous chapter that aggregate expenditure refers to total spending for *final products* by households, businesses, government and foreign buyers. Expressed as a formula, this is:

$$GDP = C + I + G + (X - M)$$

Why do changes in total spending cause the level of GDP to change? If the total spending is increasing, then businesses may find it profitable and increase output production. When firms increase production, they use more inputs of labour and capital (and sometimes more land). Hence, increased spending leads to economic growth in output, employment and incomes. When total spending falls, businesses will need to produce a lower volume of products and avoid unsold inventory accumulation. In this case, output, employment and incomes fall. These cutbacks can, in turn, lead to recession.

The situation described above assumes that the economy is operating below full employment so that any changes in total spending bring forth changes in the supply of goods and services to the market. Once the economy reaches full employment, increases in total spending will not change the production of output and hence have no impact on real GDP. Further spending at the full employment level will simply pull up the price level and 'inflate' nominal GDP, resulting in higher inflation.

In subsequent chapters will explain more about the causes of business cycles. Using the aggregate demand and supply analytical framework, described in **Chapter 14**, you will learn to understand why changes occur in national output, unemployment and the price level. Understanding the nature and determinants of the business cycle and the extent to which government policies can reduce its severity is the first primary concern of macroeconomics.

The GDP gap

When people in an economy would like to have a job but are unemployed, society forfeits the production of goods and services that such people could have helped produce. To determine the dollar value of how much society loses if the economy fails to reach full employment (see **Chapter 13** for a more detailed discussion of unemployment), economists estimate the **GDP gap**.

GDP gap
The difference between full-employment real GDP and actual real GDP.

The GDP gap is the difference between full-employment *real GDP* and *actual real GDP* in a particular period. Full employment does not mean an unemployment rate of zero. In a dynamic economy, with changing tastes and technology, there will always be some level of measured unemployment while people, having left their current employment, find suitable alternative employment. More will be discussed on this in **Chapter 13**.

The level of GDP that could be produced at full employment is also called *potential real GDP*. Because the GDP gap is estimated based on the difference between GDP at full employment and GDP at the actual unemployment rate, the GDP gap measures the cost of what is known as *cyclical unemployment*. GDP gap can be expressed as a formula:

GDP gap = cost of cyclical unemployment = potential real GDP − actual real GDP

In summary

- Business cycle expansions and recessions are generally thought to be caused by changes in aggregate spending in the economy.
- However, aggregate supply shocks can also on occasion be the cause.
- The impact of changes in aggregate spending on output and prices depends on whether the economy is operating below full employment output or at full employment output.
- **GDP gap** = a measure of the cost of what is called cyclical unemployment = potential real GDP − actual real GDP.

4 Economic growth in the longer term

While an economy may deviate from its underlying long-term growth path on account of shocks to aggregate demand for goods and services, in the longer term it is generally believed it will revert to some underlying long-term growth path. This long-term growth path can vary considerably across countries and, over time, these different trajectories can have a very significant impact on living standards.

A second central concern in macroeconomics is to understand the underlying determinants of economic growth and whether government policies may be able to influence an economy's growth trajectory. Consider **Exhibit 12.6** where the real GDP per capita data for selected years over more than half a century for a cross-section of countries is provided. These data give some sense of the sort of long-term economic growth experienced by different countries around the world, adjusted for changes in population. All data are provided in a common currency to allow useful comparisons across countries across time (thereby hopefully achieving an 'apples to apples' comparison).

As can be seen from the exhibit, Japan's real GDP per person increased almost seven-fold over the last half-century. At the same time, other countries experienced very little growth over the period! Indeed, why had Zimbabwe's GDP per capita in 2017 declined to be just over 35 per cent of what it was back in 1990?[2] And why, at the turn of the 21st century, was the US the dominant economic powerhouse in the world while the UK had held this position back in the 19th century?

Furthermore, over the last 50 years, the world has witnessed the strong and sustained economic growth of the 'Asian Tigers' (South Korea, Taiwan, Hong Kong and Singapore). Note the extraordinary 31-fold increase in GDP per capita in South Korea and the very impressive 24-fold increase in Singapore!

Finally, even though China's rapid economic ascent is evident in its 12-fold increase in real GDP per capita since 1960, in 2017, its GDP per capita was still only just about 23 per cent of that of the US. Nonetheless, it is quite incredible that China's GDP per capita has grown almost two folds (2.2 times) in the 17 years from 2000 to 2017, implying that while economic living standards in the industrialised countries represented in the exhibit have increased by between 30 and 50 per cent over that period, in the span of just one generation, economic living standards in China have improved by 260 per cent![3]

[2] Regrettably, the answer to Zimbabwe's descent since 1980 has less to do with its essential economic endowments, and more to do with the country's authoritarian and oppressive government policies in place from the early 1980s.

[3] However, this rapid growth has also produced some pretty substantial challenges for China, not least of which is a very significant increase in air and water pollution, along with very high and rapidly increasing emissions of greenhouse gases. Refer back to Chapter 11 for a refresher on why rapid economic growth does not necessarily equate to rapid increases in welfare.

Exhibit 12.6 **Long-term economic growth: an international comparison of changes in GDP per capita for selected countries, 1960–2017**

	1960 GDP per capita	1980 GDP per capita	2000 GDP per capita	2017 GDP per capita	GDP per capita growth multiples: 1960–2017	GDP per capita growth multiples: 1980–2017	GDP per capita growth multiples: 2000–2017
Australia	14 223.3	22 172.1	35 218.3	47 392.6	2.3	1.1	0.3
Bangladesh	1 429.3	1 206.9	1 370.7	3 436.0	1.4	1.8	1.5
Brazil	2 114.2	5 331.9	8 570.5	13 812.9	5.5	1.6	0.6
China	1 014.1	1 824.6	4 108.3	13 042.6	11.9	6.1	2.2
Germany	9 937.1	19 533.2	33 330.5	47 556.0	3.8	1.4	0.4
India	1 040.9	1 162.6	2 004.1	6 421.6	5.2	4.5	2.2
Indonesia	1 281.8	2 212.2	3 449.2	10 594.1	7.3	3.8	2.1
Japan	4 954.4	18 980.9	34 570.5	40 373.7	7.1	1.1	0.2
NZ	12 835.8	16 265.3	26 470.4	36 537.8	1.8	1.2	0.4
South Africa	5 913.8	8 954.6	8 718.7	11 741.2	1.0	0.3	0.3
Singapore	2 644.9	18 349.8	36 042.1	67 138.1	24.4	2.7	0.9
South Korea	111 3.251	4 830.7	21 825.2	36 264.5	31.6	6.5	0.7
UK	11 909.77	19 839.9	31 526.2	39 127.8	2.3	1.0	0.2
USA	17 462.63	29 154.8	45 859.2	54 794.7	2.1	0.9	0.2
Zimbabwe	1 990.9	2 543.6	2 600.2	1 871.8	−0.1	−0.3	−0.3

Note: GDP per capita, 2011 international $ (2011 international $ representing the purchasing power).

Source: https://ourworldindata.org/grapher/gdp-per-capita-inflation--and-ppp-adjusted-world-bank-data-vs-penn-world-table-data.

What might be the explanation for these widely differing growth performances? What are the essential factors that drive a country's long-term economic growth? Robert Solow (**Solow model of economic growth**) provided deeper understanding of the early insights contributing to long-term economic growth.[4]

Solow model of economic growth
An early growth model that sought to explain how consumption, saving, capital, labour and technological change combine in the longer term to determine a nation's economic growth.

At the most basic level, a country's economy grows over time as a result of growth in both demand and supply. At each point in time, it is the interaction of these forces that determines the quantity of output produced, consumed, saved and invested.

Over the longer term, a community's demand for goods and services cannot outstrip its productive capacity and, equally, without sufficient increase in demand, there will be no incentive for an increase to occur in the productive capacity of the country.

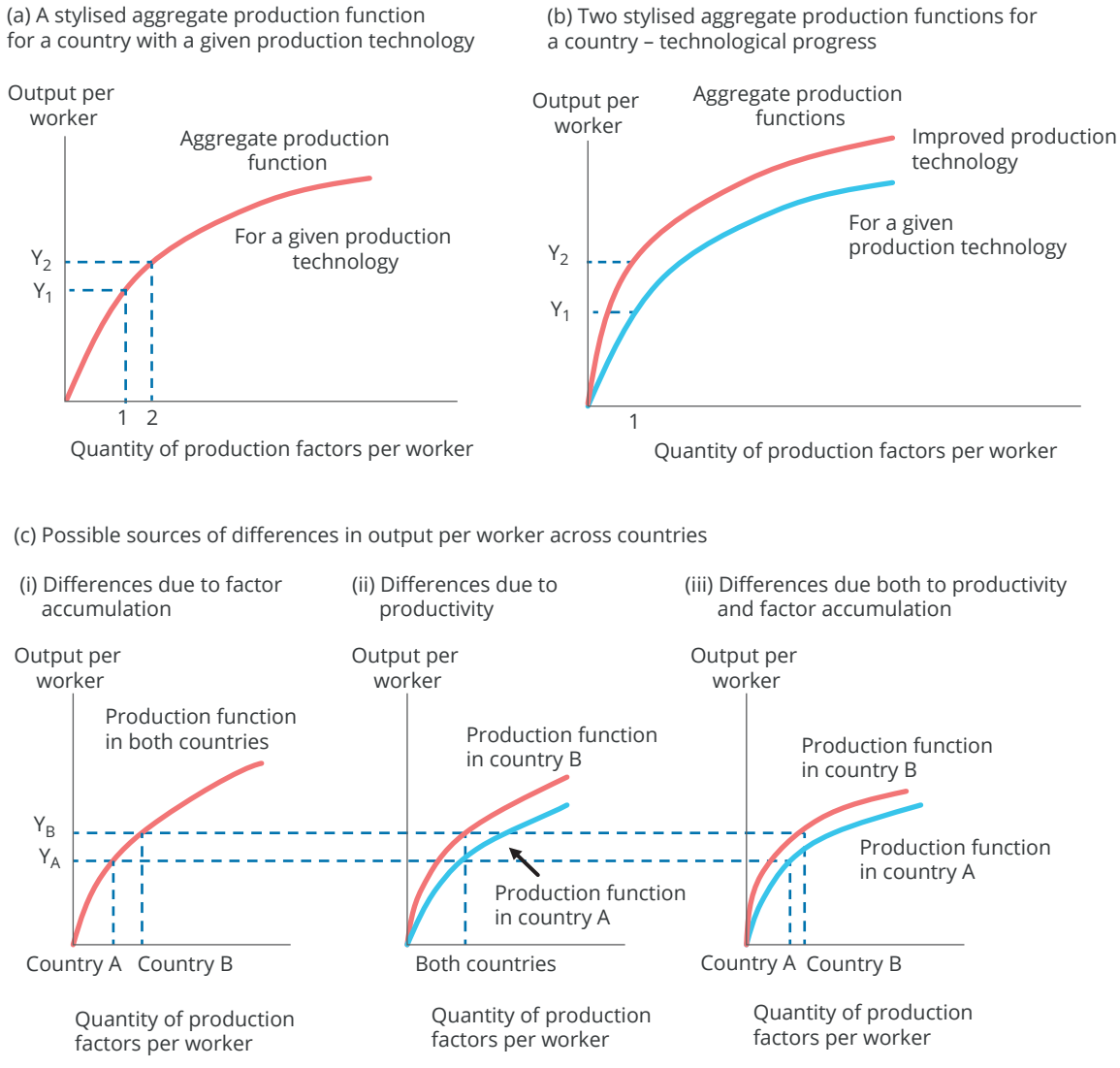
The determinants of growth

Let us consider the supply side first. Fundamentally, output growth is the result of increases in the productive inputs land, labour and capital – combined with improvements in technology – into the production process over time.

[4] His first significant article on the topic was, 'A contribution to the theory of economic growth' (*Quarterly Journal of Economics,* 70(1), 1956, p. 65–94). A more recent book by David Weil (2014), *Economic Growth* (3rd edn), provides excellent foundation reading on the topic, while a more mathematical treatment of growth theory is provided in *Macroeconomics* (9th edn) by N. Gregory Mankiw (2015). The growth discussion provided in the text borrows somewhat from both Weil's and Mankiw's presentations.

To simplify our discussion below, we will refer to the non-labour production factor inputs collectively as simply 'production factors'. Now recall the concept of the production function introduced in **Chapter 6** and imagine, for argument's sake, that all of a country's many separate existing micro-level production processes were represented by just one stylised aggregate production function, as illustrated by **Exhibit 12.7**.

Exhibit 12.7 Stylised aggregate production functions for a country



Adapted from David N. Weil, *Economic Growth*, Addison-Wesley, USA, 2005, pp. 34–5.

The aggregate production function of part (a) of **Exhibit 12.7** indicates that, as the quantity of production factors per worker increases, the output per worker in the country increases, but at a decreasing rate. This is again just the law of diminishing marginal productivity in action, which was first introduced in **Chapter 6**.

In part (a) of **Exhibit 12.7**, for a given production technology, a country may increase its output per worker only by increasing the quantity of production factors per worker. For example, by increasing production factors from point 1 to point 2, the country can increase its output per worker from Y_1 to Y_2.

Now consider the impact of a wide-ranging improvement in production technologies available in the country (perhaps someone may have invented something like the internet) that allowed any given quantity of production factors to be used more productively by the workforce to produce higher levels of output per worker.

This is illustrated in part (b) of **Exhibit 12.7** by an upward shift in the country's aggregate production function. In this part of the exhibit, the country is enjoying higher output per worker – Y_2 versus Y_1 – as a result of using a new, more productive technology when the production inputs per worker held constant (this quantity is unchanged at point 1).

Part (a) and (b) of **Exhibits 12.7** illustrate a compelling yet straightforward point: a country may enjoy a higher level of output per worker – or, in the language of the earlier discussion, a higher level of real GDP per person – by either somehow increasing the quantity of production factors available to the workforce or by improving its available production technologies.

This logic also applies when we think about different countries, and this is illustrated in part (c) of **Exhibit 12.7**. In each case, country B is experiencing higher output per worker or real GDP per person. In the first panel, both countries are using the same production technologies. However, country B has a higher output per worker because each of its workers has a higher quantity of production factors available to use. In the second panel, both countries' workers have the same quantity of production factors available to them, but country B has superior production technologies. In the third panel, a combination of these two determinants contributes to country B enjoying a higher level of output per worker than country A.

Based on the discussion, therefore, those countries mentioned earlier that have recorded relatively higher growth in real GDP per person over time have either been able to more rapidly accumulate the quantity of production factors per worker, or been able to increase the productivity of their production technologies more rapidly, or a combination of the two.

This being the case, one is naturally led to ask how a country might increase the quantity of production factors available to each worker and/or improve its production technologies. Let us take the first issue first.

Production factor accumulation

The amount of production factors available to workers will reflect the amount of investment per person a country is making, and this leads to the consideration of the other side of the economic growth equation, that is, demand.

Fundamentally, a nation's output is either consumed or used for capital accumulation. The output not consumed is the saving of the inhabitants, which in turn allows for the investment undertaken by firms.[5]

Thus, it is saving that is a crucial determinant of the quantity of investment in the economy. Higher levels of saving per person imply higher levels of investment per person, a larger quantity of production factors per person, and therefore higher levels of output per person.

Policymakers can put in place suitable incentives to assist a community in transitioning to a higher saving rate. However, the task may not be easy politically. In the longer term, the community would be better off in terms of higher ongoing consumption per capita; but in the short term, the community will have to give up current consumption! The situation is complicated even further when it is recognised that the future beneficiaries of the short-term sacrifices may be quite different from the group making the present-day consumption sacrifice.

[5] As explained briefly in Chapter 11, a nation can augment its investment by also utilising the saving of foreigners. However, for our purposes at the moment, we can abstract from this consideration since the essential ideas discussed here are applicable to the more complex (but more realistic) situation also.

The impact of technological change

The above discussion leads invariably to the exciting conclusion that, for a given state of unchanging production technologies, eventually a maximum level of saving and investment per person would be attained. Beyond that, a country's inhabitants would no longer willingly give up any further current consumption to allow even higher rates of saving and investment to occur. At this point, there would be no further growth in the country's consumption per capita; that is, living standards would, at some point, cease to improve!

What is the explanation for the persistent growth in living standards in most countries over the past couple of centuries? The ultimate answer to this must be the other fundamental determinant of growth, namely, technological change.

Furthermore, given that improvements in production technologies enhance the productivity of labour (and capital), it is important to note that there are other ways to enhance the productivity of labour than through what might be called 'pure' technological progress. Better organisation of the workforce, improved management practices and a more skilled, better educated and more highly trained workforce can add to the productivity of labour and capital. All of these things can sensibly be grouped under the general heading of technological progress.

Following on from these considerations, an important question to ask is: How does technological progress occur and can its rate be influenced by government policy?

We will now very briefly introduce a modern model of economic growth, known as the endogenous growth model.

Economics and ethics Do we need only economic growth?

In his article, 'The Moral Consequences of Economic Growth' (2006),[1] Benjamin Friedman writes as follows.

> Are we right to care so much about economic growth as we clearly do? ... Perhaps because we are never clear about just why we attach so much importance to economic growth in the first place, we are often at cross-purposes ... about what we want. The root of the problem, I believe, is that our conventional thinking about economic growth fails to reflect the breadth of what growth means ... for a society.
>
> Moral thinking, in practically every known culture, enjoins us not to place undue emphasis on our material concerns. ... We, therefore, think of economic growth in terms of material considerations versus moral ones. ... I believe this thinking is seriously, in some circumstances, dangerously, incomplete. The value of a rising standard of living lies not just in the concrete improvements it brings to how individuals live but in how it shapes the social, political and, ultimately, the moral character of a people. Economic growth – meaning a rising standard of living for the clear majority of citizens – more often than not fosters greater opportunity, tolerance of diversity, social mobility, commitment to fairness, and dedication to democracy.
>
> Ever since the Enlightenment, western thinking has regarded each of these tendencies positively, and in explicitly moral terms. ... How the citizens of any country think about economic growth, and what actions they take in consequence, are therefore a matter of far broader importance than we conventionally assume. ... The familiar balancing of material positives against moral negatives when we discuss economic growth is, therefore, a false choice. ... Economic growth bears moral benefits as well ... and it is important that we take these moral positives into account.

[1] **Reprinted by permission from Springer Nature. *Society*®, 'The Moral Consequences of Economic Growth', by Benjamin M. Friedman, Jan/Feb, 2006, pp. 15–22.**

Endogenous growth model

In early economic growth models, technological progress was simply assumed to be independent of the economic growth process. However, is this a realistic assumption?

Rather than technological progress being **exogenous** – or outside of the economic growth process – perhaps it is more realistic to regard it as **endogenous**. In addition to technological progress being a determinant of the rate of economic growth, it may also inherently be determined by the economic growth process itself. This being the case, the rate of technological advance would in part be the direct outcome of economic growth and could go on indefinitely.

Exogenous technological progress Technological progress that occurs independent of, or outside of, the economic growth process.

Endogenous growth model The view that the process of technological change is significantly endogenous to economic growth rather than being exogenous, as in earlier economic growth models.

Recognition of the dominant role played by the process of knowledge acquisition and application in the growth process implies that governments can potentially play a very significant decisive role. For example, once knowledge exists and is used in the economic process, it is difficult for any group to maintain a monopoly of that knowledge. However, the research and development (R&D) required to develop it can involve substantial investments of resources. Therefore, R&D will be carried out by private firms if it earns profits and only if they feel they can get an advantage over their rivals.

The patent system is a good example of government's recognition that it is in society's longer-term interest to grant an individual a temporary monopoly right (the granting of a patent) over discovery or technological development. This time-limited monopoly right provides some incentive for firms to engage in significant R&D.

It is nonetheless reasonably clear that an individual firm will probably not be able to capture all the resulting benefits from its R&D expenditure and there will inevitably be some positive externalities that benefit other firms in the economy. From the standpoint of society, these knowledge externalities provide benefits to the society; however, from the viewpoint of the individual firm incurring the original R&D cost, they represent a disincentive to invest.

The existence of these knowledge externalities is the basis of arguments supporting the government providing tax breaks for R&D investment. The fact that the potential societal benefit of firms' R&D expenditures may be higher than their private benefits is a justification for the government to consider some sort of tax incentive or subsidies to encourage more private R&D expenditure. Thus, we are led to the conclusion that government policies designed to encourage further knowledge development – even if such policies involve considerable subsidisation from the community's taxpayers – may pay handsome dividends to the community in the longer term.

Applicable concepts: economic growth and government policies

Global perspective

Tried and true way to economic growth

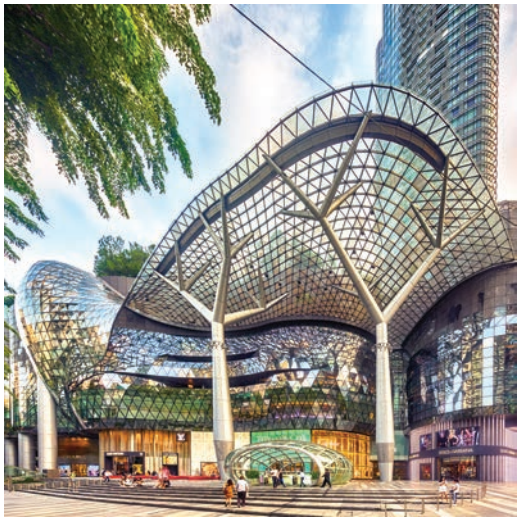
The following is an edited version of an article that appeared in the *Australian Financial Review*.[1]

> Wolfgang Kasper says that good institutional rules make for prosperous nations.
>
> The Heritage Foundation in Washington published convincing evidence in support of a new theory of economic growth, namely that 'it is the traffic rules and the institutions ... that make for poverty or prosperity!'
>
> Recent empirical research by Americans John Talbot and Richard Roll has also shown that 85 per cent of all differences between the poorest and richest societies can be explained by differences in the protection of private property, civil liberties, political and press freedom, as well as the absence of black markets, discriminatory regulations, inflation and barriers to free trade.

Post-war economic reforms in Germany and Japan, subsequent improvements in economic freedom in East Asia, even the People's Republic of China have been rewarded with economic prosperity.[2]

Interest groups that seek and extract political favours, parliaments and judges that grant them and politicians that distribute opportunistic handouts have to be seen as the main enemies of broad-based prosperity.

[One] implication for national policymakers ... is to make foreign aid conditional on institutional reform. With economic reforms that implement high standards of economic freedom, a good investment climate will be ensured to attract both domestic and foreign investors so that eventually aid becomes superfluous.



Getty Images/John Seaton Callahan

Another policy implication is that a high investment rate and high spending on research and development are not the main drivers of prosperity. Government policies that promote investment or R&D by artificial means induce waste of resources ... introduce favouritism, and detract from the equality of all before the law.

This research that looks at the fundamental institutions of economic freedom and genuine economic reforms is rapidly changing the policy scene overseas. It will eventually trickle down to Canberra and State capitals. The question is: with what time lag? And will the electorate accept the lessons?

What do you think?

Analyse the quote provided in the context of the text's discussion of growth theory. If you disagree with it, outline the basis of your disagreement. If you agree, what broad role would you see as appropriate for government and policy to play?

1 **Wolfgang Kasper, 'Tried and true way to economic growth', *Australian Financial Review*, 6 December 2001, p. 55. The use of this work has been licensed by Copyright Agency; except as permitted by the Copyright Act, you must not re-use this work without the permission of the copyright owner or Copyright Agency.**

2 **And, unfortunately, in the case of Zimbabwe, the absence of such economic and social freedoms, and the lack of protection of private property rights are very significant components of the explanation for that country's moribund economy over the past 30-plus years (refer again to Exhibit 12.8 and footnote 2).**

You're the economist

Should the government subsidise private-sector R&D expenditure?

The R&D tax incentive encourages companies to engage in R&D benefiting Australia by providing a tax offset for eligible R&D activities. It has two core components: a refundable tax offset for certain eligible entities whose aggregated turnover is less than $20 million, and eligible for a tax benefit to offset some of the cost of eligible R&D activities. This may be a refundable or non-refundable offset depending on the company.

The Morrison government's reintroduced R&D tax incentive reforms and will curtail Australian innovation and punish firms that are incurring higher costs. The government moved forward with its controversial R&D tax incentive scheme and will introduce an annual $4 million refund cap. Commonwealth coffers will save an estimated

$1.8 billion through to 2022–23 under the proposed changes, which by definition is money that will no longer subsidise private-sector research.

Deloitte said the current intensity measures, which apply to non-refundable claims, would deliver a lower net tax benefit to the companies, noting that 'The reduced benefits will result in companies opting out of the Australian R&D regime and potentially shifting operations offshore — New Zealand'.[1]

What is your assessment of this policy issue?

[1] **Deloitte, 'Revised R&D Tax Incentive Bill introduced into Parliament', 5 December 2019, https://www2.deloitte.com/content/dam/Deloitte/au/Documents/tax/deloitte-au-tax-revised-rd-tax-incentive-091219.pdf**

In summary

- Two important long-term economic growth drivers are a country's rate of saving and investment per person; and its rate of technological progress.
- A modern view of the role of technological progress is that it is **endogenous** to the economic growth process.
- Taxpayer subsidies for private-sector R&D and the existence of the patent system may be justified by the high cost of R&D and the benefits to wider society associated with knowledge development.

5 The goals of macroeconomic policy

In **Chapter 11**, we introduced important issues dealing with the measurement of the macroeconomy. In this chapter, we have extended that to consider how an economy grows over time. In particular, we have highlighted the determinants of its long-term growth path, as well as the fact that, over the shorter term, an economy experiences fluctuations around this long-term growth path, known as business cycles. In the following chapters, we will extend this by elaborating on the **goals of macroeconomic policy** in some detail, but first we will briefly outline them here.

Goals of macroeconomic policy
These are: first, to reduce the severity of the ups and downs of the business cycle; and second, to assist with the achievement of the highest possible sustainable economic growth trajectory consistent with the economic, social and cultural aspirations of the community.

First, it is desirable wherever possible to follow policies that will reduce the severity of the ups and downs of the business cycle. There are two aspects to this. For reasons that will be explained **Chapter 13**, it is desirable to reduce the severity of the recession, thereby limiting the rise in unemployment that is inevitably associated with such downturns. It is, however, also important to try to avoid situations where the economy is growing 'too fast'. Such periods often ultimately result in excessively increasing inflation (see the **Chapter 13**), leading to poor business and investment decisions, which in their turn can precipitate a subsequent recession.

Second, where possible, macroeconomic policies should be followed that enable the economy to achieve the highest possible sustainable long-term economic growth trajectory that is consistent with, and will facilitate the achievement of, the economic, social and cultural aspirations of the community.

Of course, altering a nation's long-term growth trajectory will entail a range of government policies, many of which will be more microeconomic in nature. However, the good macroeconomic policy will undoubtedly have a role to play. For example, many economists believe that business cycle downturns have not only a short-term negative economic impact on the community but, long-term effects on economic growth. These downturns may adversely affect the capital stock and the skills of the workforce and permanently affect the long-term growth path of the economy. Also, raising a nation's saving rate may well involve a raft of government expenditure and taxation policies (see **Chapter 17** on fiscal policy) that may be regarded as macroeconomics.

More will be said on these issues in later chapters.

Analyse the issue

Applicable concepts: productivity growth and economic wellbeing

If productivity growth increases unemployment, does it, in fact, reduce economic wellbeing and therefore should it be resisted?

Some technological change may, in the short run, throw people out of a job in a particular firm or industry. However, the fact that there are some losers from technological change is no reason to resist it. Notwithstanding this, there is nonetheless an essential role for government to play here to provide appropriate transitional arrangements to enable the displaced people to retrain or relocate to find gainful employment again elsewhere in the economy.

As a simple thought experiment, imagine an island economy where some of the inhabitants gain their livelihood from fishing, others from growing and gathering coconuts and so on. Let us also suppose the technology used for fishing is by a spear from the beach or off rocks. As a result, there is also a group of spear makers who make a nice living from the fishing industry by supplying their capital (the spears).

Now suppose one morning the inhabitants wake up to find a few boats (they've never seen these before) and fishing nets in good repair washed up on the beach. Some fisherfolk experiment with this new technology and find they can catch a lot more fish in half the time by getting out into deeper water to fish. Before you know it, the small number of boats and nets are appropriated.

After a few days, others in the fishing industry realise they are using outmoded technology and find they are being undercut by those using the new methods. As a result, they find they are not getting as many fish for their efforts as they used to. The spear makers are also concerned at developments.

These two groups immediately band together to decry the unfair trade practices of the few with the new technology and try to get the community's elders to ban it. They succeed, and the boats and nets are immediately burned, and things quickly return to the way they were before.

Shutterstock.com/Strahil Dimitrov

What do you think?

Given this outcome, the community is no worse off than before, but it is no better off either. Think about what alternative actions could have been taken when the little island community found the new technology potentially very beneficial, but socially disruptive. What could have happened to the community's economic wellbeing? How could short-term transitional costs have been handled? What lessons does this parable teach us? Can you think of a current or recent real-world issue to which you can see the relevance of this little allegory?

In summary

The **goals of macroeconomic policy** are twofold:

- First, to reduce the severity of the recession, as well as to try to avoid situations where the economy grows 'too fast'.
- Second, to enable the economy to achieve the highest possible sustainable long-term economic growth trajectory consistent with the economic, social and cultural aspirations of the community.

Key concepts

Business cycle
Recession
Expansion
Peak
Trough
Economic growth
Leading indicators
Coincident indicators
Lagging indicators
GDP gap
Solow model of economic growth
Exogenous technological progress
Endogenous growth model
Goals of macroeconomic policy

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand the concept of the business cycle 'rollercoaster' and its two phases

- The **business cycle** refers to alternating periods of economic growth and contraction, which can be dated by changes in output, income, sales and employment measures. Each business cycle can be divided into two distinct phases: **recession** and **expansion**. A *recession* is a downturn in the business cycle in which output, sales and employment decline. In practice, a recession is usually defined as occurring when real GDP contracts in each of at least two consecutive quarters. An *expansion* is an upturn in the business cycle during which real GDP, employment and other measures of aggregate economic activity rise. In practice, an expansion is regarded as having commenced when real GDP expands for at least two consecutive quarters. The endpoint of an expansion is referred to as the **peak**, and the endpoint of a recession is referred to as the **trough**.

2. Develop an understanding of business cycle indicators, including how they are used to determine the phase of the business cycle and to forecast where it may be headed in the near future

- Economic indicator variables are used to help better understand and predict the business cycle. They are usually classified into three categories:
  - **Leading indicators**: variables that change direction before the economy shifts from one phase into another (say, from an expansion into a recession).
  - **Coincident indicators**: variables that change at the same time as the economy shifts from one phase into another.
  - **Lagging indicators**: variables that change direction after the phase change in the business cycle.

3. Appreciate the importance of the relationship between changes in total spending and the business cycle

- Business cycle expansions and recessions are generally thought to be caused by changes in aggregate spending in the economy – also known as aggregate demand – although a shock to the production side of the economy – also known as aggregate supply – can also on occasion be the cause. In an economy operating below full employment, changes to aggregate spending translate into a change in production and employment. At full employment level of output further increases in aggregate spending result only in higher prices (inflation). The difference between the actual real GDP produced in a period and what could have been produced had the economy been operating at full employment (also known as potential real GDP) is known as the **GDP gap**, which is a measure of the cost of what is called cyclical unemployment, discussed further in **Chapter 13**.

4. Understand the long-term determinants of economic growth and what governments might do to influence a country's long-term rate of economic growth

- Two important factors that drive a country's long-term growth path are:
  - its rate of saving – and investment – per person which determines the quantum of capital at the disposal of each of the country's inhabitants
  - its rate of technological progress.
- A modern view of technological progress is that it is **endogenous** to the economic growth process; that is, not only is it a determinant of economic growth but it is also inherently determined by the economic growth process itself.
- The significant cost of R&D and the existence of significant positive externalities associated with knowledge development imply there is some justification for taxpayer subsidies to encourage private sector R&D as well as the existence of the patent system.

5. Appreciate the goals of macroeconomic policy in relation to the business cycle and economic growth

- The goals of macroeconomic policy are twofold:
  - It is desirable wherever possible to follow policies that will reduce the severity of the recession, thereby limiting the rise in unemployment, as well as to try to avoid situations where the economy is growing 'too fast', which might lead to inflation becoming a problem.
  - To enable the economy to achieve the highest possible sustainable long-term economic growth trajectory that is consistent with the economic, social and cultural aspirations of the community.

Study questions and problems

1 What is the business cycle, and what is its fundamental cause?

2 What are the two phases of the business cycle? Explain. What do you call the point at which a business cycle expansion finishes? What is the point at which a business cycle recession finishes?

3 What are the three main types of business cycle indicators? Explain.

4 Shown here are real GDP figures for each of 10 quarters.

Quarterly real GDP (billions of dollars)

Quarter	Real GDP (billions of dollars)	Quarter	Real GDP (billions of dollars)
1	250	6	400
2	300	7	500
3	350	8	600
4	325	9	575
5	300	10	620

Plot these data points and identify the two phases of the business cycle. What are some of the consequences of a prolonged decline in real GDP? Is the decline in real GDP from $600 billion to $575 billion a recession? Why or why not?

5 If actual real GDP moves well below potential GDP, what do you call this difference between the actual real GDP and the potential GDP? What would you say might be the impact on measured unemployment from such a movement?

6 Using models of aggregate production functions, explain the effects of increasing production factors per worker and the effects of improved production technology?

7 Indicate whether each of the following is likely to be associated with a high or low level of long-term economic growth.

		High	Low
a	Highly skilled workforce		
b	Political instability		
c	High savings rate		
d	Low rate of capital accumulation		
e	Advanced technology		

8 What might be the two fundamental explanations for the increase in economic living standards (output per capita) in most developed countries than in developing countries around the world?

9 Does the pursuit of high economic growth have positive or negative moral consequences? Explain.

10 Describe the goals of macroeconomic policy.

Answers to 'You're the economist'

Where are we on the business cycle rollercoaster?

Steven the car dealer's sales in the first and second quarters of this year confirm that the economy is in a recession phase of the business cycle. The fact that sales increased in the third and the fourth quarters of the second quarter can be viewed as the end of the recession phase – trough. This suggests that the economy may have entered a recovery phase in the third quarter. The fact that sales in these quarters were below the corresponding quarters in the previous year may simply be a reflection of the severity of the recession from which the economy was just beginning to recover. If you can see that real GDP during the first few quarters of an expansion, but it can be lower than real GDP in corresponding quarters in previous years, YOU ARE THINKING LIKE AN ECONOMIST.

Should the government subsidise private-sector R&D expenditure?

The government is providing tax incentives for R&D expenditures since the existence of significant spill-overs to the rest of the community when a private firm attempts to produce new knowledge as a result of its R&D. In theory, then, such a policy has a considerable justification since spending on R&D is considered as a positive externality. The government provides these tax incentives and benefits from the taxpayer money. However, some private investors will misuse the system in order to improperly claim expenses as R&D expenditures when they are just ordinary business expenses. To the extent that this occurs, taxpayers are unnecessarily and unfairly increasing the profits of such firms with no wider benefit to society. If the government were to reintroduce or continue with a substantial tax incentive, there would need to be reliable measures included in the policy to guard against such attempted rorts. If you thought about the issue in these terms, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

(Note: For some of the following more than one answer may be correct)

1 The phases of a business cycle are:
- **a** upswing and downswing.
- **b** full employment, depression, expansion and plateau.
- **c** full employment and unemployment.
- **d** recession and expansion.

2 Which of the following is *not* a coincident indicator?
- **a** Personal income.
- **b** Industrial production.
- **c** Money supply, M3.
- **d** Manufacturing and trade sales.

3 The point at which a recession ends is called the:
- **a** expansionary phase.
- **b** trough.
- **c** contractionary phase.
- **d** peak.

4 The difference between the real GDP produced at full employment and the real GDP measured at the actual unemployment rate measures the cost of:
- **a** full employment.
- **b** actual unemployment.
- **c** cyclical unemployment.
- **d** natural rate of unemployment.

5 The goal(s) of macroeconomic policy is/are:
- **a** to maximise economic growth consistent with the community's social and cultural aspirations.
- **b** ensure the stock market grows on average at the rate of growth of nominal GDP.
- **c** to keep interest rates as low as possible.
- **d** to reduce the severity of the ups and downs of the business cycle.

6 The two fundamental determinants of growth in output per capita in the longer term are:
- **a** aggregate demand and aggregate supply growth.
- **b** consumption and investment per capita growth.
- **c** growth in production factors per person and technological change.
- **d** industrial production and sales per capita growth.

7 The growth model, which posits that the process of technological change is assumed to be influenced by the economic growth process itself is known as:
- **a** the exogenous growth model.
- **b** the aggregate demand/aggregate supply model.
- **c** the endogenous growth model.
- **d** the business cycle economic growth model.

8 Which of the following do you think best characterises the vicious cycle of poverty in many less developed countries?
- **a** People cannot save while accumulating capital and capital cannot be accumulated while people are saving.
- **b** Poverty, saving and investment go through cycles just like the business cycle.
- **c** GDP per capita must rise before people can save and invest, but saving and investment are vital contributors to rising GDP per capita.
- **d** It is a cyclical certainty that the children of poor people will always be poor, just like their parents.

9 The steady and sometimes rapid increase in living standards in many countries over the 20th century can be explained:
- **a** by a steady increase in production factors per worker.
- **b** by steady and sometimes rapid technological progress.
- **c** by steadily increasing government involvement in actively assisting selected industries and companies predicted by government industrial experts to produce significant economic benefits to the community.
- **d** By all of these options given.

10 Because economic growth means greater material wealth for a country's inhabitants, a goal of higher economic growth:
- **a** should not be pursued because it inevitably leads to eroding community morality.
- **b** more often than not fosters greater opportunity, tolerance of diversity, social mobility, and dedication to democracy, all usually regarded as morally positive.
- **c** needs to be considered in terms beyond the material improvements it brings to how individuals live but in how it may also shape the social, political and, ultimately, the moral character of a people.
- **d** both b and c.

13 Inflation and unemployment

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand the meaning and measurement of inflation.
2 Develop an understanding of the consequences of inflation.
3 Examine the meaning and measurement of unemployment.
4 Understand the different types of unemployment.
5 Examine whether there is a relationship between inflation and unemployment.

Getty Images/Anthony Devlin – PA Images

Maintaining an economy at or near full employment with strong, sustainable economic growth and keeping prices reasonably stable are probably the most important economic goals of a nation. Around the world, the Great Depression of the 1930s produced profound and long-lasting changes in people's lives, leaving millions of people unemployed – or underemployed – for years.

And while Australia escaped recession, and experienced only relatively modest increases in unemployment as a result of the 2008–09 global financial and economic crisis (GFEC), many people in many countries around the world were thrown into long-term unemployment from the synchronised international recession of those years. Some may never again be re-employed, at least not in their preferred pre-GFEC job and salary. Australia and most countries in the world experienced very high unemployment rates due to the COVID-19. It is yet too early to comment on the full impact on the level of unemployment due to the pandemic (at least at the time of writing the book, we have seen a significant increase in unemployment). Similarly, the 'Great Inflation' of the 1970s and early 1980s left memories of the miseries of inflation. As with high unemployment, voters are very quick to blame any government that fails to keep inflation under control.

This chapter explains what is meant by unemployment and inflation. You will study how the government measures unemployment and how it computes the rate of inflation. The chapter includes a discussion of the consequences of both unemployment and inflation. For example, you will learn about the types of unemployment, what 'full employment' is, and what the non-monetary and demographic consequences of unemployment are. Similarly, you will learn about demand-pull and cost-push inflation, who wins from inflation and who loses. After studying this chapter, you will have a much clearer understanding of why unemployment is such a burden on society and why inflation is so feared.

1 Meaning and measurement of inflation

Sixty years ago a typical small bottle of soft drink sold for about 10 cents in Australia (actually 10 pence in those days). Nowadays, a similar drink sells for say 20 times that price. However, the soft drink has not become 20 times more intrinsically valuable than it was in the 1950s.

Much of the price rise is simply accounted for by a general increase in the prices of goods and services produced in the economy. **Inflation** is an increase in the general (average) price level of goods and services in the economy. The opposite of inflation is **deflation**, which is a decrease in the general (average) price level of goods and services in the economy (negative inflation). Note that inflation does not mean that all prices of all products in the economy rise during a given period (similarly for deflation).

You will sometimes also hear the word **disinflation**. This is a situation in which the rate of inflation is dropping (from, say, 10% per annum to 5% per annum).

Inflation
An increase in the general (average) price level of goods and services in the economy.

Deflation
A decrease in the general (average) price level of goods and services in the economy.

Disinflation
A reduction in the rate of inflation.

The consumer price index

The most widely reported measure of inflation is the **consumer price index (CPI)**. The CPI is an index that measures changes in the average price of consumer goods and services. It is sometimes called the cost-of-living index. The CPI includes only consumer goods and services in order to determine how rising prices affect the income of consumers. CPI includes over 100000 goods and services purchased by households. As such, the CPI does not cover items purchased by businesses and government, nor does it include items that are exported. However, it does measure changes in the prices of imported consumer goods and services. Such imported products are nonetheless of interest to consumers, so their price movements are incorporated into the calculation of the CPI.

Consumer price index (CPI)
An index that measures changes in the average prices of consumer goods and services.

The Australian Bureau of Statistics (ABS) prepares and publishes the CPI every quarter. Of all Organisation for Economic Co-operation and Development (OECD) countries, only Australia and New Zealand compile CPI quarterly. All others compile it monthly.

Be that as it may, each quarter the ABS contacts a sample of homeowners, tenants and businesses supplying consumer products or services in Australia's capital cities. Based on these quarterly sample surveys, the ABS estimates current average prices for a defined 'market basket' of different items purchased by the typical urban family in each capital city.

Exhibit 13.1 presents a breakdown of the significant consumption categories and the relative importance of each as a percentage of total expenditure. The table reveals, for example, that in the 'weighting base period' of 2017–18, about 22.7 cents out of each consumer dollar was spent on housing and about 10.3 cents on transportation.

The composition of the market basket generally remains unchanged from one period to the next, so the CPI is called a fixed-weight price index. If 22.7 per cent of consumer spending was on housing in 2017–18 (the 'weighting base period' for the 17th Series of the CPI), the fixed-weight assumption amounts to assuming that 22.7 per cent of spending is still spent on housing in, say, 2020.

Of course, the higher the time between the weighting base year and the current period for which the CPI is being compiled, the more likely it is that the current period spending weights may have changed significantly. It is also of some considerable interest to note how our spending patterns have changed over the last 60 to 70 years since the Second World War. For instance, the ABS historical weights tell us that in 1948–49 the average urban family spent about 31 per cent of their income on food and 12 per cent on housing, whereas in 2017–18 these percentages had changed very significantly to 16.1 per cent and 22.7 per cent respectively.

The weighting base year is changed roughly every six to seven years, although the 2009–10 expenditure weights were being used until the September quarter, 2017, and revised only in December quarter 2017. Over time, as well as the broad category weights changing considerably, the composition of items included in the CPI has also changed markedly. The ABS looks at data from the Household Expenditure Survey and additional

Exhibit 13.1 **Composition of the Australian consumer price index (weighted average of eight capital cities): March quarter, 2020**

Category (all items 100.0%)	Percentage weight in index	% price change Dec qtr. 2019 ‡ Mar qtr. 2020	% price change Mar qtr. 2019 ‡ Mar qtr. 2020
Food and non-alcoholic beverages	16.1	1.9	3.2
Alcohol and tobacco	7.1	1.6	7.9
Housing	22.7	0.3	0.6
Housing furnishings, supplies and services	9.4	0.8	2.2
Clothing and footwear	3.6	−0.7	2.0
Transportation	10.3	−1.9	2.6
Health	5.4	1.7	2.9
Recreation	12.7	−1.7	1.3
Education	4.3	2.6	2.7
Communication	2.7	−0.3	−3.4
Financial and insurance services	5.8	0.7	1.6
All groups	**100**	**0.3**	**2.2**

Note: In December 2017, the ABS was using its 17th CPI Series, which employed the estimated spending patterns for 2017–18.

Source: Australian Bureau of Statistics, Cat. No. 6401, http://www.abs.gov.au/ausstats.

industry-related data when computing the relative weights. In December 2017, the ABS switched to the 17th Series which was appropriately linked to the earlier CPI series to give one continuous CPI series back to 1948.

Exhibit 13.1 shows that in the March quarter 2020, inflation, as measured by the change in the CPI from March 2019 to March 2020, registered 1.6 per cent. This was a very low rate of annual inflation and represented a continuation of low recorded inflation in evidence in Australia since 2011.

How the CPI is computed

Exhibit 13.2 illustrates the basic idea behind the CPI and how this price index measures inflation. Suppose that in 2018 a hypothetical family in Australia purchased a market basket of only hamburgers, petrol and jeans. Column 1 shows the quantity purchased for each of these items, and column 2 lists the corresponding average selling price. Multiplying the price by the quantity gives the market basket cost in column 3 of each consumer product purchased in 2018. The total cost paid by our typical family for the market basket, based on 2018 prices and quantities purchased, is $245.

Two years later, in 2020, we wish to know the impact of rising prices on consumer purchases. To calculate the CPI, we determine the cost of the same market basket, but valued at 2020 current-year prices, and compare this to the cost at 2012 **base-year** prices. Expressed as a general formula:

Base year
A year chosen as a reference point for comparison with some earlier or later year.

$$\text{CPI} = \frac{\text{current year cost (2020 prices)}}{\text{cost of same market basket at base year (2018 prices)}} \times 100$$

As shown in **Exhibit 13.2**, the 2020 cost for our market basket example is calculated by multiplying the 2020 price for each item in column 4 by the 2018 quantity purchased in column 1. Column 5 lists the result for each item in the market basket – the total market basket cost in 2020 is $335.

Exhibit 13.2 Consumer price index for a simple economy

Products in consumers' market basket	(1) 2018 quantity purchased	(2) 2018 price	(3) Market basket cost in 2018 (1) × (2)	(4) 2020 price	(5) Market basket cost in 2020 (1) × (4)
Hamburgers	50	$0.80	$40.00	$1.00	$50.00
Litres of petrol	250	$0.70	$175.00	$0.90	$225.00
Jeans	2	$15.00	$30.00	$30.00	$60.00
		Total 2012 cost = $245.00		Total 2020 cost = $335.00	

$$2020\ CPI = \frac{\$335}{\$245} \times 100 = 136.7$$

Note: All figures are fictitious and may be very different from actual prices in 2012 and 2020.

The CPI value of 136.7 is computed as the ratio of the current 2020 cost of the 2018 market basket of quantities ($335) to the cost of the same market basket in the 2018 base year ($245) multiplied by 100.

The value of the CPI is always 100 in the base year because the numerator and the denominator of the CPI formula are the same in the base year. At the time of writing, the base year for Australia's CPI, referred to by the ABS as the *index reference period*, was 2018 (so that the value of the CPI averages to 100 during the four quarters of that year). However, as indicated above, the ABS was using 2017–18 as its 'weighting base period'. This means that the spending patterns in 2017–18 were used to weight price changes in the component groups of the CPI (see **Exhibit 13.1**) to get the value of the overall index for each period.

Once the ABS selects the base year (i.e., the index reference year) and uses the market basket technique to generate the CPI numbers, the annual inflation rate is simply computed as the percentage change in the official index from one year to the next. (Note: It is calculated as the *percentage change* in the index numbers, not the simple difference in the index numbers.)

Interpretation of a CPI number

You're the economist

The CPI for 2018–19 was quoted as 116.2. How would you interpret this index number?

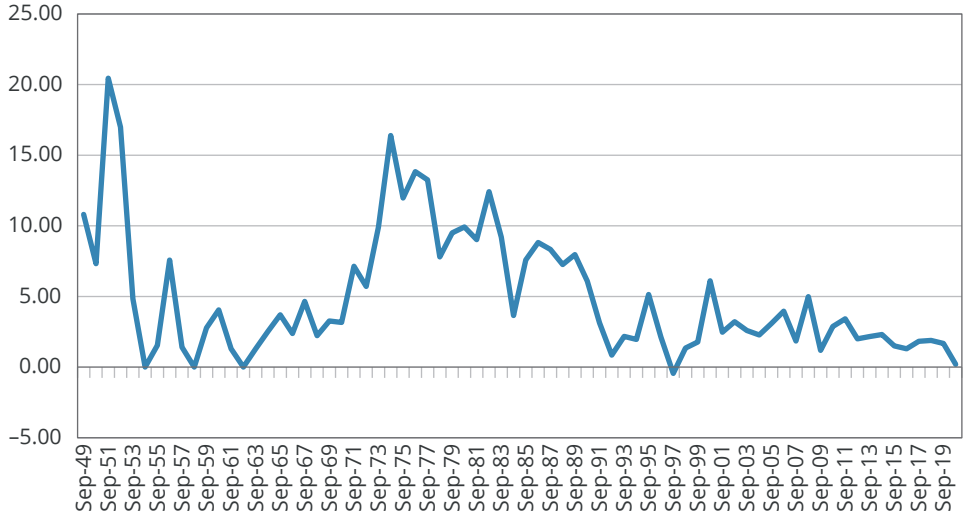
Hint: Remember, in the text, it is stated that the ABS was still using 2011–12 as its index base period for the calculation of the CPI in 2018–19.

In 2017–18, for example, the average CPI for the year was 114.1, while in 2018–19 it was 116.2. The rate of inflation for 2018-19 (over 2017–18) was therefore 1.84 per cent, obtained as ([116.2/114.1] − 1) × 100.

History of Australia's inflation rates

Exhibit 13.3 records how rapidly prices have changed in Australia since 1949, as measured by quarterly 'annual percentage changes' in the CPI (calculated as the change in the CPI value in a quarter of one year from the value of the CPI in the same quarter of the previous year). Considering the historical data of underlying inflation (abstracting from the temporary four-quarter impact on measured inflation from the introduction of the GST in July 2000), the inflation is lower since the onset of the 1990s than at any time since the 1960s. Also of interest

Exhibit 13.3 **Australia's inflation rate, 1949–2019**



Source: Australian Bureau of Statistics, Cat. No. 6401, Tables 1 and 2, http://www.abs.gov.au.

are the very high rates of inflation experienced in the 1970s. These were mainly the result of very high wage growth, which occurred in the early to mid-1970s and were well over increases in labour productivity, thereby significantly increasing production costs. Australia was experiencing deflation in mid-2020 due to the impacts of COVID-19, and, therefore, can observe a significant drop in the price levels and inflation from 2019 to 2020.

Criticism of the consumer price index

The CPI is not a perfect measure of inflation, and it has been the subject of much public debate over the years. Here we discuss the key reasons.

The use of one 'typical' or 'average' basket

Changes in the CPI are based on a 'typical' basket of products purchased, but it is often argued that it does not match the actual market basket purchased by many consumers. Suppose you spend your nominal annual income entirely on lemonade, meat pies and jeans. During the year, the measured CPI inflation rate is 5 per cent, but assume that the prices of lemonade, meat pies and jeans all went up by 20 per cent. In this case, your cost of living has gone up by 20 per cent – not by 5 per cent – and the official inflation rate based on the CPI will understate the impact of price increases on your cost of living.

Retired people, for example, buy a bundle of products that differs quite markedly from that of the 'typical' family. Because retired people purchase proportionally more medical services than the typical family – and fewer children-related items – the CPI-based inflation rate may understate the impact of inflation on older people if the cost of medical services is rising faster than other items in the CPI.

While this is a valid criticism, it is a complicated issue to remedy, short of the ABS trying to construct a separate CPI for all the various special-interest groups in the community, and the cost of doing so would be so prohibitive as to be impractical.

Adjusting for changes in quality

The ABS has difficulty adjusting the CPI for changes in *quality*. As noted in the chapter on measuring GDP, how do you compare a personal computer made in the past with a new personal computer? The new computer may cost more, but it is much better than the old computer. A portion of the price increase, therefore, reflects better quality rather than only a higher price for the same item.

If the quality of items improves along with price increases in the items, increases in the CPI will overstate inflation. Similarly, deteriorating quality means that increases in the CPI understate inflation. The ABS attempts to make adjustments for quality changes in automobiles, electronic equipment and other products in the market basket, but, in most cases, precisely quantifying the size of the necessary adjustments is a challenging exercise. Thus, whatever adjustments are made are necessarily inaccurate and probably inadequate.

The impact of the substitution effect

The use of a single unchanging 'weighting base period' market basket ignores the law of demand. If the relative price of a product rises, consumers will tend to purchase available close substitutes, and so a smaller quantity is demanded. Because the market basket does not automatically change over time, the result is that by not reducing the percentage weight applying to the items whose relative price is increasing. The CPI will overstate the impact of higher prices of such products on the measured change in the overall price level.

To deal with this substitution bias problem, the ABS regularly (about every six years) updates the year it uses as its 'weighting base period' and uses household expenditure survey data for that year to try to keep up with changing consumption patterns and to correct for the fixed market-basket limitations of the CPI. The 'weighting base period' was last updated in 2017–18 using 2015–16 expenditure patterns to determine the weights for its 17th CPI Series.

The university education price index

You're the economist

Suppose Anne's market basket for a university education consisted of only the four items (fictitious costs) listed in the following table, with quantities of each item fixed in all periods and indicated in the notes below the table.

Using 2017 as the base year, what is the percentage change in Anne's university education price index in 2020 compared to 2019?

Item	2017	2019	2020
Tuition and fees[1]	$8 000	$8 500	$8 250
Room and board[2]	$7 000	$7 400	$7 800
Books[3]	$2 000	$2 500	$2 800
Soft drinks[4]	$250	$275	$285

[1] Tuition for two semesters
[2] Payment for nine months
[3] Twenty books of 800 pages with full colour
[4] One hundred 350-ml Coca-Colas

In addition to the limitations mentioned above, CPI may not reflect the actual cost of living since the ABS is surveying major stores to get price information. Consumers are purchasing goods and services from cheap online stores and discount stores where prices are a lot cheaper. Also, given the goods and services in the CPI basket are fixed for several years, newly introduced popular consumer products may not be included in the basket. The above two limitations may lead to overestimation of the actual rate of inflation.

Demand-pull and cost-push inflation

Economists often conceptually distinguish between two basic types of inflation, depending on whether it originates from the buyers' or the sellers' side of the market.

Demand-pull inflation

Demand-pull inflation
A rise in the general price level resulting from an excess of total spending (demand) over supply.

Perhaps the most familiar type of inflation is called **demand-pull inflation**. This is a rise in the general price level resulting from an excess of total spending (demand). Demand-pull inflation is often expressed as 'too much money chasing too few goods', and sellers typically respond by raising prices. If such excess in nominal demand is widespread across many markets, the general price level in the economy will be 'pulled up' by the pressure from buyers' total expenditures.

Since aggregate spending includes consumer spending (C), business investment (I), government spending (G) and net exports ($X - M$), even foreigners may contribute to domestic inflation by bidding up the price of a country's exports. The resulting increase in export income will feed through into additional demand in other parts of the economy, thereby possibly resulting in increases in domestic prices.

This export income effect flowing from the impact of the commodity price boom was almost certainly a significant cause of the rising inflation in Australia in the second half of 2007 and into early 2008 (subsequently considerably ameliorated by the impact of the GFEC in Australia).

Cost-push inflation

Cost-push inflation
A rise in the general price level resulting from an increase in the cost of production, irrespective of demand conditions.

An excess of total spending is not the only possible explanation for rising prices. In 1973, and again in 1979, for example, the price of oil increased dramatically. This meant a significant increase in the cost of producing many goods and services (not just in the price of petrol). The result was **cost-push inflation**, which is a rise in the general price level resulting from a general increase in the costs of production.

More generally, any increased costs to businesses are a potential source of cost-push inflation. This means that upward pressure on prices may be caused by cost increases for labour, raw materials, construction, equipment, borrowing and so on.

The role of expectations

Note that the influence of expectations on both demand-pull and cost-push inflation is an important consideration.

Suppose buyers see prices rise and believe they should purchase that new television set or car today before these items cost much more tomorrow. At or near full employment, this demand-pull pressure results in a rise in prices. On the suppliers' side, firms may expect their production costs to rise in the future, and this causes them to raise prices in anticipation of the higher costs. The result is cost-push inflation.

So, in reality, the two types of inflation can be closely interlinked and, in any given inflationary episode, it may be impossible to ascribe the cause to one or the other type of mechanism. The vital point of distinguishing between the two in this discussion is to provide some insights into the different types of factors that may result in increasing prices.

In summary

- **Inflation/deflation** is an increase/decrease in the overall average level of prices.
- **Disinflation** is a reduction in the rate of inflation.
- The **consumer price index (CPI)** measures changes in the average price of consumer goods and services.
- Some criticisms of the CPI include its use of a single 'typical' basket of products, the impact of quality changes, and the impact of the 'substitution effect'.
- Rising prices can be the result of demand-side pressures or cost-push factors or a combination of both.

2 Consequences of inflation

We now turn from the issue of measuring inflation to its effects on people's income, **wealth** and purchasing power. As we will see, inflation can quite arbitrarily create benefits for some and losses for others. It can also potentially impede proper business decision-making and the efficiency with which resources are allocated in the economy.

Wealth
The value of the stock of assets owned at some point in time.

Inflation shrinks income

Inflation can reduce our standard of living through declines in the purchasing power of our income. The higher the rate of inflation, the more significant the decline in the quantity of products we can purchase with a given **nominal (or money) income**. Nominal income is the actual number of dollars received in a period of time from sources such as wages, salary, rent, dividends, interest or pensions.

However, nominal income does not measure real purchasing power. To evaluate the real worth of a given increase in nominal income requires converting nominal income to **real income**. Real income is the actual number of dollars received (nominal income) adjusted for changes in the general level of prices as measured by, for example, the CPI.

Nominal (or money) income
The actual number of dollars received as income during a period of time.

Real income
The actual number of dollars received (nominal income) adjusted for changes in the average price level as measured by, for example, the CPI.

Real income measures the quantity of goods and services that can be purchased with one's nominal income. If the CPI increases and a person's nominal income remains the same, their real income (purchasing power) falls.

Suppose nominal income in 2017 is $40 000 and the 2017 CPI value is 136. Real income relative to the CPI base year is:

$$\text{real income} = \frac{\text{nominal income}}{\text{CPI (as decimal, i.e. CPI / 100)}}$$

$$\text{2017 real income} = \frac{\$40\,000}{1.36} = \$29\,412$$

Now assume nominal income rises in 2018 by 10 per cent, from $40 000 to $44 000, and the CPI increases by 5.1 per cent, from 136 to 143 (143/136 = 1.051). Thus, more money is earned, but how much is it really worth?

To answer this question, we compute the 2018 real income as follows:

$$\text{2020 real income} = \frac{\$44\,000}{1.43} = \$30\,769$$

Using the difference between the preceding two computed real income figures, the percentage change in real income between 2017 and 2018 was 4.6 per cent ($1357/$29412 × 100). This means the increase in nominal income from 2017 to 2018 translates into an extra $1357 of real purchasing power in 2018 to spend on movies, clothes, travel and so on.

A good approximation for calculating changes in real income can be obtained through the following simple formula:

$$\text{percentage change in real income} = \text{percentage change in nominal income} - \text{percentage change in CPI}$$

Economics and ethics Importance of inflation-adjusted analyses

In the previous section, we concentrated on using CPI movements to inflation-adjust nominal income changes in order to gauge changes in real income. Movements in the CPI are also very useful when one is interested in comparing particular prices or expenditure over time. For instance, in the 'Analyse the issue' box that follows, we look at the inflation-adjusted AUD price of oil over time. Also, when we look at announced increases in levels of annual government spending in areas like, for example, health and education, it is crucial to look at the inflation-adjusted increases to judge the extra real spending increases appropriately involved.

If nominal spending on a government program increased by 5 per cent when inflation was running at 8 per cent, then the government committed less in real resources to that program. In regard to the economics and ethics considerations mentioned in **Chapter 1**, unfortunately too often government ministers will announce program expenditure increases in nominal terms rather than in real, inflation-adjusted terms. Without adjustment, such announcements can give quite a misleading impression of the implied funding level commitments.

Inflation and the real interest rate

Nominal interest rate
The actual rate of interest earned or paid over a period of time.

Real interest rate
The nominal rate of interest minus the inflation rate.

It is also essential to distinguish between the **nominal interest rate** and the **real interest rate**. The nominal interest rate is the actual rate of interest earned over a period of time, and banks announce this. If you borrow $10 000 from a bank at a 10 per cent annual interest rate for five years, this is more accurately called a 10 per cent annual nominal interest rate.

The real interest rate is approximately the nominal interest rate minus the inflation rate. Suppose the inflation rate during the year is 3 per cent. This means that a fixed-term deposit for $5000 you may have at a bank that yields 8 per cent annual nominal interest-only earn approximately 5 per cent real interest. Even though you receive $5400 at the end of the year, the general level of prices of goods and services has gone up by 3 per cent, which has reduced the real purchasing power of $5400.

Thus, the effective real purchasing power of your $5400 at the end of the year (in terms of the prices prevailing at that time) will be 5400/1.03, or $5243, representing growth in real purchasing power of 4.86 per cent (approximately 5 per cent).

$$\text{Real Interest Rate} = \text{Nominal Interest Rate} - \text{Inflation Rate}$$

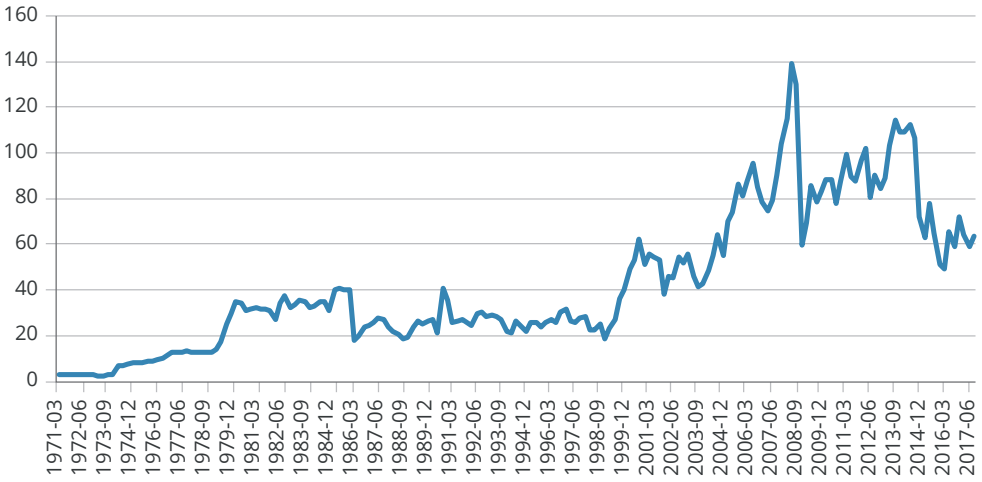
Analyse the issue Applicable concept: inflation-adjusting nominal price data

Just how high really was the price of oil in 2017?

As can be seen from **Exhibit 13.4**, after a long period of a relatively stable crude oil price from the mid-1980s, the AUD price of oil began shooting up in the early 2000s, staying very high through until mid-2008, after which it fell dramatically with the full impact of the GFEC beginning to be felt. In late 2017, the price was still well below its 2008 peak of A$139 (monthly average terms).

Apart from the impact of changes in the price of oil being felt by all motorists at the petrol pump, the oil price is a significant factor in production costs and is, therefore, a potentially very significant supply-side shift factor (this is discussed again in **Chapter 14**). The price of oil is widely accepted as the cause of the first 'stagflation' episode of 1973–74, where a synchronised international recession occurred along with high inflation, and is also considered a significant cause of the 1979–80 international recession.

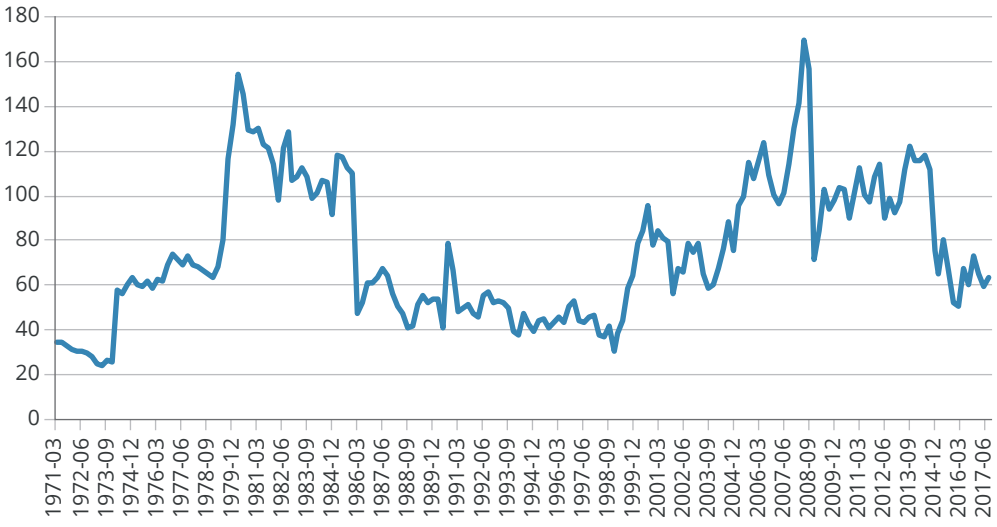
Exhibit 13.4 Crude oil price, $A, 1971–2017



Source: Oil price data supplied by Economagic, http://www.economagic.com.

Putting aside the recession concern issue, from a historical perspective, just how high were oil prices in the 2000–17 period? In the first graph (**Exhibit 13.4**), the quarterly nominal AUD price of oil per barrel is charted for the years 1971–2017 (to the September quarter). As can be seen, the nominal AUD price of oil averaged around A$63 per barrel in the first nine months of 2017 and was considerably higher than the average price of oil in the oil price crises of 1973–74 and 1979–80. From this graph, it would seem that the price of oil in the 2000s was considerably higher than in the 1970s, 1980s and throughout the 1990s.

Exhibit 13.5 Crude oil price, inflation adjusted, 1971–2017



Source: Oil price data supplied by Economagic, http://www.economagic.com.

However, it is important to remember that the general level of prices in the Australian economy rose steadily throughout the 1970s, 1980s and 1990s. As measured by the CPI, the average price level in 2017 was more than four times as high as it was around the time of the second oil price shock in 1979–80. Therefore, to gauge the relative 'real price' of oil over this long period of time, we have to adjust the nominal oil price for general inflation. This is done in the next graph (**Exhibit 13.5**).

As can be seen from **Exhibit 13.5**, a very different picture emerges once the oil price data are inflation-adjusted. Here, all prices are adjusted into September 2017 terms; that is, September 2017 was used as the base period.

Now the real price of oil in 1979–80 becomes very clear.

What do you think?

Using the information provided in **Exhibit 13.5**, answer the following questions:

1. What was the approximate price of oil in 1979–80 equivalent to in 2017 terms? This is the price that oil would have needed to be in 2017 for it to be considered as high in real terms as it was back in 1979–80.
2. Compare this to the AUD price of crude oil when you are reading this chapter. How high was the real price of oil at that time? To answer this question you will need to find out the USD price of oil (West Texas Intermediate, or WTI), the value of the AUD in USD terms, and the value of the CPI at the time.

This should demonstrate just how important it is to adjust for general price-level changes when you want to compare the real price of something over different time periods.

Inflation can turn lenders into losers

Suppose someone decides to extend a one-year loan to a close friend in order to start a business. Both parties anticipate that the inflation rate will be 5 per cent during the year. The loan is made, and the agreement is for the borrower simply to repay the principal plus 5 per cent to offset inflation. In short, both parties assume payment of a zero real interest rate (the 5 per cent nominal interest rate minus the 5 per cent expected rate of inflation).

Consider what happens if the inflation rate is unexpectedly actually 8 per cent during the year of the loan. The clear unintentional winner is the borrower because the lender is only paid the principal plus 5 per cent interest, resulting in the purchasing power of the money they lent falling by 3 per cent owing to the actual inflation rate turning out to be 8 per cent.

Instead of zero, the real interest rate paid by the borrower was −3 per cent (the 5% nominal interest rate minus the 8% rate of inflation). In real terms, the friend who extended the loan *paid for the privilege* of lending their money to their friend!

Interestingly, during the 1970s, in several countries around the world – including Australia – real rates of interest were sometimes negative. This was because lenders had not fully adjusted their inflationary expectations to the higher rates of inflation prevalent in the 1970s compared with the low inflation rates of the 1960s. Many therefore lent money at nominal rates of interest, which were insufficient to cover the eventual rate of actual inflation that had occurred by the time they received their principal and interest back. The lenders were the losers, and the borrowers were the winners!

Inflation and investment and business decisions

Business and investment decisions should be based on assessments of relative real rates of return to alternative investments. When inflation is low, it is also usually stable, thereby making such judgements of relative rates of return more straightforward.

However, when the rate of inflation is higher, it is also usually more unstable. This makes it more difficult to abstract correctly from future expected general price-level changes to focus on relative real rates of return

from alternative investments. This can lead to greater numbers of inappropriate decisions being made by firms and individuals, resulting in reduced efficiency with which the community's scarce resources are allocated.

Inflation in other countries

Exhibit 13.6 reveals that inflation rates can vary widely among nations. In 2017, for example, Malaysia, India, Russia and Indonesia experienced relatively high inflation rates while most other countries listed had quite modest inflation rates. While Russia's inflation rate was high in 2017, it had, fortunately, come down quite significantly from the 7.0 per cent recorded in 2016. Five countries listed – Japan, France, Greece, China and Germany – had very low rates of inflation (less than 1.7 per cent) and two of them – Greece and Japan – actually experienced deflation in 2016. None of the countries listed experienced deflation in 2017.

Exhibit 13.6 **Annual inflation rates in selected countries, 2017 and 2016**

Country	2017	2016
Australia	2.0	1.3
Brazil	3.4	8.7
China	1.6	2.0
France	1.2	0.3
Germany	1.7	0.4
Greece	1.1	–0.8
India	3.6	4.5
Indonesia	3.8	3.5
Japan	0.5	–0.1
Malaysia	3.8	2.1
New Zealand	1.9	0.6
Russia	3.7	7.0
South Korea	1.9	1.0
United Kingdom	2.7	0.7
United States	2.1	1.3

Source: Central Intelligence Agency, *World Factbook*, https://www.cia.gov

Inflation on a rampage

Hyperinflation is an extremely rapid rise in the general price level. Most economists would agree that an inflation rate of about 50 per cent per month or more is hyperinflation.[1] Runaway inflation often leads to rapid and violent social and political change stemming from four causes.

Hyperinflation An extremely rapid rise in the general price level, generally defined as prices rising by 50 per cent or more each month.

First, individuals and businesses develop an inflation psychosis, which causes them to buy quickly today in order to avoid paying even more tomorrow. Second, substantial unanticipated inflation jeopardises debtor–lender contracts, such as credit cards, home mortgages, life-insurance policies, pensions, bonds and other forms of savings.

[1] P. Cagan (1956) first provided this definition in his *The Monetary Dynamics of Hyperinflation*, a work often regarded as the first serious study of hyperinflation.

Wage–price spiral
A situation that occurs when increases in nominal wage rates are passed on in higher prices, which in turn result in even higher nominal wage rates and prices.

Third, hyperinflation sets a vicious **wage–price spiral** in motion. A wage–price spiral can undoubtedly occur in quite normal economic circumstances, but it is deadly in times of hyperinflation. It occurs as a result of increases in nominal wage rates being very rapidly passed on in higher prices, which in turn result in even higher nominal wage rates and prices, and so on. Fourth, in order to hedge against the high losses of purchasing power from hyperinflation, people invest in non-productive things like gold, silver, stamps, jewels, art, antiques and other currencies, rather than new factories, machinery and technological research, which expand an economy's production possibilities frontier.

History reveals numerous hyperinflation examples. One of the most famous occurred during 2007–08 in Zimbabwe. It was challenging to measure the hyperinflation in Zimbabwe since the government stopped collecting inflation statistics. The Zimbabwe government simply printed money to pay its bills. By 2008, the month-to-month inflation rate in Zimbabwe had reached about 79 billion per cent per month, with prices rising many times daily. No one was willing to make new loans, credit markets collapsed, wealth was redistributed, and people's lifetime savings were wiped out. The Zimbabwe dollars became worthless, and they stopped printing its currency in 2009 with currencies from other countries being used.

Global perspective — Applicable concept: hyperinflation

When the inflation rate is 116 000 per cent, prices change by the hour

A 1985 *Wall Street Journal* article describes hyperinflation in La Paz, Bolivia:

> A courier stumbles into Banco Boliviano Americano with a large sack containing 32 million pesos. At that moment, the 32 million pesos – enough bills to stuff a mail sack – were worth only $500. Life's like that with quadruple-digit inflation ...
>
> In 1984, prices zoomed 2700 per cent, compared with a mere 329 per cent the year before! Experts predict the inflation rate could soar as high as 40 000 per cent.
>
> The 1000-peso bill, the most commonly used, costs more to print than it purchases. It buys one bag of tea. To purchase an average-size television set with 1000-peso bills, customers have to haul money weighing more than 68 pounds into the showroom. To ease the strain, the government came out with a new 100 000-peso note, worth $1.[1]


Getty Images Plus/iStock Editorial/jacus

A 1993 *Charlotte Observer* article reported a rate of inflation in the billions for Belgrade:

> The number on Wednesday was 286 125 293 792. It was not the day's winning lottery figures nor the number of miles to the Hubble space telescope. It was the latest calculation of Yugoslavia's nearly incalculable inflation rate ...To cover the costs of war and pay off the unemployed, the government has resorted to indiscriminately printing money. That has rendered the national currency, the dinar, practically worthless.[2]

A 1998 Dow Jones Newswire story reported:

> The Russian government is likely to try to solve the country's economic impasse by printing money and thus cause hyperinflation. If this happens, Russia's inflation rate will reach 450 per cent to 500 per cent this year, Gaidar, former Prime Minister, said. Even if the government rejects such measures, consumer prices will rise by 250 per cent to 300 per cent this year.[3]

What do you think?

1 Can you relate inflation psychosis to these excerpts? Give an example of a debtor–lender relationship that is jeopardised by hyperinflation.
2 Why do you think that, in 1984/85, workers in Bolivia were striking even though wages rose at an annual rate of 1500 per cent? Do you see any connection between hyperinflation and the political system?

[1] S. L. Nazario, 'When inflation rate is 116 000%, prices change by the hour', 7 February 1985, p. 1. *The Wall Street Journal* by Dow Jones & Co; News Corporation. Reproduced with permission of DOW JONES COMPANY in the format Book via Copyright Clearance Center.

[2] S. Lekic, 'Belgrade puts rate of inflation in billions', *Charlotte Observer*, 2 December 1993, p. 24A.
[3] P. Munter, 'Russia's Gaidar: Government to print money, trigger hyperinflation', *Dow Jones Newswire*, 2 October 1998.

In summary

- **Percentage change in real income** = approximately the percentage rise in nominal income − the rate of inflation.
- **Real rate of interest** on a loan = approximately the nominal interest rate charged − the rate of inflation.
- Higher and more unstable inflation rates often result in reduced efficiency with which the community's scarce resources are allocated.
- Problems are amplified manifold during periods of **hyperinflation** (inflation of more than 50 per cent each month).

3 The meaning and measurement of unemployment

AAP Image/Tracey Nearmy

Since the abyss of the Great Depression, a significant economic goal of Australia and most other developed countries around the world has been to achieve full employment. However, what actually is full employment?

The methods used in Australia to measure employment and unemployment are virtually identical to those used in most other developed nations. Each month the ABS surveys a random sample of households. Each member of the family who is 15 years of age or older (and, as discussed shortly, who satisfies a few other conditions to be regarded as a member of the labour force) is asked whether he or she is employed or unemployed. If a person works at least one hour per week for pay or at least 15 hours per week as an unpaid worker in a family business, that person is counted as employed.

Unemployment rate
The percentage of people in the labour force who are without jobs and are actively seeking jobs.

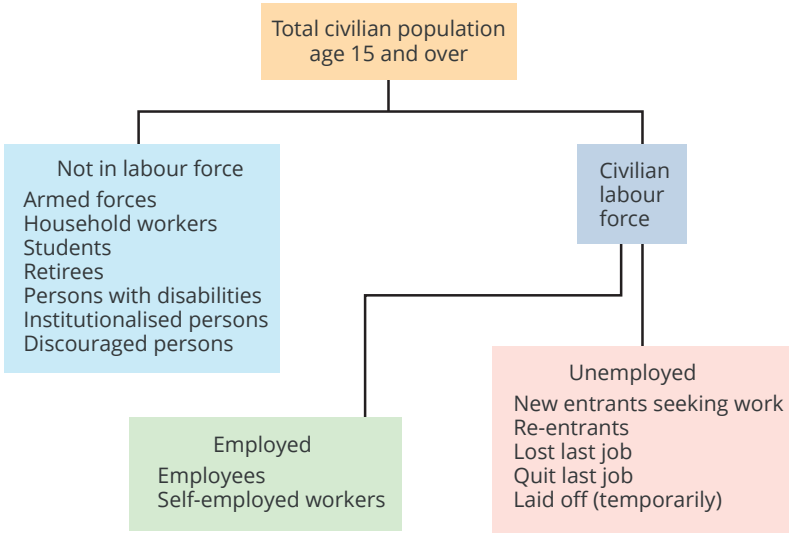
Civilian labour force
Those people 15 years of age and older who are employed or who are actively seeking a job, excluding those in the armed forces, homemakers, students, discouraged workers and other persons not in the labour force.

If the person is not employed, the question is then whether or not he or she has looked for work in the past month. If they have, the person is said to be unemployed. Using its survey data, the ABS publishes the **unemployment rate** and other employment-related statistics monthly.

The unemployment rate is the percentage of people in the labour force who are without jobs and are actively seeking jobs. Nevertheless, who is actually counted as an unemployed person and which people are counted as being in the labour force? Indeed, all people without jobs do not rank among the unemployed. Students (older than 15) and retired persons are not counted as unemployed. Likewise, people who are chronically ill or severely disabled are not included as unemployed.

The **civilian labour force** – which is usually what is reported in the press and referred to in this text as merely the 'labour force' – are those people (excluding students) aged 15 and over who are either employed or unemployed, excluding members of the armed forces, specific other categories of people (refer to **Exhibit 13.7**), as well as people in institutions, such as prisons and psychiatric hospitals.

Exhibit 13.7 **Population, employment and unemployment**



Based on survey data, the ABS computes the civilian unemployment rate – referred to in this text as merely the 'unemployment rate' – as the ratio of the number of unemployed to the labour force. In 2020, until June, the unemployment rate in Australia averaged 7.4 per cent over the year.

The unemployment rate is widely reported and analysed and is always a significant political issue. Indeed, governments can fall as a result of increased unemployment rates. **Exhibit 13.8** charts the historical record of the Australian unemployment rate since 1978.

Note that the highest unemployment rate reached in the 20th century was actually over 25 per cent during the years of the Great Depression (not charted). Since then, the highest rate of unemployment was experienced relatively recently in 1993. At the other extreme, the lowest unemployment rate we have attained was below 2 per cent in some years during the 1960s. However, we may experience an even higher rate of unemployment in late 2020 due to the impact of COVID-19.

Probably the starkest feature of **Exhibit 13.8** is the very dramatic effect that the early 1980s and early 1990s recessions had on Australia's unemployment rate – and, to a much lesser extent, the impact of the GFEC on Australia's unemployment rate in 2009. This is typical of the impact of recessions. Also, notice the very long time it typically takes to get the unemployment rate back down after a recession.

Exhibit 13.8 The Australian unemployment rate, 1978–2020

Based on Australian Bureau of Statistics data, Cat. No. 6202.0, https://www.abs.gov.au.

Unemployment in other countries

Exhibit 13.9 shows that, among the listed countries, Japan and Malaysia had the lowest unemployment rates in 2017. China, Germany and South Korea also had quite low measured unemployment rates. However, in the case of China, the reported unemployment rate is based on urban areas only, and many believe China's actual unemployment rate (when estimated using a similar methodology as used in developed countries) could be as high as 8–9 per cent or even more. At 5.6 per cent, Australia had a reasonably low unemployment rate compared with other developed countries in both 2016 and 2017.

The most striking feature of **Exhibit 13.9** is Greece's unemployment rate. Since Greece entered the Eurozone in 2001, it has not fared very well. By 2008–09, as the GFEC impacted the world's economies, Greece had already developed severe sovereign debt problems brought about primarily due to the Greek government's foreign borrowing to meet growing welfare and public sector wages bills, along with very high military spending. The onset of the GFEC brought the country to a crisis point. More will be said about this in **Chapter 19**. However, in seeking EU assistance to avoid defaulting on its sovereign debt, the country had to agree to a range of economic austerity measures, which have severely impacted on the short-term economic growth of the country, with the result of escalating unemployment, borne disproportionately by the young.

Criticism of the unemployment rate

The unemployment rate is criticised for both understating and overstating the 'true' unemployment rate.

An example of *overstating* the unemployment rate occurs when respondents to the ABS falsely report that they are seeking employment when they are not. The motivation may be that unemployment benefits depend on actively pursuing a job, or an individual may be 'employed' in illegal activities. Besides, if a person is engaged in underground or illegal economic activities for income generation, they will not reveal accurate information. This makes people provide false information that may be considered as unemployed.

Exhibit 13.9 Unemployment rates for selected nations, 2016 and 2017

Country	2016	2017
Australia	5.7	5.6
Brazil	11.3	12.8
China	4.0	3.9
France	10.0	9.4
Germany	4.2	3.8
Greece	23.6	21.5
India	8.0	8.5
Indonesia	5.6	5.4
Japan	3.1	2.9
Malaysia	3.5	3.4
New Zealand	5.1	4.7
Russia	5.5	5.2
South Korea	3.7	3.7
United Kingdom	4.9	4.4
United States	4.9	4.4

Source: Central Intelligence Agency, *World Factbook*, http://www.cia.gov.

Discouraged worker
A person who wants to work, but who has given up searching for work because he or she believes there will be no job offers.

The other side of the coin is that the official definition of unemployment *understates* the unemployment rate by not counting so-called **discouraged workers**. A discouraged worker is a person who would like to work, but who has given up searching for a job because he or she believes there will be no offers.

The ABS counts a discouraged worker as anyone who has looked for work within the last six months but responds that he or she is no longer actively looking. The ABS simply includes discouraged workers in the 'not in the labour force' category listed in **Exhibit 13.7**. Because the number of discouraged workers is likely to rise during a recession, the degree of underestimation of the official unemployment rate is expected to increase during a downturn.

What is the approximate size of the discouraged worker effect in Australia? The ABS does not appear to publish such estimates. However, a recent US study suggests that, in that country, the impact of the US recession of 2008–09 was that appropriately incorporating the discouraged worker effect would have raised the measured unemployment rate by up to 0.35 per cent.[2]

Underemployment
Another reason why some argue the measured unemployment rate understates the degree of labour unemployment in the economy. It is defined as those employed who would nonetheless like to work more hours if they could, expressed as a percentage of the labour force.

Underemployment is another reason why some argue the measured unemployment rate understates the degree of labour unemployment in the economy. It is defined as those employed who would nonetheless like to work more hours if they could, expressed as a percentage of the labour force. This could also be viewed as people with high skills engaging in low-skilled work, or people who are not engaging in work that suits their qualifications. Such under-utilisation of work potential is again higher during a recession but is again not reflected in the measured unemployment rate.

[2] B. Ravikumar and L. Shao, 'Accounting for discouraged workers in the unemployment rate', *Economic Synopses*, No. 9, Federal Reserve Bank of St Louis, 2014, http://research.stlouisfed.org/econ/ravikumar, accessed 25 May 2018.

The ABS does publish measures of underemployment.[3] For example, in June 2020, the official measured unemployment rate was 7.4 per cent. However, the underemployment rate in that quarter was 11.7 per cent. This implies that in June 2020, what the ABS calls the labour underutilisation rate was 19.1 per cent (7.4 + 11.7).[4]

In summary

- The **civilian labour force** includes those people 15 years of age and older who are employed or who are actively seeking a job, excluding those in the armed forces, homemakers, students, discouraged workers and other persons not in the labour force.
- The **unemployment rate** includes the percentage of people in the labour force who are without jobs and are actively seeking jobs.
- Two criticisms of the measured unemployment rate are:
  - it excludes **discouraged workers**
  - the existence of **underemployment** – those employed who would like more hours of work.

4 Types of unemployment

The overall total number of unemployed persons arises from four different causes: *seasonal*, *frictional*, *structural* and *cyclical*. Understanding these conceptual categories of unemployment aids in understanding and formulating policies to ease the burden of unemployment.

Seasonal unemployment

Seasonal unemployment is unemployment caused by changes in hiring due to recurring changes in weather conditions, demand and/or production patterns. Seasonal variations in employment are inevitable.

Seasonal unemployment Unemployment caused by recurring changes in hiring due to recurring changes in weather conditions, demand and/or production patterns.

Demand for specific jobs rises and falls with seasons of the year. Ski instructors find jobs at Australia's southern ski resorts during the winter, but not in the summer. Construction workers are sometimes seasonally unemployed – or laid off for short periods – during the wet season. When crops are harvested, farms hire many workers, and after the harvest, these people become seasonally unemployed. As a final example, the Christmas season is generally a time of peak employment in the wholesale and retail trades areas.

Frictional unemployment

For some unemployed workers, the absence of a job is only temporary. At any given time, some people with marketable skills lose their jobs, and others voluntarily quit jobs to accept or look for new ones. Moreover, there are always young people who leave school and search for their first job.

Workers in some industries, such as construction, experience short periods of unemployment between projects and temporary lay-offs are common. Because jobs are available requiring their skills such people are considered to be 'between jobs' and will find new employment once their skills are matched to available jobs.

This type of unemployment, called **frictional unemployment**, is usually a quite short term and is therefore not of great concern from either public policy or a social welfare perspective. Frictional unemployment is unemployment caused by the normal search time required by people with marketable skills who are changing jobs, initially entering the labour force, or re-entering the labour force.

Frictional unemployment Unemployment caused by the normal search time required by workers with marketable skills who are changing jobs, initially entering the labour force, or re-entering the labour force.

[3] ABS, Statistics Catalogue No. 6202.0, *Labour Force, Australia*, Table 22.

[4] Typically, underemployment is a phenomenon affecting women more than men. In June 2020, for example, ABS data suggested underemployment among women was 12.6 per cent, and for men, it was 10.8 per cent.

The cause of frictional unemployment is either the transition time to a new job or the lack of information required to match a job applicant immediately with a job vacancy. Because it takes time to search for the information required to match employers and employees, there will always be some number of workers frictionally unemployed. However, sound public policy can certainly reduce frictional unemployment. Improved methods of distributing job information through a national data bank can help unemployed people find jobs more quickly and reduce frictional unemployment. The introduction of the Job Network initiative by the Australian government in the late 1990s may be thought of as an example of trying to increase the efficiency of the job search and placement process.

Structural unemployment

Structural unemployment Unemployment caused by a mismatch of the skills of workers out of work and the skills required for existing job opportunities.

Unlike frictional unemployment, **structural unemployment** is long-term or possibly permanent unemployment resulting from the non-existence of jobs for some unemployed people. Structural unemployment is unemployment caused by a mismatch of the skills of people out of work and the skills required for existing job opportunities.

Changes in the structure of the economy over time, resulting from changing tastes and demand conditions or from changed production processes, can lead to structural unemployment. In a rapidly evolving economy, people may become structurally unemployed due to poor educational levels, and skill sets becoming inappropriate for new production technology or due to the introduction of new products. Foreign competition, geographic immobility, restricted entry into specific jobs, and shifts in government priorities over time can also lead to structural unemployment.

Because of the great dynamism of modern economies, with rapidly changing tastes and production techniques, and the associated numerous sources of mismatching between skills and jobs, it is generally considered that a certain level of structural unemployment is inevitable. Public and private programs that retrain and/or re-educate employees to fill available job openings can decrease structural unemployment.

Another potential cause of structural unemployment is the existence of a minimum wage law. This may happen if the regulated minimum wage is above the productivity of particular groups of unskilled workers. Rather than guarantee a minimum standard of living for the workers, which is the intention of the policy, it may, paradoxically, actually work against them by causing them to be unemployed.

One policy approach intended to offset such undesirable effects of a minimum wage is a sub-minimum wage paid during a training period to give employers an incentive to hire unskilled workers. An alternative is to offer a separate financial incentive to the employer (thereby reducing the employment cost of the marginal employee) to hire people who have been unemployed for a long time. However, another alternative is to remove the minimum wage provisions entirely but ensure a reasonable minimum standard of living for the lowly paid through the tax and welfare system.

Cyclical unemployment

Cyclical unemployment Unemployment caused by the lack of a sufficient number of jobs during a recession.

Cyclical unemployment is attributable to the business cycle and is caused by the lack of a sufficient number of jobs during a recession or a growth slowdown. When real GDP falls, many companies cut back on production, other companies close altogether, and jobs disappear. People scramble for fewer available jobs (exactly as happened in the wake of the GFEC in 2008–09 and the COVID-19 crisis in 2019–20 in many countries around the world).

Look at what has happened to unemployment during Australia's first two recessions in the 1980s and 1990s (refer again to **Exhibit 13.8**). In each case, unemployment rose dramatically and in a short time. Furthermore, after the recession was over, it seemed to take a long time for enough jobs to be generated to get the unemployment rate back down again. For example, in the 1990s recession, it took less than two years for the unemployment rate to rise from under 6 per cent to over 11 per cent, but then it took eight years to get it back down to anything like 6 per cent.

To smooth out these swings in unemployment, a focus of the macroeconomic policy is to moderate cyclical unemployment.

The goal of full employment

In this section, we take a closer look at the meaning of **full employment**. Because seasonal, frictional and structural unemployment are present in good and bad times, *full employment* does not mean 'zero per cent unemployment'.

Full employment The situation in which an economy operates at an unemployment rate equal to the sum of the seasonal, frictional and structural unemployment rates.

Full employment is the situation in which an economy operates at an unemployment rate equal to the sum of the seasonal, frictional and structural unemployment rates. Full employment, therefore, is that rate of unemployment that involves zero cyclical unemployment.

Unfortunately, economists cannot state with certainty what percentage of the labour force is seasonally, frictionally and structurally unemployed at any point in time. In practice, therefore, the full employment rate of unemployment – sometimes referred to as the 'natural rate of unemployment' or the 'non-accelerating inflation rate of unemployment' (NAIRU) – is not only difficult to measure but also quite clearly changes over time.

In Australia in the 1960s, 2 per cent unemployment was generally considered to represent full employment. During the 1970s it seemed to rise to around 4 per cent, in the 1980s and 1990s it was probably around 6 to 7 per cent, and in the 2000s up to 2020 (the time of writing this text) it had probably dropped back down to about 5 per cent.[5]

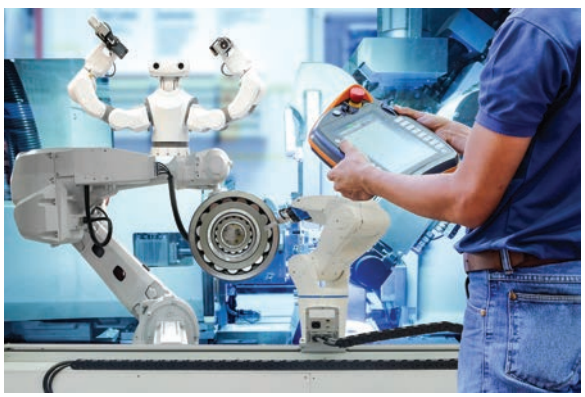
Several reasons are given for why the full-employment rate of unemployment is not fixed. One reason may be that, during the period between the early 1960s and the early 1980s, the participation of women in the labour force increased significantly. This change in the labour force composition is thought to have increased the full-employment rate of unemployment because women typically experience higher unemployment rates than men.

Another frequently cited but not completely uncontroversial explanation for the rise in the full-employment rate of unemployment since the 1960s is that relatively more substantial unemployment benefits, welfare and the interaction of the tax and welfare systems have operated to make unemployment less painful and employment a relatively less attractive option than in the 1950s and 1960s.

Applicable concept: types of unemployment

Analyse the issue

Are machines and robots taking our jobs?



Shutterstock.com/FeelGoodLuck

Consider that, throughout the last century, the manufacturing and services sectors were incrementally impacted by the increased use of machines and automation. This has inexorably introduced more considerable uncertainty as far as employment in the sectors is concerned. Today, industrial robots are wiping out the manufacturing jobs, and least-skilled jobs are affected more. This has meant the demand for workers with low-skills has decreased considerably, and regions where more people have lower skills tend to have higher unemployment rates due to robots. Commonly, workers from manufacturing move to transport, construction and maintenance to get new jobs, but these area are also vulnerable to automation.

Of course, taking an even longer time perspective, increasing automation also boosts economic growth and jobs. On average, installing a robot in lower-skilled regions could lead to twice as much job losses than installing a robot in a high skilled region. Even though the negative impacts of automation are felt more by the regions with lower-skilled

[5] See T. Cusbert, 'Estimating the NAIRU and the Unemployment Gap', *RBA Bulletin*, June Quarter, 2017, pp. 13–22.

workers, overall, automation has increased global GDP and jobs. However, this will create more issues in terms of economic inequality and political stability.[1]

Jobs that are of repetitive nature will be hard hit, whereas jobs that require more compassion, creativity and intelligence are likely to be exclusively carried out by humans. The massive challenge for the governments is to find ways to encourage innovation, such as robots, while making sure that they do not create divides in society. The policymakers, business leaders, teachers and workers need to think about how to develop workforce skills to adapt to growing automation.

We have seen that robots are putting every job from the factory worker to the white-collar executives at stress as all jobs are vulnerable to automation. The following also places the issue in a global perspective by reporting on 21st-century crystal-ball gazing:

> Will robots be stepping into the operating room? Someday [we may] see surgeons operating on patients thousands of miles away across the globe by sending 'microrobots' inside patients' bodies to clamp and cut and sew and transferring medical information and images anywhere in seconds.[2]

Interestingly, we already see elements of this in medicine. The development of such devices as digital stethoscopes has allowed nurses and paramedical staff in remote areas of Australia to relay essential and quite detailed medical information on patients with critical conditions to major hospitals many hundreds of kilometres away, often saving lives but also removing the need for as many doctors in such areas.

What do you think?

1. Are the manufacturing workers experiencing seasonal, frictional, structural or cyclical unemployment? Explain.
2. What solution would you, therefore, propose for any manufacturing worker displaced by such changing technology?

[1] BBC, 'Robots to replace 20 million factory jobs by 2030', 26 June 2019, https://www.bbc.com/news/business-48760799

[2] R. Saltus, '"Telemedicine" foresees robots as surgeons', *Boston Globe*, 8 April 1996, sec. 3, p. 2.

Hysteresis
Occurs when the full-employment rate of unemployment increases (decreases) as the actual unemployment rate increases (decreases).

Finally, it is argued that the full-employment rate of unemployment is itself dependent on the recent path of actually measured unemployment. This is called unemployment **hysteresis**. Hysteresis occurs when the full-employment rate of unemployment increases as the actual unemployment rate increases, or vice versa.

There are a couple of explanations suggested as to why hysteresis might exist.

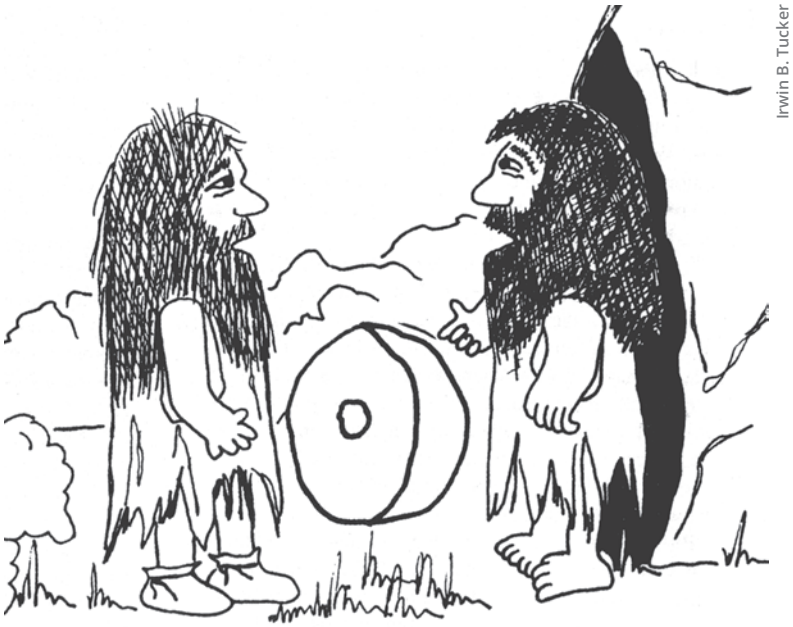
First, if people become cyclically unemployed but stay unemployed for an extended period, then their job skills can atrophy, and their job readiness can reduce. The very fact that they have been unemployed for an extended period of time can make them unattractive to prospective employers. As a result, they can become structurally unemployed and almost unemployable. This may well be an outcome of the GFEC of 2008–09 and also the COVID-19 pandemic in 2019–20. Given the severity of the downturn in some countries, many people in those countries thrown out of work by the crisis may find themselves out of work for many years.

Another argument – referred to as the insider/outsider argument – is that when demand again begins to take off at the end of a recession, those with jobs (insiders) will bid up to market wage rates, thereby making them relatively unattractive recently unemployed (the outsiders) more expensive to hire and increasing the likelihood of their remaining unemployed.

You're the economist

What kind of unemployment did the invention of the wheel cause?

Would the invention of the wheel have caused frictional, structural or cyclical unemployment?



But Caleb, what about the effect on labour?

Non-monetary and demographic consequences of unemployment

The burden of unemployment is more than the loss of potential output measured by the output that could have been produced by the unemployed. There are also non-monetary costs of unemployment.

Without work, many people lose their feeling of self-worth. A person's self-image suffers when they cannot support a family and be a 'valuable' member of society. Research has found high unemployment to be associated with despair, family break-ups, suicides, crime, political unrest, mental illness, heart attacks and other health problems.[6]

Various labour-market groups share the impact of unemployment unequally. **Exhibit 13.10** presents the unemployment rates experienced by selected demographic groups 2019–20. The average overall unemployment rate was 7.4 per cent, but the figures in the exhibit reveal that this unemployment rate is far from equal across different demographic groups.

First, it may surprise some that the unemployment rate for males was higher than that of females. Second, young people (15- to 19-year-olds) experienced a relatively much higher unemployment rate because they are

[6] M. Deady, L. Tan, N. Kugenthiran, D. Collins, H. Christensen and S. B., 'Harvey Unemployment, suicide and COVID-19: using the evidence to plan for prevention', *Medical Journal of Australia*, 2020; Vol. 213, No. 4, pp. 153–154.e1. doi: 10.5694/mja2.50715 P; also, L. Farré, F. Fasani and H. Mueller, 'Feeling useless: The effect of unemployment on mental health in the Great Recession', *IZA Journal of Labor Economics*, 2018, Vol. 7, No. 8, https://doi.org/10.1186/s40172-018-0068-5

new entrants to the workforce who have little employment experience, high quit rates and little job mobility. A second quite exciting feature of the exhibit is that in 2019–20 the unemployment rate among those above 35 right through to government pension age of 65 was quite similar. Also, notice the very low unemployment rate among the over-65s. This is because the vast majority of people in that age bracket either voluntarily leave the labour force to retire or become discouraged in seeking work and subsequently also leave the labour force.

Exhibit 13.10 **Unemployment rates by selected groups, 2019–20**

Demographic group	Unemployment rate (%)
Overall	7.4
Sex	
Male	7.6
Female	7.3
Age (years)	
15–19	19.4
20–24	13.9
25–34	7.7
35–44	5.2
45–54	5.0
60–64	5.5
65 and over	3.0
By unemployment duration (weeks)	**% of total unemployed**
Under 4	17.7
4 to under 13	30.0
13 to under 26	26.1
26 to under 52	10.1
52 and over	16.1

Based on Australian Bureau of Statistics data, Cat. No. 6291.0.55.001, Tables 1 and 14, https://www.abs.gov.au

Analyse the issue

Applicable concepts: The human and social cost of unemployment

Brother, can you spare a dime?

Prolonged unemployment not only means lost wages but also impaired health and social relationships. Developed nations around the world fought their toughest battle against unemployment during the Great Depression of the 1930s and, as a result, the beginnings of today's modern welfare systems were put in place.

As noted earlier, the unemployment rate reached a maximum of 25 per cent during the Great Depression and stayed in the 20 per cent-plus region for quite some years during the 1930s. For comparison, at the low point of Australia's 1990–92 recession – the worst for Australia since the Second World War – the unemployment rate 'only' reached a high of 11 per cent.[1]

However, these statistics tell only part of the horror story. Hundreds of thousands of workers were 'discouraged workers' who had simply given up looking for work because there was no work available, and these people were not counted. During the Great Depression, people were standing in line at soup kitchens, selling produce on the street, or roaming the outback looking for whatever work they could find.

Getty Images/Rolls Press/Popperfoto

The Australian movie *The Sundowners* told the story of the men who would arrive at a farm or station at dusk and try to get a meal and a place to sleep that night by doing some odd jobs. In America, 'Brother, can you spare a dime?' was a common greeting.

A 1992 US study estimated the frightening impact of sustained unemployment not reflected in official unemployment data. Mary Merva, a University of Utah economist, co-authored a study of unemployment in 30 selected big cities from 1976–90. The finding of this research was that a 1 percentage point increase in the national unemployment rate resulted in 6.7 per cent more murders, 3.1 per cent more deaths from stroke, 5.6 per cent more fatal heart disease and a 3.9 per cent increase in suicides.[2]

There is no question to analyse here, but it is essential to understand the broad social and human cost of unemployment in addition to the direct economic cost.

[1] Of course, for many other countries, the GFEC recession of 2008–09 has been the worst recession since the Great Depression.
[2] R. Davis, 'Recessions cost lives', *USA Today*, 16 October 1992, News, p. 1A.

Finally, an analysis of unemployment by unemployment duration reveals that in 2019–20, approximately 16 per cent of the unemployed had been unemployed for over a year. Referred to as 'the long-term unemployed', it is this group for whom it is most challenging to find new jobs.

Interestingly, at the time of writing of the first edition of this book in 2001, the latest data then available for 1999–2000 indicated that this group accounted for 29 per cent of all unemployed persons. By 2003–04, significant progress had evidently been made in reducing the proportion to 21 per cent, and by 2006–07 the proportion fell again to 17 per cent and was again at this rate in 2009–10. Regrettably, due to the relatively weaker Australian economy since 2009–10, the percentage of long-term unemployed in the pool of unemployed in 2016–17 had risen back up to above its 2003–04 level. Even though the unemployment duration is in a decreasing trend, we may see that this duration would increase due to the negative impacts of the COVID-19 pandemic. By the time of writing this edition, the long-term unemployed had peaked at 16 per cent.

Other indicators

Labour force participation rate and **employment to population ratio** can be used to understand the key trends in the labour force in Australia. Labour force participation rate is calculated as:

$$\text{labour force participation rate} = \frac{\text{labour force}}{\text{civilian population 15 years and above}} \times 100$$

Labour force participation rate
The percentage of people aged 15 years and over that are in the labour force.

Employment to population ratio
The number of people aged 15 years and over that are employed as a percentage of the population that are 15 years and above.

It is important to calculate employment to population ratio and calculated as:

$$\text{employment to population ratio} = \frac{\text{employment}}{\text{civilian population 15 years and above}} \times 100$$

Exhibit 13.11 shows the statistics for labour force participation rate and employment to population ratio by gender. Labour force participation rate and employment ratio for the male is higher than the overall and female rates are lower than the overall. On average, the labour force participation rate has increased over time, and an increase in female labour force participation played a significant role in this increase. However, male labour force participation rate has declined over time. Although males have higher employment to population ratio than females, the gap between males and females has decreased over time.

Exhibit 13.11 **Labour force participation rate and employment to population ratio by gender, 2019–20**

Demographic group	Indicator (%)
Labour force participation rate – Overall	64.0
Labour force participation rate – Male	69.1
Labour force participation rate – Female	59.0
Employment to population ratio – Overall	59.2
Employment to population ratio – Male	63.9
Employment to population ratio – Female	54.7

Based on Australian Bureau of Statistics data, Cat. No. 6291.0, https://www.abs.gov.au.

In summary

- **Seasonal unemployment** is caused by changes in hiring due to recurring changes in weather conditions, demand and/or production patterns.
- **Frictional unemployment** is caused by the normal search time required by people changing jobs or entering/re-entering the labour force.
- **Structural unemployment** is caused by a fundamental mismatch of the skills of the unemployed and the skills required for existing job opportunities.
- **Cyclical unemployment** is caused by the lack of sufficient jobs during a recession or a growth slowdown.
- Demographically, those under 25 experience relatively higher rates of unemployment.
- **Full employment** is that rate of unemployment that involves zero cyclical unemployment.
- The full employment rate of unemployment – or the 'non-accelerating inflation rate of unemployment' (NAIRU) – while being dependent on the recent path of actually measured unemployment (unemployment hysteresis), was, in 2017 in Australia, estimated to be around 5 per cent.
- **Labour force participation rate** is the percentage of people aged 15 years and over that are in the labour force.
- **Employment to population ratio** shows the number of people aged 15 years and over that are employed as a percentage of the population that are 15 years and above.

5 Inflation and unemployment: is there a relationship?

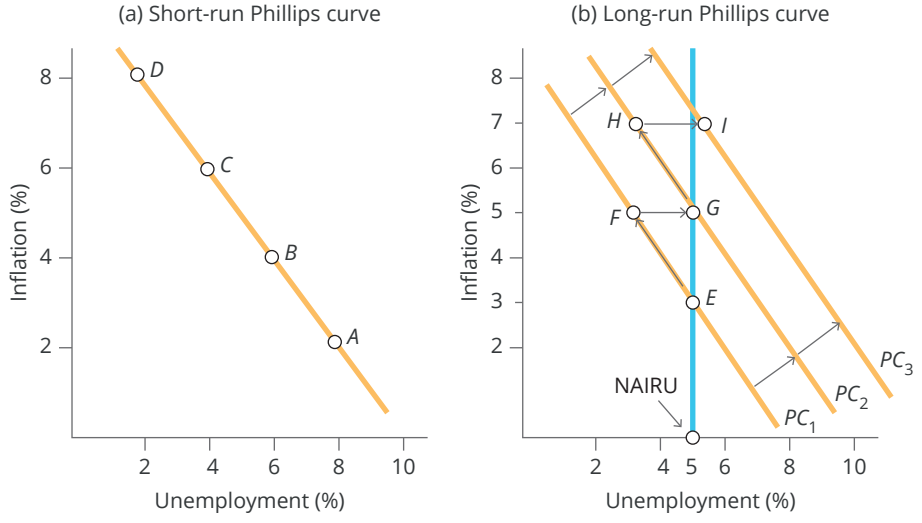
Before we leave this chapter, it is natural to ask whether there may be a relationship between the rate of inflation in a country and its unemployment rate. In the 1950s a famous New Zealand economist Bill Phillips, who spent much of his academic career at the London School of Economics, argued that, under certain circumstances, there could be expected to be a trade-off between them. His argument was based on the demand-pull inflation outlined above.

At its simplest level, the argument is that as unemployment increases, demand pressures in the economy come down and induce suppliers to slow the rate of increase in product prices. Moreover, on the production side, due to the reduced demand, the rate of wage inflation would slow as producers resisted wage rises, thereby easing cost pressures. This would take the pressure off prices, and the rate of inflation could be expected to reduce.

On the other hand, as the rate of unemployment reduces and approaches the 'full employment' rate of unemployment, demand pressures would put pressure on the product supply chain, bottlenecks would occur, wage inflation would accelerate, production costs would be forced up, and suppliers would respond by increasing prices more rapidly. Thus, lower unemployment rates could be expected to be associated with higher rates of inflation.

This leads to the conclusion that there may exist an inverse, or trade-off, relationship between unemployment and inflation. This is depicted stylistically in part (a) of **Exhibit 13.12** where, for example, points *A*, *B*, *C*, and *D* all represent different inflation/unemployment combinations on the stylised Phillips curve. Inflation and unemployment data from Britain in the 1950s and 1960s, and from other countries as well, provided empirical support for this theoretically argued inverse relationship. This led policymakers in several countries to conclude that a conscious policy choice could be made whereby a country's macroeconomic policymakers could 'manufacture' a persistently very low unemployment rate provided they were willing to accept a relatively high rate of inflation. For example, in the graph (a), a conscious policy choice could consist of say, point *C* or even *D*.

Exhibit 13.12 **The short-run and long-run Phillips curves**



However, towards the end of the 1960s, another very famous economist – an American by the name of Milton Friedman – argued that any such trade-off could only be expected to be very short-run. He argued that any attempt by policymakers to manufacture a lower rate of unemployment than the so-called 'natural rate' would, after a time, raise the rate of inflation, which in turn would increase inflationary expectations, which in turn would further increase the actual inflation rate. This would destabilise business decision-making, which would undo any short-run lowering of the unemployment rate below the natural rate, with the result merely being increasing rates of inflation.

Thus, Friedman argued there would be no long-run trade-off. In other words, he argued the long-run Phillips curve is vertical at the natural rate of unemployment or NAIRU. The evidence from the 1970s and 1980s supported him. The dynamics over time are stylistically represented in part (b) of **Exhibit 13.12**. Initially, the economy is at point *E* with inflation – and inflationary expectations – at 3 per cent and unemployment at the NAIRU of 5 per cent. The authorities then seek to lower unemployment below the NAIRU by stimulating increased spending and economic activity. In the short-term, unemployment is reduced, with higher inflation the result. This is represented by point *F* on the first short-run Phillips curve, labelled PC_1.

However, the higher inflation outcome increases people's inflationary expectations, and this is represented by shifting out of the short-run Phillips curve from PC_1 to PC_2. The economy has now moved back to the NAIRU but with a higher inflation rate (point *G*). If the authorities again sought to reduce unemployment below the NAIRU, even higher inflation would again result (point *H*), but then again the short-run Phillips curve would shift out (PC_3) due to people now having even higher inflationary expectations. In the next period, the economy is depicted as moving back to the NAIRU but with even higher inflation (at point *I*). Thus, points *E*, *G* and *I* lie on a vertical long-run Phillips curve centred on the NAIRU. While this is highly stylised, as noted above, the evidence from the 1970s and 1980s supported this theoretical proposition.

Finally, both Phillips' and Friedman's arguments were primarily driven by considerations of the demand side of the economy. However, inflation can also originate from increasing costs on the supply side of the economy – like the oil price shocks of the 1970s. As will be seen in more detail in the next chapter, such supply-side shocks can be expected to produce both increasing price levels and higher unemployment, a phenomenon known as 'stagflation'. This was a term coined in the late 1970s to reflect the simultaneously occurring high inflation and increasing unemployment rates experienced by many countries around the world in that decade.

In summary

- The so-called '*Phillips curve*' posits an inverse relationship between inflation and unemployment. The supporting argument was based on the demand-pull-type inflation outlined earlier in this chapter.
- Some policymakers concluded it might be possible to 'manufacture' a persistently very low unemployment rate along with a relatively high rate of inflation.
- Milton Friedman argued that any such trade-off would only be very short-run, arguing that any attempt to manufacture a lower rate of unemployment than the NAIRU would, after a time, raise inflation, increase inflationary expectations, which in turn would further increase the actual inflation rate.
- According to Friedman, the *long-run Phillips curve* would be vertical at the NAIRU with no long-run trade-off between inflation and unemployment.
- Both Phillips' and Friedman's arguments were driven by considerations of the demand side of the economy only.

Key concepts

Inflation
Deflation
Disinflation
Consumer price index (CPI)
Base year
Demand-pull inflation
Cost-push inflation
Wealth
Nominal (or money) income
Real income
Nominal interest rate
Real interest rate
Hyperinflation
Wage–price spiral
Unemployment rate
Civilian labour force
Discouraged workers
Underemployment
Seasonal unemployment
Frictional unemployment
Structural unemployment
Cyclical unemployment
Full employment
Hysteresis
Labour force participation rate
Employment to population ratio

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand the meaning and measurement of inflation

- **Inflation** is an increase and **deflation** is a decrease in the overall average level of prices of goods and services in the economy.
- Disinflation is a reduction in the rate of inflation from some higher rate to a lower rate.
- The most widely reported measure of inflation is the **consumer price index (CPI)**, which measures changes in the average price of consumer goods and services. It is sometimes called the cost-of-living index since it measures the change in the cost of a fixed basket of products over time as the prices of those products change.
- Some criticisms of the CPI relate to its use of a single 'typical' basket of products, the impact of quality changes and the impact of the so-called 'substitution effect'.
- Rising prices can be the result of demand-side pressures or cost-push factors or a combination of both.

2. Develop an understanding of the consequences of inflation

- People whose **nominal income** rises more slowly than the rate of inflation lose real purchasing power.
- A change in **real income** can be approximated by subtracting the rate of inflation from the percentage rise in nominal income.
- Similarly, the **real rate of interest** on a loan is approximately the nominal interest rate charged less the rate of inflation.
- Sometimes, if lenders underestimate future inflation, they can experience a negative real rate of return on loans they have made.
- With a higher and more unstable inflation rate, greater numbers of inappropriate decisions are made by firms and individuals, resulting in reduced efficiency with which the community's scarce resources are allocated.
- All of these problems are amplified manifold during periods of **hyperinflation** where prices can be rising at more than 50 per cent each month.

3. Examine the meaning and measurement of unemployment

- The **civilian labour force** includes those people 15 years of age and older who are employed or who are actively seeking a job, excluding those in the armed forces, homemakers, students, **discouraged workers** and other persons not in the labour force.
- The **unemployment rate** is the percentage of people in the labour force who are without jobs and are actively seeking jobs.
- There are two criticisms of the measured unemployment rate:
  - it regards **discouraged worker** – those who would like to work, but who have given up searching for a job because they believe there will be no offers – as being out of the labour force
  - the existence of **underemployment** – those employed who would like to work more hours if they could (expressed as a % of the labour force) – may understate the degree of labour unemployment in the economy.

4. Understand the different types of unemployment

- The four conceptual categories of unemployment are:
  - **Seasonal unemployment** is due to recurring changes in weather conditions, demand and/or production patterns.
  - **Frictional unemployment** arises from the normal search time required by people with marketable skills changing jobs or entering/re-entering the labour force.
  - **Structural unemployment** is caused by a fundamental mismatch of the skills of the unemployed and the skills required for existing job opportunities.
  - **Cyclical unemployment** is caused by the lack of sufficient jobs during a recession or a growth slowdown.
- Demographically, those under 25 experience relatively higher rates of unemployment.
- **Full employment** is that rate of unemployment that involves zero cyclical unemployment. The full employment rate of unemployment – sometimes referred to as the 'natural rate of unemployment' or the 'non-accelerating inflation rate of unemployment' (NAIRU) – changes over time and, in 2017 in Australia, it was estimated to be around 5 per cent. The NAIRU is itself dependent on the recent path of actually measured unemployment (unemployment hysteresis).

5. Examine whether there is a relationship between inflation and unemployment

- A. W. (Bill) Phillips argued that, under certain circumstances, there could be expected to be a trade-off – an inverse relationship – between inflation and unemployment (such an inverse relationship being called a *Phillips curve*). His argument was based on the **demand-pull inflation** outlined earlier in this chapter. This led some policymakers to conclude they could 'manufacture' a persistently very low unemployment rate along with a relatively higher rate of inflation.
- Milton Friedman argued that any such trade-off would only be very short run in nature arguing that any attempt by policymakers to manufacture a lower rate of unemployment than the NAIRU would, after a time, raise the rate of inflation, thereby increasing inflationary expectations, which in turn would further increase the actual inflation rate. This would destabilise business decision-making, which would undo any short-run lowering of the unemployment rate below the natural rate, with the result merely being increasing rates of inflation. The *long-run Phillips curve* would, therefore, be vertical at the NAIRU.
- Both Phillips' and Friedman's arguments were driven by considerations of the demand side of the economy. However, inflation can also originate from increasing costs on the supply side of the economy. Such supply-side shocks can be expected to produce both increasing price levels and higher unemployment.

Study questions and problems

1 Compare and contrast 'disinflation' and 'deflation'.

2 Suppose your annual income in 2020 was $135 000. Assuming the base year for the CPI was 2011, calculate your real 2020 income measured in 2011 dollars, assuming the CPI was 125 in 2020.

3 Explain the following statement: 'If everyone expects inflation to occur, it will!'

4 Explain the 'substitution bias effect' criticism of the CPI.

5 What is 'demand-pull' inflation, and how does it differ from 'cost-push' inflation?

6 Suppose the annual nominal rate of interest on a fixed-term bank deposit is 5 per cent. What would be the real rate of interest over the term of the deposit if the inflation rate turned out to be 7 per cent?

7 Describe the relevant criteria that government statisticians use to determine whether a person is 'unemployed'.

8 Describe two ways in which the official unemployment rate has been criticised for underestimating unemployment?

9 If, over two periods, the unemployment rate had reduced from 6 per cent to 5.5 per cent and the measured labour underutilisation rate had reduced from 12 per cent to 10 per cent, has the underemployment rate increased or decreased and by what percentage?

10 Why is frictional unemployment inevitable in an economy?

11 Explain the full-employment rate of unemployment (alternatively known as the natural rate of unemployment)?

12 Do you agree that one goal of macroeconomics should be to reduce the measured unemployment rate to zero? Why or why not?

Answers to 'You're the economist'

Interpretation of a CPI number

If you said the average level of consumer prices – as measured by the CPI – in 2018–19 was 116.2 per cent of that of 2011–12, or, that the average level of consumer prices in 2018–19 was 16.2 per cent higher than that of 2011–12, then YOU ARE THINKING LIKE AN ECONOMIST.

The university education price index

If you said the price of Anne's university education increased by 2.4 per cent in 2020 compared to 2019, YOU ARE THINKING LIKE AN ECONOMIST.

What kind of unemployment did the invention of the wheel cause?

The invention of the wheel represented a new technology for people in ancient times. Many workers who transported goods lost their jobs, even in the primitive era, to the more efficient cart with wheels. If you said the invention of the wheel caused structural unemployment, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 As shown in **Exhibit 13.12**, the calculated rate of inflation for Year 4 is closest to:
- **a** 12 per cent.
- **b** 1.07 per cent.
- **c** 3.7 per cent.
- **d** 7 per cent.

Exhibit 13.13 Consumer price index

Year	Consumer price index
1	100
2	104
3	108
4	112
5	116

2 As shown in **Exhibit 13.13**, the actual measured rate of inflation from Year 2 to Year 5 is nearest to:
- **a** 16 per cent.
- **b** 11.5 per cent.
- **c** 3.6 per cent.
- **d** 12 per cent.

3 Consider an economy with only two goods: bread and wine. Let's say that, in 2008, the typical family bought four loaves of bread at $1.50 per loaf and two bottles of wine for $9 per bottle. In 2018, let's say bread cost $4.50 per loaf and wine cost $16 per bottle (fictitious prices). The CPI for 2018 (using 2008 as the base year) is approximately:
- **a** 195.
- **b** 208.
- **c** 239.
- **d** 110.

4 Suppose a typical car tyre cost $50 in 2005 and had a useful life of 40 000 km. In 2015, let's suppose the typical car tyre cost $75 and had a useful life of 50 000 km. Which statement do you think is most correct?
- **a** It is not possible to estimate 'tyre inflation' between the two years due to the difference in expected useful life.
- **b** The best measure of tyre inflation is to ignore the change in expected useful life and estimate it as 50 per cent.
- **c** An approximate 'quality-adjusted' rate of tyre inflation would be 25 per cent.
- **d** While some adjustment might be possible for the longer expected life, when quality changes occur, it is best not to try to calculate an inflation measure in such a situation.

5 Last year the Carter family earned $50 000. This year their income is $52 000. In an economy with an inflation rate of 3 per cent, which of the following is correct?
- **a** The Carters' nominal income and real income have both risen.
- **b** The Carters' nominal income and real income have both fallen.
- **c** The Carters' nominal income has fallen, and their real income has risen.
- **d** The Carters' nominal income has risen, and their real income has fallen.

6 The seasonal unemployed are those people who:
- **a** always quit their jobs at Christmas time owing to the oppressive heat.
- **b** lose their jobs at regular times of the year because the work they do occurs at certain times of the year and not at other times.
- **c** find it difficult to find work because, for medical reasons, they can't work in different climates.
- **d** are thrown out of work when people's tastes for particular food changes.

7 'Full employment' is defined as a situation when:
- **a** both structural and cyclical unemployment are zero.
- **b** cyclical unemployment is zero.
- **c** all people who want jobs have them.
- **d** all of the above.

8 Hannah is working as a waitress in a restaurant and she lost her job due to the economic recession. What type of unemployment could she be classified into? Unemployment due to a recession is:
- **a** frictional unemployment.
- **b** recessionary unemployment.
- **c** structural unemployment.
- **d** cyclical unemployment.

9 Structural unemployment is caused by:
- **a** the impact of the business cycle on job opportunities.
- **b** temporary lay-offs due to long periods of bad weather in the construction industry.
- **c** shifts in the economy that make certain job skills obsolete.
- **d** workers finding a job in a different sector of the economy after a short period of unemployment.

10 If, over two periods, the unemployment rate had reduced from 6.5 per cent to 5.5 per cent and the measured underemployment rate had increased from 7.5 per cent to 8.2 per cent, the change to the labour underutilisation rate would have been:
- **a** an increase of 0.7 per cent.
- **b** an increase of 0.3 per cent.
- **c** a decrease of 0.3 per cent.
- **d** a decrease of 0.7 per cent.

PART 5

MACROECONOMIC THEORY AND POLICY

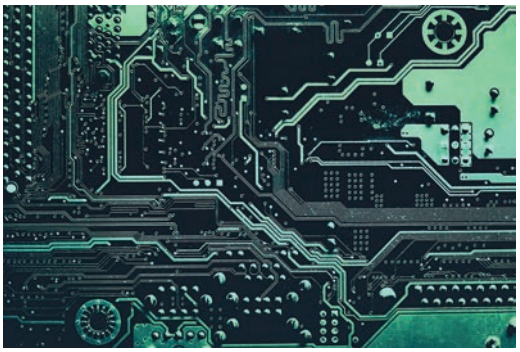
14 A simple model of the macro economy

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1. Understand the economic determinants of the four key expenditure components of aggregate demand in the economy.
2. Develop an understanding of the aggregate demand–aggregate supply (*AD–AS*) macroeconomic model.
3. Appreciate how the *AD–AS* macroeconomic model can be used to understand the business cycle and how macroeconomic policy can affect prices, GDP and employment.

In the 20th century, the 1920s were known as the 'Roaring 20s'. This was a time of optimism and prosperity. Australia had one of the highest levels of real gross domestic product (GDP) per capita in the world. Stock prices rose strongly year after year and unemployment was low. Then, after October 1929, the business cycle took an abrupt downturn. Following the lead from New York, stock markets around the world crashed, and the most severe world recession of the 20th century began within a year. During this terrible economic contraction, which became known as the Great Depression, stock prices fell, wages fell, real output fell, banks failed, businesses closed their doors, and the unemployment rate soared to 25 per cent.



Shutterstock.com/Elena11

Similarly, in the 21st century, while the global financial and economic crisis (GFEC) of 2008–09 did not produce hardship on the scale of the Great Depression, it nonetheless ushered in a relatively rare macroeconomic event – a synchronised international recession. Most of the countries of the developed world quickly went into recession, literally within a few months of each other, and the world's share markets crashed to levels less than half of what they had been less than a year earlier. In many countries, unemployment soared quickly to the highest levels since the Great Depression. In the US, commentators began calling the episode as it played out in that country the 'Great Recession'.

Moreover, an unexpected event – the COVID-19 pandemic – affected almost all the countries in the world in 2019–20. It is yet too early to analyse the scale of the hardship (at least during the time of writing the book). It is expected that this event will not lead to another Great Depression, but the impact will be more severe than the GFEC. Most of the countries who were greatly affected by the virus quickly went into recession. In many countries, unemployment soared quickly to the highest levels since the GFEC.

The misery of the Great Depression created a revolution in economic thought which, some 75 years later, arguably allowed better international policy responses to the GFEC. The result was much less severe outcomes than may otherwise have occurred had policymakers employed the same responses as were used in 1930–34. Had it been otherwise, the so-called Great Recession of 2008–09 may well have become the *Second Great Depression*.

Before the Great Depression, economists had recognised that over the years, business downturns would interrupt a nation's prosperity, but they believed these episodes were temporary and self-correcting. They argued that in a short time the price system would reliably and automatically restore an economy to full employment without government intervention. However, even after four years or more, the economy didn't self-correct back to its 1929 level of real output. What went wrong? The stage was set for some new ideas offered by the great British economist John Maynard Keynes (pronounced 'canes'). Keynes argued that the economy was not self-correcting and that unemployment could indeed continue indefinitely because of inadequate aggregate spending. Keynes's work not only explained the Depression but also offered policy prescriptions that required the government to play an active role in managing the economy.

While not all modern economists agree with Keynes, his ideas still influence macroeconomics today. Indeed, in responding to the very severe synchronised international recession of 2008–09 and 2019–20, many governments and central banks around the world – including in the US, China, Japan, Germany, the UK and Australia – actively pursued stimulatory macroeconomic policies that owed much to the ideas of Keynes.

This chapter begins with a brief discussion of economic thinking before Keynes and then explains his reasons for rejecting that thinking. Then you will learn more about what determines the various components of aggregate demand expenditures (C, I, G and $X - M$) and how these components are combined to form a simple demand-oriented macroeconomic model. This simple model offers an understanding of a cure for an economy in a severe recession where inflation is not a problem. The policy prescription for such an economy is for government to expand aggregate demand (through its spending or by cutting taxes), raise output and production, create jobs and restore full employment.

Later, this demand-oriented model is extended to incorporate issues relating to aggregate supply, and you will learn how aggregate demand and supply analysis can be used to understand more fully the forces that generate the business cycle. We will open with a presentation of the aggregate demand curve and then the aggregate supply curve. Once these concepts are developed, the analysis will show why modern macroeconomics teaches that shifts in aggregate supply or aggregate demand can influence the price level, the equilibrium level of real GDP and employment.

This chapter is essential in providing you with the framework you will need to understand modern macroeconomic debate and policymaking. You will probably return to this chapter regularly because it provides the necessary tools with which to organise your thinking about the macro economy.

1 Economic determinants of the four key expenditure components of aggregate demand

We begin by first elaborating on why an understanding of the economic determinants of the four key expenditure components of aggregate demand in the economy is so important. To set the scene, we briefly explain the significance of the Great Depression in precipitating a revolution in the way economists understood the workings of the macro economy; the revolution in question was the Keynesian revolution.

Introducing the Keynesian revolution

Before the Great Depression, a group of economists known as the *classical economists* dominated economic thinking. The founder of the classical school of economics was the great Adam Smith. The classical economists believed that the forces of supply and demand, which you studied in **Chapter 4**, would naturally achieve full employment in the economy because flexible prices (including wages and interest rates) in competitive markets bring all markets to equilibrium. Markets – perhaps after a relatively short adjustment period – always clear, enabling firms to sell all goods and services offered for sale.

In brief, their position can be explained as follows. Recall the circular flow model explained in **Chapter 11** on GDP. Suppose a firm produces 1000 computer laptops that it intends to sell for $500 each to consumers. Suppose $100 000 is needed to pay input costs (including interest costs), leaving $400 000 of the expected revenue to be paid as income for the household sector, paid out, say, in wages to the firm's employees at going rates of pay, and profits. If all of the production (supply) is sold at the expected price, this, therefore, creates an equivalent amount of income to facilitate purchases (demand) in product markets.[1] What if the 1000 laptops could not be sold for $500 each? The classical economists argued that prices *and most importantly wages* would quickly drop to clear the laptop market and maintain full employment (but at a lower wage). If unemployment were to occur, it would be the result of a temporary adjustment period during which wages and prices declined in order to clear markets. Anyone unemployed after the short-run adjustment period had passed would have chosen that state voluntarily rather than work at the new prevailing wage rates.

In 1936, John Maynard Keynes published *The General Theory of Employment, Interest and Money*.[2] In his book, Keynes challenged the classical thinking by placing great primacy on the demand side of the economy rather than on the production or supply side. He observed the very high rates of unemployment in the Great Depression, that these high rates persisted for many years, and that the unemployed in question were not choosing to be unemployed voluntarily. Keynes, therefore, argued persuasively that, over long enough periods to be material of concern, *aggregate expenditure* could indeed be inadequate for an economy to achieve full employment, and that the economy could well settle into a prolonged period of high involuntary unemployment.

Aggregate expenditure is the sum of consumption (C), investment (I), government spending (G) and net exports ($X - M$). Aggregate expenditure is also known as *aggregate spending* or *aggregate demand*. Recall from **Chapter 11** on GDP that C, I, G and ($X - M$) are national accounting categories used to calculate GDP following the expenditure approach. The next few pages are devoted to outlining the economic determinants of the consumption, investment, government and net exports expenditure components of aggregate demand briefly.

Consumption demand Consumption is the most significant component of aggregate expenditure (around 60%), and the most important determinant of a household's demand for consumer goods and services is disposable income.

Marginal propensity to consume (MPC) The change in consumption resulting from a given change in disposable income.

Consumption demand (C)

What determines your family's willingness to buy food, clothing, cars, education and other consumer goods and services; that is, what determines your family's **consumption demand**? The most important factor in determining consumption demand is considered to be disposable income (income after taxes). If take-home pay increases, consumers can be expected to increase their spending – but not necessarily by the full amount of the extra income (see the following for further discussion). Recall from **Chapter 11** that consumption (C) is the most significant single component of aggregate expenditures. In most modern developed economies it now accounts for about 60 per cent of total production.

Marginal propensities to consume and save

As income grows, so does consumption, but by less than income. This crucial concept, called the **marginal propensity to consume (MPC)**, is the change in consumption resulting from a given change in disposable income. For example, an MPC of 0.75 means that for every dollar increase (decrease) in disposable income, consumption should increase (decrease) by 75 cents.

---

1 In the early 1800s, Jean-Baptiste Say developed what came to be known as Say's Law which helped convince classical economists that a prolonged depression was impossible. Say's Law states that the production of goods and services (supply) generates an equal amount of total spending (demand) for these goods and services. Put merely, Say's Law is the theory that 'supply creates its demand'.

2 J. M. Keynes, *The General Theory of Employment, Interest and Money*, Macmillan, London, 1936. *Inter alia*, Keynes repudiated Say's Law and, in effect, argued that 'demand creates supply' rather than 'supply creates demand'.

In each model developed throughout this text, we assume that the MPC is constant for all income levels. This is admittedly an unrealistic assumption, like those on lower incomes will undoubtedly have higher MPCs than those on higher incomes. Nevertheless, none of the fundamental ideas discussed in the text is changed by recognising that MPCs do vary for different income groups.

You're the economist

What is your MPC?

As your income increases over time, your MPC can remain constant, or it can change. Would you expect your MPC to increase, decrease or remain constant as your income increases throughout your career?

What do households do with extra disposable income if they do not spend it? They save it. The **marginal propensity to save (MPS)** then is the change in savings resulting from a given change in disposable income; it is, therefore 1 – MPC. In our example, the MPS is 0.25; so, for every dollar increase in disposable income, savings should increase by 25 cents.

Marginal propensity to save (MPS) The change in savings resulting from a given change in disposable income, namely 1 – MPC.

Factors other than income affecting consumption behaviour

Expectations

Consumer expectations are subjective views of the future (optimistic or pessimistic), which can change consumption spending in the present. Expectations may involve future price changes, the likelihood of becoming unemployed, the likelihood of receiving higher income, or the possibility of a future shortage of products. Suppose households believe prices will be much higher next year and they, therefore, decide to buy now. This would trigger current spending and increase consumption across income groups. On the other hand, the anticipation of a recession and fears about losing jobs would make families more frugal in their current spending, which decreases consumption spending across all income groups. This was experienced during the COVID-19 pandemic, whereby people were pessimistic about the economy and their jobs, so they decreased their spending.

Wealth

Holding all other factors constant,[3] the more wealth households accumulate, the more they spend at any current level of disposable income. The surge in share prices in the late 1990s, particularly in the US, is widely considered to have been an essential cause of the very rapid growth in consumption in that country during that period. Similarly, the sharp and prolonged increases in the Australian stock market in the mid-2000s is widely regarded as being a significant factor in increased consumption spending in Australia. Furthermore, the dramatic crash in world stock markets beginning in September 2008 decreased household spending remarkably. In addition, most recently, the adverse effects on the world stock market and the property market were widely regarded as potentially having a severely depressing effect on household consumption spending.

Wealth owned by households includes both *real* assets, such as homes and cars, electronic goods and so on, and *financial* assets, such as, cash, savings accounts, shares, bonds, superannuation, managed funds and insurance policies. Changes in stock market prices, real estate prices and prices of other assets affect the value of wealth and, in turn, can change the willingness of the nation's consumers to spend.

[3] This is a favourite phrase of economists. The words *ceteris paribus* explained in Chapter 1, mean the same thing. Of course, in the real world, other things are never constant but are changing all the time. Economists realise this, but if you are to have any hope of understanding how changes in a whole host of things can be expected to affect others, and then to grasp the fundamental causal mechanisms at work, it is useful to isolate each different determinant and analyse what its impact would be if nothing else were to change.

The price level

Any change in the general price level changes consumption demand by reducing or enlarging the *purchasing power* of financial assets (wealth) with fixed nominal value. If the real value of financial wealth falls, families feel less well-off and spend less at any level of current disposable income. The next section discusses this phenomenon in more detail.

The interest rate

Some consumption – for example, spending on consumer durables like expensive kitchen appliances, electronic gear, boats and cars – often requires borrowing to finance the spending. A lower rate of interest on loans encourages consumers to borrow more, and a higher interest rate discourages consumer borrowing. If interest rates fall, households across the various income groups will be inclined to use more credit to finance consumer purchases. Interest rates are often considered as policy variables, and the central banks in the country play a significant role in determining the interest rates. **Chapter 15** discusses this in greater detail.

Investment demand (*I*)

It is widely accepted that changes in the private-sector components of aggregate expenditures (personal consumption and investment spending) are the major cause of the business cycle. Moreover, the more volatile of these two components is investment spending.[4] Recall from **Chapter 11** that investment expenditure ('gross private domestic investment') consists of spending on newly produced residential and non-residential structures, plant and equipment, and changes to inventories.

There are two significant determinants of investment spending, namely expectations of future returns from a potential investment and the interest rate, which is the financing cost of any investment proposal. By way of illustration, suppose a local pharmacist is considering borrowing $3000 to purchase a new desktop computer/printer platform that, just for argument's sake, she knows will be obsolete in a year and will need to be replaced. The pharmacist expects the new computer will help her increase the pharmacy's annual sales revenue and hopes the increase might be as much as $3300. Thus, assuming no taxes and other expenses exist, the expected (hoped for) rate of return from this investment spending is 10 per cent ($300/$3000).

However, what about the impact of the interest cost of borrowing the $3000 to purchase the computer? If the interest rate is less than 10 per cent per annum, the business will expect to earn a positive return from the investment, so the pharmacist will be inclined to buy the computer. A rate of interest higher than 10 per cent means the investment will result in a loss, so she would not go ahead with the purchase. In other words, in general, businesses can be expected to undertake all planned investment projects for which the expected rate of return exceeds the interest rate. In aggregate, and ignoring other determinants of investment demand, this suggests that aggregate investment spending in the economy will increase as interest rates drop.

Why, then, is investment demand so unstable? The reason is that there are also many volatile determinants influencing investment demand other than interest rates. Any factor that decreases the expected rate of return on planned investment spending can be expected to lower actual investment spending. Moreover, for entirely understandable reasons, businesspeople are susceptible to moods of optimism and pessimism concerning economic conditions. Their *expectations* (as Keynes defined it, their 'animal spirits') about the future translate into forecasts of future sales, future costs and future profitability of investment projects.

These forecasts involve a clouded crystal ball, requiring a high degree of intuition or subjective analysis. There are always so many ever-changing factors, such as consumer sentiment, government spending and tax

[4] The relatively greater stability of consumption spending has been widely observed in many countries over many different periods. It may be due to people being reluctant to change their consumption habits and, if necessary, being willing to draw on their accumulated savings to maintain their consumption.

policies, monetary policy changes by central banks, national and world events, population growth and stock market conditions, that forecasting is invariably complicated.

When a wave of pessimism becomes pervasive, businesspeople reduce their expectations of profitability at any available interest rate. Such a pessimistic attitude can become contagious and reduce investment spending in the entire economy. There were indeed elements of this at work with the onset of the GFEC in 2008–09. During this time, all of the world's major economies entered a synchronised international economic recession, and business expectations about the immediate future became quite pessimistic. During the COVID-19 pandemic, business pessimism spread through to many businesses and investors in the world and decreased the investment spending immensely. At other times businesspeople can become very optimistic and revise their expected rate of profit from investment at any given interest rate upward. Thus, changes in business confidence (expectations) are a significant cause of fluctuations in investment spending.

Other factors affecting investment behaviour

Technological progress includes the introduction of new products and new ways of doing things. A greater level of output can be produced with a given level of inputs with technological progress. Robots, personal computers, fax machines, mobile phones, the internet and similar new inventions provide less costly means of production. New technologies create a flurry of investment spending as firms buy the latest innovations in order to improve their production capabilities, causing investment demand to rise.

Capacity utilisation is also a factor. During a recession, many businesses operate well below their maximum productive capacity. Since much of the nation's capital stock stand idle during such a time, firms have little incentive to invest in more. For example, during the COVID-19 pandemic in 2020, all the airline industries operated below their capacity. Conversely, after a few years of expansion, firms may be operating their plants at a high rate of capacity utilisation and the outlook for sales growth is optimistic. In this case, there is pressure on firms to invest in new projects to meet sales demand, and investment spending increases.

Business taxes can shift investment demand. Business decisions, in reality, depending on the expected *after-tax* rate of profit. An increase in business taxes, therefore, would lower profitability and reduce investment at any given interest rate. On the other hand, the government may wish to encourage investment by allowing, say, a *tax credit* of some form for new investment. A 10 per cent *investment tax credit* means that if a company decided to invest $10 million in significant new equipment, then in addition to the usual allowance for depreciation, the company's tax bill will be cut by $1 million. The effect of this tax policy is to increase the after-tax profitability of new investment projects, and investment demand could be expected to increase.

In summary, *ceteris paribus*, the quantity of **investment demanded** by businesses increases as rates of interest drop. Shifts in investment demand – that is, desired investment spending at any given interest rate – result from changes in profit expectations, technological change, capacity utilisation and business taxes.

Investment demand
Ceteris paribus, the quantity of investment demanded by businesses increases as rates of interest drop. Shifts in investment demand – that is, desired investment spending at any given interest rate – result from changes in profit expectations, technological change, capacity utilisation and business taxes.

Government demand (*G*)

Government consumption and investment expenditures comprise the second-largest component of aggregate expenditure in most modern developed countries. Government spending can be considered as an **autonomous expenditure**. Expenditures that do not respond in any systematic way with identifiable economic determinants, such as interest rate changes or income changes, are said to be autonomous. The reasoning is that government spending is primarily the result of a political decision (often considered as a policy variable), which may be largely independent of the level of national output or interest rates.

This category was crucial to Keynes's policy prescription for dealing with severely recessed economies. In such situations, Keynes thought that, due to negative sentiment about the future, the private sector would be very reluctant to expand its consumption and investment spending. What would be needed would be a 'kick start' from new spending by the government to stimulate the economy, improve sentiment about the future and thereby get private-sector activity going (more on this below).

Autonomous expenditure
Spending that does not vary in any systematic way, with identifiable economic determinants, such as, the current level of disposable income or interest rates.

Net export demand ($X - M$)

Like government spending, exports can be treated – at least as a first approximation – as autonomous expenditures unaffected by a nation's domestic level of real GDP. Economic conditions in the countries that buy Australian products affect the level of our exports. On the other hand, the level of imports purchased by Australian residents is significantly influenced by the economic conditions in Australia. As Australian real GDP increases then – other things being equal – we can expect imports to increase and net exports ($X - M$) to reduce. Of course, for a given level of Australian real GDP (thereby determining a certain level of import demand in Australia), we can expect net export demand to increase if world economic activity increases (or if the world economies experience significant economic growth).

Other important determinants of net export demand are changes in exchange rates and the terms of trade. More will be said on this in **Chapter 18** on international trade and finance. For the time being, however, it is useful to sketch out the main points.

As the international price of a country's currency reduces (exchange rate depreciates), the income in domestic currency terms generated by a nation's export sales increases and this would be likely to encourage more significant exports. At the same time, imports into the country become more expensive in domestic currency terms, thereby discouraging spending on imports. Thus, in this situation, other things being equal, net exports can be expected to increase. The opposite can be expected when the international price of a country's currency increases (exchange rate appreciates).

The terms of trade are defined as the ratio of a country's export prices to the prices of its imports. As the terms of trade increase – for example, they did sharply in the 2000s in Australia on account of very significant increases in world commodity prices – it means the country's exports are becoming relatively more valuable internationally. This encourages exporters to expand production, with the result that net exports can be expected to increase. The opposite result could be expected when the terms of trade deteriorate.

The aggregate demand–output model

Aggregate demand (AD) The total of the four categories of demand, namely $C + I + G + (X - M)$.

Aggregate demand (AD) is simply the total of the above categories of demand, namely $C + I + G + (X - M)$. This is precisely the same way in which we defined GDP in **Chapter 11**, but do not be confused; there, we were simply defining the accounting identity: that GDP was the sum of all the various types of expenditures in the economy in any given period. Any discrepancy between what was produced and what people actually bought was accounted for by intended or unintended accumulation or decumulation in inventories.

While here we are still using these same categories, now we are talking about them as categories of demand. Moreover, it is quite possible that, in a given period, the total aggregate demand by households (C), businesses (I), government agencies (G) and foreigners ($X - M$) will not be precisely equal to the total production that firms had expected to sell. This is because the factors feeding into the spending decisions of all the various economic players are quite different from those factors influencing those who are making the production decisions.

Indeed, as noted earlier, it was Keynes who first persuasively argued in the 1930s that aggregate demand might not necessarily always equal the level of total production required to maintain full employment of the workforce. If aggregate demand is left entirely to private markets, it may remain below the full-employment level of production for extended periods. It thereby debunks the classical economics view of self-correction of the economy in the long-run, which is still prevailing in the first quarter of the 20th century.

It is perhaps worth repeating once more. At any given time, as a result of expectations about likely demand for their products, firms will be offering a quantity of their output for sale on the market at prices they suppose will be acceptable to buyers and which will yield a reasonable return over and above all explicit costs (including the cost of borrowing). We refer to this total, newly produced output in the economy as aggregate production or *aggregate supply*.

At the same time, buyers (i.e., households, government, other firms and foreigners) will be in the market offering to take up this output. The extent to which they will be willing to purchase the output of the firms will depend upon their needs, their expectations about the future, the prices of the goods in relation to their income

and the prices of close substitutes, and, in the case of large purchases, the cost of borrowing. We refer to the total of this 'willingness to purchase' goods and services as *aggregate demand*.

In any given period, these two aggregates will not necessarily be equal. Any divergence between them will be accounted for in that period by an unintended change in firms' inventories and will precipitate adjustments in subsequent periods. Adjustments can take the form of production or price changes or both. For the moment, in the simple **aggregate demand–output model**, also called the *Keynesian model*, we assume that the response is an output one only.

Aggregate demand–output model
Also called the Keynesian model; the model that determines the equilibrium level of real GDP by equating output produced to aggregate demand.

To fix our ideas, let us consider a specific hypothetical situation. Imagine an economy operating at a particular point in time at its full-employment productive capacity of, say, $450 billion. What would happen if, in the next period, households were not willing to purchase all the output at the prices offered? Suppose aggregate demand turned out to be only $440 billion, perhaps because of emerging consumer pessimism, for example? Firms will find goods left on their shelves and in their warehouses and will have to make an *unplanned inventory investment* of $10 billion. They may be expected to respond to this unhappy state of affairs by subsequently laying off staff and reducing output. As a result of this process, the economy moves below its full-employment equilibrium.

In the above hypothetical example, it is reasonable to ask why the adjustment to the disequilibrium would be by way of reduced output and employment. Why not a price adjustment? This is what the classical economists believed would happen. In the classical world, prices and wages would drop so as to clear the product markets and given the lower wages, firms would be able to continue to employ the same numbers of people.

Keynes simply argued that this type of adjustment was just not the way he believed the real world worked. In reality, he believed prices – particularly the price of labour (wages) – tended to be resistant to downward adjustments. Thus, the adjustment could be expected to be on the output side rather than the price side. Furthermore, once the output is cut, people are laid off, and expectations about the future are curtailed, aggregate demand in subsequent periods may well be stuck for long periods at a level below the economy's full-employment production level, thereby resulting in continuing involuntary unemployment.

Now it's time to pause, take a breath and appreciate fully the point we have just reached. It is a powerful idea!

The basic explanation offered by Keynes for the Great Depression is as follows: Once equilibrium is established between aggregate demand and aggregate output, in contrast to what pre-Keynesian economists had believed, there may well be no tendency for the economy to change, even when that equilibrium is well below the full-employment output level.

The solution Keynes offered western economies facing the Great Depression was to increase aggregate demand through government spending until the full-employment equilibrium was reached. Otherwise, prolonged unemployment might persist indefinitely, and the economy might never self-correct.

We now turn to a critical idea behind changes in aggregate spending to stabilise the macro economy.

The spending multiplier effect

Changes in aggregate demand in the Keynesian model make things happen. The simple demand-oriented Keynesian model assumes that there is sufficient excess capacity in the economy to allow any increase in aggregate demand to be met by an increase in production. The existence of such excess capacity would undoubtedly have been a reasonable assumption in the severely recessed economy of the Great Depression that Keynes was witnessing in the 1930s, and would also be quite a reasonable assumption for the major western economies in the wake of the GFEC in 2008–09. However, it would also, more generally, be appropriate for any modern-day economy significantly below its full-employment level of economic activity. During the COVID-19 pandemic in 2020, most economies around the world operated below the full-employment level and capacity.

The crux of the Keynesian macroeconomic policy depends on a change in aggregate government expenditures or taxation, which is then multiplied or amplified by rounds of spending and re-spending throughout the economy. To understand this, you should note that the categories of aggregate demand outlined above (C, I, G and $(X - M)$) are not independent of each other.

For example, a stimulus to consumption demand (brought about perhaps by a generally more optimistic view of future economic conditions) could lead firms to expand production, increase their expenditure on plant and equipment (thereby increasing investment demand and providing a stimulus to those firms producing capital equipment) and hire more staff. These expenditures, in turn, would further stimulate consumption expenditure by virtue of the increased employment and income of the employees of the consumption-producing and capital-goods-producing firms. This would further stimulate investment, and so the process would continue.

Spending multiplier
The induced rounds of spending that occur in an economy after some initial stimulus to spending.

Thus, the original consumption demand stimulus would have a 'ripple effect' throughout the economy. The induced rounds of expenditure are often collectively referred to as the **spending multiplier**. The fact that the initial expenditure increases (or decreases, for that matter) are multiplied in their macroeconomic effect provides critical insight into Keynes's prescription for a severely recessed economy. In such an economy, private-sector demand needs to be stimulated. By increasing its spending on goods and services, the government can provide this stimulus, with the final stimulatory impact on the economy being much more significant than the initial increase in government spending.

The multiplier is important in the Keynesian model because it means that the initial change in aggregate expenditures results in an amplified change in the equilibrium level of real GDP. This is both good and bad news for an economy. The bad news occurs, for example, when the multiplier amplifies small declines in total spending from, say, consumer and business pessimism, into downturns in national output, income and employment. The good news is that, at least in theory, the macroeconomic policy can manage, or manipulate, the economy's performance by a relatively small initial change in aggregate expenditures.

In summary

- **Aggregate demand** is the sum of *C*, *I*, *G* and (*X* − *M*) and their economic determinants have been outlined.
- Classical economists believed a continuing recession was impossible: the operation of the price mechanism would eliminate persistent shortages or surpluses of products and labour.
- John Maynard Keynes believed insufficient aggregate demand could pull aggregate output away from full-employment equilibrium for long periods, resulting in long-term involuntary unemployment.
- The **spending multiplier** refers to the induced rounds of spending that occur in an economy after some initial stimulus to spending.

2 The aggregate demand–aggregate supply model

As explained above, Keynes's prescription for the Great Depression was simple: increased government spending would lead to an amplified impact on aggregate demand in the economy and jobs would be created. Although Keynes was not concerned with the problem of inflation, his theory has implications for fighting demand-pull inflation. In this case, the government would cut spending or increase taxes to reduce aggregate demand.

However, Keynes placed little or no analytical emphasis on aggregate supply considerations. To properly understand modern business cycles, the simple demand-oriented Keynesian framework outlined in the previous section needs to be augmented.

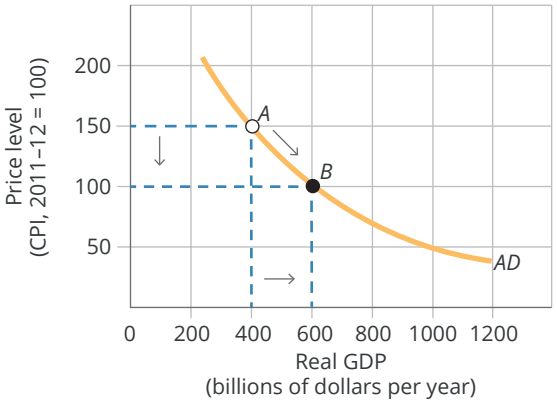
The discussion begins with a presentation of the aggregate demand curve and then the aggregate supply curve. Once these concepts are developed, the analysis shows why modern macroeconomics teaches that shifts in aggregate supply or aggregate demand can influence the price level (and thereby the rate of price inflation), the equilibrium level of real GDP, and employment.

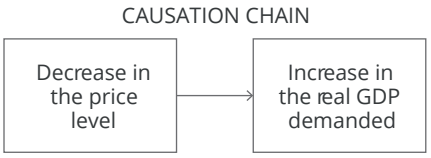
The aggregate demand curve

Here we are considering the collective economy-wide demand for *all* goods and services, rather than the *market* demand for a particular good or service. **Exhibit 14.1** shows the **aggregate demand (AD) curve**, which slopes downward. The aggregate demand curve shows the level of total real GDP that households, businesses, government and foreigners (net exports) would be willing to purchase at different possible average price levels during the time period, *ceteris paribus*. Just as we have seen with the demand curve for an individual product, other factors remaining constant, the lower the economy-wide price level, the greater the aggregate quantity demanded of real goods and services.

Aggregate demand (AD) curve
A curve showing the level of total real GDP that households, businesses, government and foreigners (net exports) would be willing to purchase at different possible average price levels during a time period, *ceteris paribus*.

Exhibit 14.1 The aggregate demand curve





The aggregate demand curve shows the relationship between the price index level and the level of real GDP, other things being equal. The lower the price level, the larger the GDP demanded by households, businesses, government and foreigners. If the price index level is 150 at point *A*, real GDP of $400 billion would be demanded. If the price index level is 100 at point *B*, real GDP demanded would be $600 billion.

The downward slope of the aggregate demand curve shows that people are willing to buy more goods and services at a lower average price level. While the horizontal axis in individual market supply and demand model measures *physical* units, such as a tonne of wheat, the horizontal axis used in the aggregate demand and supply model measures the value of all *final* goods and services included in real GDP. The horizontal axis, therefore, represents the quantity of aggregate production demanded, measured in some base year's dollars. The vertical axis is an *index* of the overall price level, such as the GDP deflator or the CPI rather than, say, the price per unit of a good. As shown in **Exhibit 14.1**, if the price level measured by the CPI is 150 at point *A*, real GDP of $400 billion would be demanded in, say, the year 2018. If the price level is 100 at point *B*, a real GDP of $600 billion would be demanded.

Although the aggregate demand curve looks like an individual product demand curve, these concepts are very different. As we move along a product demand curve, the price of other related products is assumed to be unchanging. However, when we deal with changes in the general or average price level in an economy, this assumption is meaningless because we are using a market basket measure for *all* goods and services. In other

words, when we move along the aggregate demand curve, it is the prices of all goods and services in the economy that we are assuming are changing.

Reasons for the aggregate demand curve's shape

As you will recall from **Chapter 3**, in the case of the demand curve for an individual product, the reason, the demand curve slopes downwards is that it is assumed that prices of the product's close substitutes remain unchanged. Thus, as the price of the product reduces, it becomes relatively cheaper compared with other similar products, and we, therefore, expect that *ceteris paribus*, more of it will be demanded.

The reasons for the downward slope of an aggregate demand curve are very different and include *the real balances* or *wealth effect*, *the interest-rate effect* and *the net exports effect*. Furthermore, the key to the real balances or wealth effect and the interest-rate effect is that, for the time period in question to which a particular *AD* curve is assumed to relate, the total nominal quantity of liquid financial assets in the economy is assumed to be fixed.

A liquid financial asset is one that can be converted easily – with little or no loss of nominal value – into a form that can be used to buy goods and services. Cash in your pocket is the most liquid, while a fixed-term deposit you have in your bank, or some shares you may have in a company, are much less liquid.

Real balances or wealth effect

Cash and bank deposits are just a couple of examples of financial assets whose real value (in terms of purchasing power) changes with the price level. If prices are falling, households are more willing and able to spend. Suppose you have $2000 in a cheque account with which to buy 10 weeks' worth of groceries. If prices fall by 20 per cent, $2000 will now buy enough groceries for 12 weeks. This rise in real wealth may make you more willing and able to purchase out of current income (perhaps a new iPhone?).

Real balances or wealth effect
The impact on total spending (real GDP) of the inverse relationship between the price level and the real value of financial assets with fixed nominal value.

The effect of a change in the price level on real consumption spending is called the **real balances or wealth effect**. The real balances or wealth effect is the impact on total spending (real GDP) caused by the inverse relationship between the price level and the real value of financial assets with fixed nominal value.

Interest-rate effect

Interest-rate effect
The impact on total spending (real GDP) of the direct relationship between the price level and the interest rate.

A second reason that the aggregate demand curve is downward-sloping involves the **interest-rate effect**. The interest-rate effect is the impact on total spending (real GDP) caused by the direct relationship between the price level and the interest rate. Again, remember that a critical assumption of the aggregate demand curve is that the supply of the total nominal quantity of liquid financial assets available for borrowing in the economy in the time period relevant to the *AD* curve remains fixed. A high price level means people must take more dollars from their wallets and cheque accounts in order to purchase goods and services that they may want. At a higher price level, the demand for borrowed funds to buy products will also increase and given a fixed available supply of nominal liquid financial assets, this will push up the cost of borrowing; that is, nominal interest rates will rise.

Rising interest rates discourage households from borrowing to purchase homes, cars, electronic equipment and other consumer products. Similarly, at higher interest rates, *ceteris paribus*, businesses cut investment projects because the higher cost of borrowing diminishes the likely profitability of these investments. Thus, assuming the availability of fixed nominal liquidity, an increase in the price level translates, through higher interest rates, into a reduced real GDP demanded.

Net exports effect

Whether Australian-made goods have lower prices than foreign goods is another important factor that impacts on the aggregate demand curve. A higher domestic price level tends to make Australian goods more expensive compared with foreign goods, and the quantity of imports rises because consumers substitute imported goods for domestic goods. An increase in the general price level in Australia will also imply higher production costs

of Australia's exports, reduce their profitability, and thereby tend to reduce Australian exports. Consequently, a rise in the domestic price level in an economy tends to increase imports and decrease exports, and thereby reduce the net exports component of the aggregate demand for real GDP.

This is the **net exports effect**. The net exports effect is the impact on total aggregate demand (for real GDP) caused by the inverse relationship between the price level and the net exports of an economy.

Net exports effect The impact on total aggregate demand (for real GDP) caused by the inverse relationship between the price level and the net exports of an economy.

Note that here we are assuming that the foreign exchange rate remains unchanged (the *ceteris paribus* assumption again). Thus, a particular *AD* curve is drawn not only with a fixed stock of available nominal liquid financial assets in play but also with a particular exchange rate in play.

Exhibit 14.2 summarises the three effects that explain why the aggregate demand curve in **Exhibit 14.1** is downward-sloping.

Exhibit 14.2 Why the aggregate demand curve is downward-sloping

Effect	Causation chain
Real balances or wealth effect	Price level decreases → Real purchasing power of financial assets with fixed nominal value rises → Consumers buy more products → Real GDP demanded increases
Interest-rate effect	Price level decreases → Increase in the real value of the fixed nominal supply of liquid financial assets → Nominal interest rates fall → Businesses and households borrow and buy more products → Real GDP demanded increases
Net exports effect	Price level decreases → Home country's goods become less expensive than foreign goods → Domestic residents and foreigners buy more home country products → Exports rise and imports fall → Real GDP demanded increases

Non-price-level determinants of aggregate demand

As was the case with individual demand curves, we must distinguish between *changes in real GDP demanded*, caused by changes in the price level, and *changes in aggregate demand*, caused by changes in one or more of the *non-price-level determinants*.

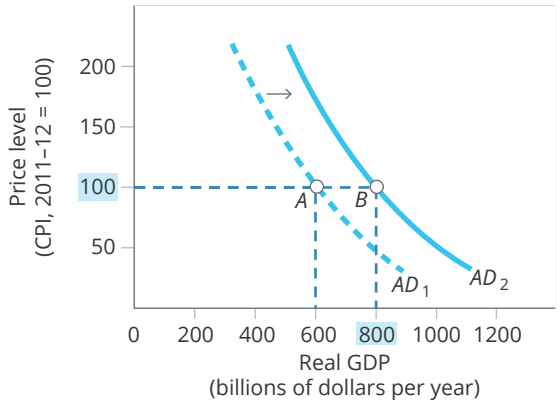
Once the *ceteris paribus* assumption is relaxed, changes in variables other than the price level would be represented by a change in the location of the aggregate demand curve. Thus, any non-price-level-related change in consumption (C), investment (I), government spending (G) and/or net exports ($X - M$) components of aggregate expenditures (refer again to the earlier initial discussion of aggregate expenditure components provided in the chapter) would all be represented by a shift in the aggregate demand curve.

Exhibit 14.3 illustrates the link between an increase in aggregate demand and an increase in aggregate expenditure demanded. Begin at point A on aggregate demand curve AD_1, with a price level of 100 and a real GDP of $600 billion. Assume that the price level remains constant at 100 and that there is then an increase in aggregate demand represented by a shift from AD_1 to AD_2. Consequently, the level of real GDP demanded rises from $600 billion (point A) to $800 billion (point B) at the price level of 100.

What might be the cause of such a shift in one or more of the components of aggregate demand? The cause might be that consumers have become more optimistic about the future or increase in consumer confidence and their demand for consumption expenditures (C) have risen. Another factor bringing about an increase in consumer demand would be an increase in the wealth of consumers unrelated to the general price level. This may happen, for instance, as a result of increases in house prices or significant rises in stock markets. More generally, any increase in the value of assets held by consumers can be expected to have a positive effect on their spending.

Alternatively, an increase in business optimism may have increased profit expectations, and the desired level of investment (I) has risen because businesses seek to spend more on plant and equipment. Another factor bringing about an increase in investment demand would be a decrease in the interest rates unrelated to the general price level.

Exhibit 14.3 **A shift in the aggregate demand curve**



CAUSATION CHAIN

Increase in non-price-level determinants: *C*, *I*, *G*, (*X* – *M*) → Shift out in the aggregate demand curve from AD_1 to AD_2

At the price level of 100, the level of real GDP demanded is $600 billion at point *A* on AD_1. An increase in one of the non-price-level determinants of consumption (*C*), investment (*I*), government spending (*G*) or net exports (*X* – *M*) causes the level of real GDP demanded to rise to $800 billion at point *B* on AD_2. Because this effect occurs at any price level, an increase in aggregate demand is represented by a shift in the *AD* curve rightward. Conversely, a decrease in aggregate demand is represented by a shift in the *AD* curve leftward.

The same increase in aggregate demand could also have been caused by a boost in autonomous government spending (*G*). Changes in government spending are considered as a policy variable and will be discussed in detail in **Chapter 17**.

Finally, a rise in net exports (*X* – *M*) may have been responsible. This is particularly relevant to an economy like Australia's, which is dependent to a great extent on international trade. In this case, the cause may be due to a boost to the economic activity of the country's trading partners, resulting in higher demand for Australian exports in those countries. Alternatively, such an increase in net exports could arise from a reduction in the international price of the country's currency; that is, a depreciation in its exchange rate. The reduced value of the currency means that imported products become more expensive in domestic currency terms, thus stimulating demand for import-competing domestic products. Also, export revenues in domestic currency terms increase, thereby stimulating increased export activity. Thus, a fall in the exchange rate can also lead to increased aggregate demand, represented by an outward shift in the *AD* curve. The determinants and economic effects of changes in a country's exchange rate are discussed in greater detail in **Chapter 18**.

Conversely, analogous factors to the above leading to reduced aggregate demand would be represented by inward or leftward shifts in the *AD* curve.

The aggregate supply curve

The aggregate supply (*AS*) curve is similar to the individual product supply curve in that – at least in the short run – it is also considered to be upward-sloping. However, the arguments in support of an upwardly sloping *product market supply* curve do not at all apply directly to the *aggregate supply* curve.

The supply curve for an individual product is upward-sloping because profit-maximising firms will be induced to bring more product to the market if the market price of their product rises. The same mechanism cannot be at work in the case of the aggregate supply curve because, as the general price level rises along the curve, all prices in the economy are assumed to be rising. This would include not only product selling prices but also the prices of inputs into the production process (including the cost of labour).

We will come back to this issue a little later but, for now, putting this issue to the side for this introduction to the concept, we define the **aggregate supply (AS) curve** as follows. The aggregate supply curve represents, *ceteris paribus*, the total dollar amount of goods and services – that is, real GDP – that firms would in aggregate seek to produce in an economy at various price levels during a particular time period.

Aggregate supply (AS) curve The curve that represents the level of real GDP that firms would be willing to produce at different possible price levels during a particular time period, *ceteris paribus*.

Given this general definition, in understanding the interaction between aggregate demand and aggregate supply – which will be discussed further below – it is useful to distinguish in principle three different ranges of the aggregate supply curve: the horizontal range, the intermediate range and the vertical range.

Three ranges of the aggregate supply curve

In this section, we will discuss the economics underpinning the two extreme cases of a horizontal and vertical characterisation of the aggregate supply curve. Then we will complete the discussion by adding in the intermediate range of the aggregate supply curve to arrive at what might be called an eclectic or general view of how the shape of the aggregate supply curve varies as real GDP expands or contracts.[5] The aggregate supply curve, *AS*, in **Exhibit 14.4** has three distinct ranges or segments, labelled: (1) **horizontal (or Keynesian) range**, (2) **intermediate range** and (3) **vertical (classical) range**.

The horizontal range is the horizontal segment of the aggregate supply curve, which represents an economy in a severe recession or depression as was the situation with which Keynes was particularly concerned in the 1930s. In **Exhibit 14.4**, real GDP Y_K the price level remains constant as the level of demand for real GDP rises. In a significantly recessed economy, prices, including wages, could be expected to exhibit stickiness in the face of increased aggregate demand. The economy will have many idle resources, and due to significantly unused production capacity, producers are willing to produce and sell additional output at the current prices.

Horizontal (or Keynesian) range The horizontal segment of the aggregate supply curve, which represents an economy in a severe recession.

The intermediate range is the upward-sloping segment of the aggregate supply curve, which represents an economy as it approaches full-employment output. Between Y_K and the full-employment output of Y_F, as the level of demand for real GDP rises, not only does production increase but the price level also rises moderately.

Intermediate range The upward-sloping segment of the aggregate supply curve, which represents an economy as it approaches full-employment output.

Finally, at Y_F, the level of real GDP remains constant, and only the price level rises with further increases in aggregate demand. This vertical segment of the aggregate supply curve represents an economy at the full-employment output. Increase in demand for aggregate output will increase factor prices and thereby the cost of production to the producers. The firms will respond by increasing their selling prices to maintain profit margins. As prices rise across the economy, the quantities of goods and services demanded will eventually drop back to the full-employment level of output. This represents the classical economists' view that the economy is self-correcting at the full-employment level of output.

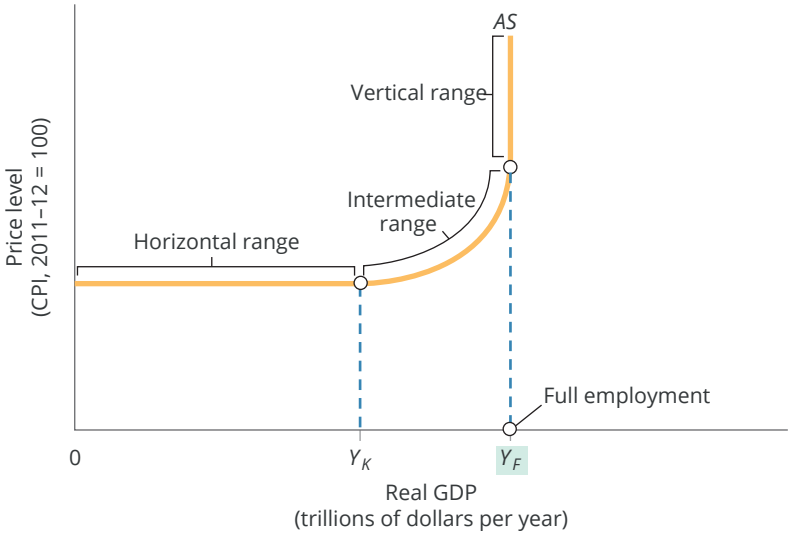
Vertical (classical) range The vertical segment of the aggregate supply curve, which represents an economy at full-employment output.

Aggregate demand and aggregate supply macroeconomic equilibrium

As in **Chapter 3**, where a short-run equilibrium occurred in an individual product market at the intersection of the product demand and supply curves, we also depict a short-run equilibrium state for the whole economy as the unique combination of price level and output level that equates how much people in aggregate are wanting to buy with the aggregate amount businesses want to produce and sell. In **Exhibit 14.6**, such a short-run *macroeconomic equilibrium* is depicted at point *E*, where the level of real GDP is $300 billion and the equilibrium price level is 100.

[5] This eclectic or general view of the *AS* curve is sometimes referred to as the 'Keynesian–Neoclassical Synthesis'.

Exhibit 14.4 **The three ranges of the aggregate supply curve**



The aggregate supply curve represents the relationship between the price level and the level of real GDP supplied. It consists of three distinct ranges: (1) a horizontal (sometimes known as Keynesian) range up to Y_K wherein the price level is constant for an economy in severe recession; (2) and intermediate range between Y_K and Y_F wherein both the price level and the level of real GDP vary as the economy approaches full-employment demand; and (3) vertical (sometimes known as the classical) range wherein price level varies, while the level of real GDP remains constant at the full-employment level of output, Y_F.

Analyse the issue

Applicable concepts: aggregate demand and aggregate supply analysis

Was Keynes right?

In *The General Theory of Employment, Interest and Money*, Keynes wrote:

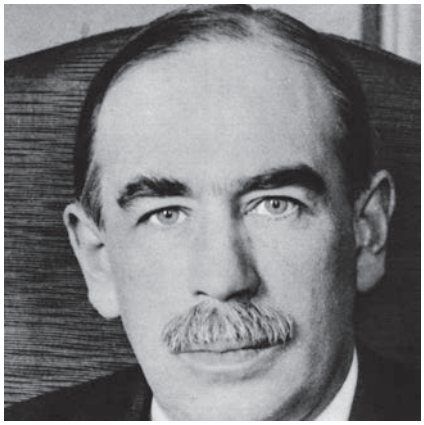
> The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist.[1]

John Maynard Keynes (1883–1946) is regarded as the father of modern macroeconomics. He was educated at Eton and Cambridge in mathematics and probability theory, but ultimately selected the field of economics and accepted a lectureship in economics at Cambridge. Keynes was a multi-faceted man. He was an honoured and supremely successful member of the British academic, financial and political upper class. For example, Keynes amassed a $2 million personal fortune (equivalent to more than $60 million in 2018 dollars – check by using CPI index numbers from 1945–2018) by speculating in stocks, international currencies and commodities. In addition to making a sizable fortune for himself, Keynes served as a trustee of King's College and built its endowment from £30 000 to £380 000.

Keynes was a prolific scholar who is best remembered for *The General Theory*, published in 1936. Keynes argued that in a severe recession, increasing aggregate demand would achieve full employment, while prices and wages remain inflexible. Moreover, his bold policy prescription was that the government raise its spending and/or reduce

taxes in order to increase the economy's aggregate demand and put the unemployed back to work.

Was Keynes correct? Even today, some 80-plus years later in our 21st-century world, this is still a very important and relevant question since many governments around the world (Australia included) deployed Keynesian-type aggregate demand stimulation as they struggled to ameliorate the impact of the 2008–09 'Great Recession'.



Alamy Stock Photo/ClassicStock

John Maynard Keynes

What do you think?

Plot the following data from **Exhibit 14.5** on the CPI (vertical axis) and real GDP (horizontal axis) for the US for different years following the onset of the Great Depression in 1929–30. Then draw in four separate parallel *AD* curves, one for each of the years in the table below with the relevant corresponding data point lying somewhere along its length. Does the plot seem approximately consistent with a flat *AS* curve and shifting *AD* curves over time, or an approximately vertical *AS* curve and shifting *AD* curves? Does your analysis lend support to Keynes's theory that, in a severely recessed economy, increases in aggregate demand propel an economy towards full employment without undue upward pressure on prices?

Exhibit 14.5 **Price level, real GDP and unemployment rate, 1933–41**

Year	CPI (1982–84 = 100)	Real GDP (billions of 1992 $)	Unemployment rate (%)
1933	13.0	577	24.9
1939	13.9	867	17.2
1940	14.0	941	14.6
1941	14.7	1102	9.9

Sources: US Department of Labor, *CPI Detailed Report: October 1998*, Table 24; *Survey of Current Business*, https://www.bea.doc.gov/bea/dn1.htm, Table 2A; *Economic Report of the President*, 1999, https://www.gpo.ucop.edu/catalog/erp99.html, Table B-31.

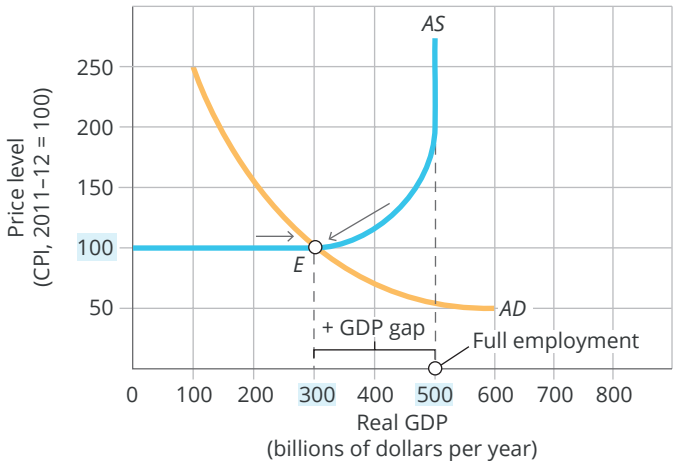
[1] J. M. Keynes, *The General Theory of Employment, Interest and Money*, Macmillan, London, 1936, p. 383.

Because the entire real GDP value of final products is bought and sold at the price level of 100, there is no upward or downward pressure for the macroeconomic equilibrium to change – at least in the short run, until one or more of the previously discussed determinants of aggregate demand and supply change. Such changes would then, of course, cause the macro economy to move away from its initial equilibrium state. Various examples of movements from one macroeconomic equilibrium to another are depicted and discussed in the following.

Note that the economy depicted in **Exhibit 14.6** is operating on the edge of the horizontal range, with a GDP shortfall of $200 billion compared with full-employment output.

To better understand the economic forces at work that lead the macro economy to gravitate towards a point like *E* in **Exhibit 14.6**, suppose for a moment that the level of production in the economy happens to be below $300 billion and that the *AD* curve remains unchanged, as in **Exhibit 14.6**. At a price level of 100, the real GDP demanded exceeds the real GDP supplied. Under such circumstances, businesses cannot fill orders quickly enough and inventories are drawn down unexpectedly. Business managers react by hiring more staff and producing more output, so that aggregate real GDP supplied would increase to equal aggregate real GDP demanded (point *E*). Because the economy is characterised as operating in the horizontal range of the aggregate supply curve, the entire response is in increased production, and the price level remains constant at 100.

Exhibit 14.6 **The aggregate demand and aggregate supply model**



Macroeconomic equilibrium occurs where the aggregate demand curve, *AD*, and the aggregate supply curve, *AS*, intersect. In this case, equilibrium, *E*, is located at the far end of the horizontal range, where the price level is 100, and the equilibrium output is $300 billion. In macroeconomic equilibrium, businesses neither overestimate nor underestimate the real GDP demanded at the prevailing price level.

Another scenario can be illustrated by supposing, for argument's sake, that the level of aggregate real GDP supplied was somewhere in the intermediate range on the *AS* curve between $300 billion and $500 billion and therefore exceeded the level of real GDP demanded. In this aggregate supply segment, the price level is between 100 and 200, and businesses face sales that are less than expected. In this case, unintended inventories of unsold goods pile up on the shelves and businesses will need to cut back on production and reduce prices, and may also need to lay off staff.

This adjustment process continues until the equilibrium price level and output level are reached at point *E*, and there is no upward or downward pressure for the price level to change. Here, the aggregated quantity of real GDP arising from the production decisions of sellers in the economy equals the aggregated outcome of the spending decisions of buyers during the particular given period of time.

In summary

- The **aggregate demand (*AD*) curve** represents the level of real GDP demanded in the economy at different price levels and slopes downward to the right due to the **real balances**, the **interest rate** and the **net exports effects**.
- The **aggregate supply (*AS*) curve** represents the level of real GDP that firms will be willing to supply at different price levels and can be thought of as having three segments: the horizontal segment, representing an economy significantly below full employment; a vertical segment representing an economy at the full-employment level of GDP; and an intermediate segment where the *AS* curve is upward sloping, representing an economy approaching full employment GDP.
- Macroeconomic equilibrium is represented in the model at the point of intersection of the *AD–AS* schedules.

3 Changes in the *AD–AS* macroeconomic equilibrium

In these final sections of the chapter the objective is to develop an understanding of how the *AD–AS* macroeconomic model can be used to understand the business cycle and how macroeconomic policy can affect prices, GDP and employment.

Equilibrium shifts arising from aggregate demand shifts

The economy might shift away from some short-run macroeconomic equilibrium due to changes in aggregate demand. The next step in our analysis is, therefore, to *shift* the aggregate demand curve along the three ranges of an assumed stable aggregate supply curve and observe and explain the impact on real GDP and the price level. As will be evident, as the macroeconomic equilibrium changes, the economy experiences more or fewer problems with inflation and unemployment.

Horizontal range

Keynes's *General Theory* offered a powerful solution to the Great Depression. Keynes perceived the economy as driven by aggregate demand and part (a) of **Exhibit 14.7** demonstrates this perspective. The range of real GDP below $300 billion is consistent with the Keynesian presumption of price and wage stickiness in such circumstances.

Assume the economy is in equilibrium at E_1, with a price level of 100 and a real GDP of $200 billion. In this case, the economy is in recession far below the full-employment GDP of $500 billion. The Keynesian prescription for responding to a recession is to increase aggregate demand until the economy achieves full employment. Because the aggregate supply curve is presumed horizontal in such circumstances, demand creates its own supply.

Suppose the increase in aggregate demand is represented by a rightward shift in the AD curve from AD_1 to AD_2 and a new equilibrium is established at E_2. Even at the higher real GDP level of $300 billion, the price level remains at 100. Stated differently, the aggregate output can expand throughout this range without raising prices. This is because, in the horizontal range, substantial idle production capacity (including spare plant and unemployed workers competing for available jobs) can be put to work at existing prices.

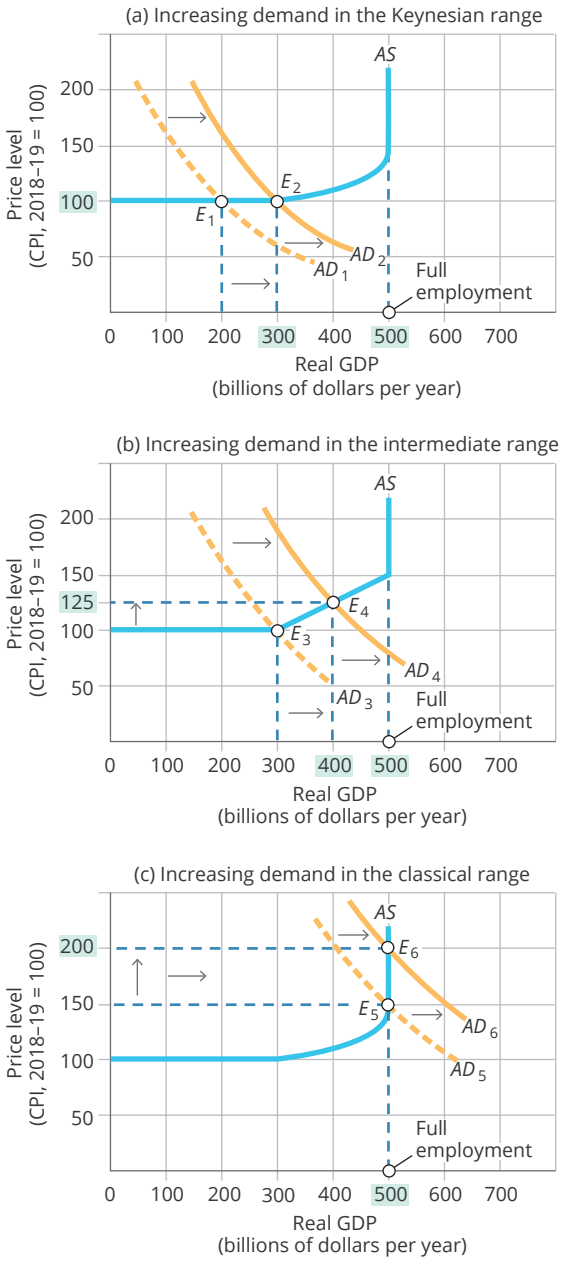
Intermediate range

The intermediate range in part (b) of **Exhibit 14.7** is between $300 billion and $500 billion worth of real GDP. As output increases in the range of the AS curve near the full-employment level of output, the considerable slack in the economy disappears. Assume an economy is initially in equilibrium at E_3 and aggregate demand increases as represented in the exhibit by a shift in the AD curve from AD_3 to AD_4. As a result, the level of real GDP rises from $300 billion to $400 billion but the price level also rises from 100 to 125. In this output range, several factors contribute to inflating prices.

First, *bottlenecks* (obstacles to increased output production) develop when some firms have no unused capacity and other firms operate below full capacity. Suppose the steel industry is operating at full capacity and cannot fill all of its orders for steel. An inadequate supply of one resource, such as steel, can hold up car production even when the car-making industry is well below capacity. Consequently, the bottleneck may cause steel firms to raise the price of steel and, in turn, car manufacturers may then need to raise the price of cars.

Second, a shortage of certain labour skills while firms are earning higher profits may lead skilled workers to exert their market power to obtain sizeable wage increases, so businesses raise prices to offset these production cost increases. Wage demands are more difficult to reject when the economy is prospering because businesses

Exhibit 14.7 Effects of increases in aggregate demand



The effect of an increase in aggregate demand on price and output levels depends on the range of the aggregate supply curve in which it occurs. In part (a), an increase in aggregate demand causing the equilibrium to change from E_1 to E_2 in the horizontal range will increase real GDP from $200 billion to $300 billion, but the price level will remain unchanged at 100.

In part (b), an increase in aggregate demand causing the equilibrium to change from E_3 to E_4 in the intermediate range will increase real GDP from $300 billion to $400 billion but the price level will also rise from 100 to 125.

In part (c), an increase in aggregate demand causing the equilibrium to change from E_5 to E_6 in the vertical range will increase the price level from 150 to 200, but real GDP will not increase beyond the full-employment level of $500 billion.

will want to hold on to good employees who may consider leaving because of attractive alternative employment opportunities. Besides, in these economic circumstances, businesses believe the higher wage costs can be passed on quite easily in the form of higher product prices because consumers will be experiencing rising incomes as output expands to near full capacity.

Third, as the economy approaches full employment, firms must use less-efficient productive capacity (including possibly less-productive employees). This less-efficient operation creates higher production costs, which may also then be passed on to consumers in the form of higher prices.

Vertical range

While inflating prices resulting from an increase in aggregate demand is no issue in the horizontal range and only a moderate consequence in the intermediate range, it becomes a very significant outcome in the vertical range of the *AS* curve.

Assume the economy is shown in part (c) of **Exhibit 14.7** is in equilibrium at E_5, where aggregate demand is just equal to full-capacity output production in the economy (represented in the exhibit by the intersection of AD_5 and the *AS* curve at the full-employment level of production). Now suppose aggregate demand in the economy increases – as a result, perhaps, of excessively easy monetary policy on the part of the central bank (more on monetary policy in **Chapters 15 and 16**). This increase in aggregate demand would be represented by a rightward shift in the *AD* curve from AD_5 to AD_6.

Because *AS* is vertical at $500 billion, this increase in aggregate demand boosts the price level from 150 to 200, but fails to expand real GDP. The explanation is that, once the economy is operating at full capacity, businesses raise their prices in order to ration fully employed productive capacity to those willing to pay the highest prices.

Equilibrium shifts arising from aggregate supply shifts

Our discussion so far has explained changes in real GDP supplied resulting from changes in aggregate demand, given unchanging aggregate supply in the economy. Now we consider a situation of unchanging aggregate demand in the economy and changes in the economy's aggregate supply caused by changes in one or more of its non-price-level determinants. Such factors affecting aggregate supply include resource costs (domestic and imported), technological change, taxes and/or subsidies and regulations.

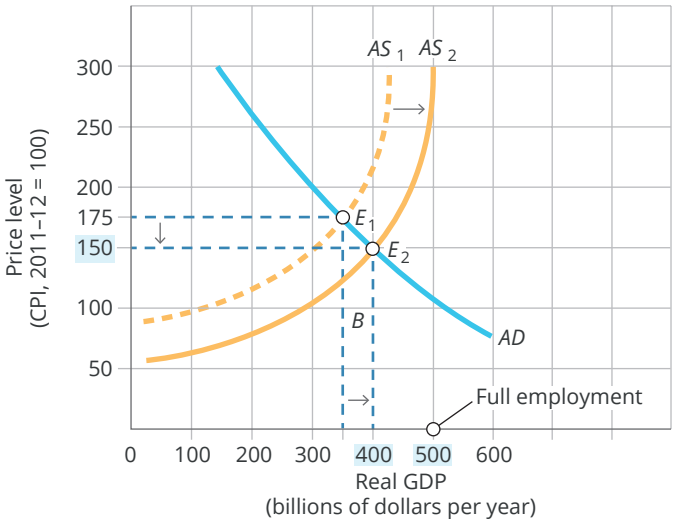
Note that each of these factors affects production costs. At a given price index level, the profit businesses make at any level of real GDP depends on production costs. If costs change, firms respond by changing their output or prices (or both). Lower production costs increase the willingness of firms to increase production levels for any set of selling prices. New production technology can reduce costs, as can higher labour productivity of a firm's workforce, as well as reduced government regulation (eliminating so-called 'red tape' – and, more recently, 'green tape'), which reduces a firm's compliance costs). Lower energy costs would also reduce production costs. For example, the recent use of 'fracking' to cheaply access vast quantities of shale oil and gas in the US has quite significantly reduced energy costs of US firms. All of these situations would be represented by a shift in the aggregate supply curve rightward, indicating that greater real GDP will be supplied by firms at all different possible price levels.

Conversely, higher production costs would be expected to lead to decreased production levels for any given set of selling prices. Increased government regulatory compliance costs, or wage rises far in excess of labour productivity (as happened in Australia in the 1970s), or raw materials and/or energy cost increases would all lead to increased production costs and would be represented by a shift in the aggregate supply curve leftward, meaning that less real GDP would be supplied at all price levels.

Exhibit 14.8 represents a supply-side explanation of the business cycle, in contrast to the demand-side case presented in **Exhibit 14.7**. (Note that for simplicity of exposition the supply curve is drawn using only the intermediate segment.) The economy begins in equilibrium at point E_1, with real GDP at $350 billion and the price level at 175. Then suppose better employer–employee industrial relations cause increased labour productivity.

This reduces the labour costs of producing output. With lower labour costs per unit of output, businesses seek to increase profits by expanding production at any price level.

Exhibit 14.8 **A rightward shift in the aggregate supply curve**



CAUSATION CHAIN

Change in one or more non-price-level determinants: resource prices, technological change, taxes, subsidies and regulations → Shift in aggregate supply in the economy, represented by a shift in the aggregate supply curve

Holding the aggregate demand schedule unchanged, the impact on the price level and real GDP depends on whether the aggregate supply curve shifts to the right or the left. A rightward shift of aggregate supply from AS_1 to AS_2 will increase the real GDP from $350 billion to $400 billion and reduce the price level from 175 to 150.

This would be represented by a shift in the aggregate supply curve rightward from AS_1 to AS_2, with equilibrium changing from E_1 to E_2. As a result, the economy's real GDP increases $50 billion and the price level *decreases* from 175 to 150.

Changes in many other things can also cause an increase in aggregate supply. Lower oil prices, better management practices, improvements in basic infrastructure inputs into the production process (microeconomic reform), lower taxes and reduced government regulation are just a few other examples of developments that lower production costs and therefore cause an increase in aggregate supply, represented by a rightward shift of the aggregate supply curve.

What kinds of events might raise production costs and be represented by a shift in the aggregate supply curve leftward? Perhaps there is war in the Middle East, or the Organization of the Petroleum Exporting Countries (OPEC) restricts supplies of oil and higher energy prices spread throughout the economy. Under such a 'supply shock', businesses would decrease their production of goods and services at any price level in response to higher production costs per unit.

Similarly, larger-than-expected wage increases (in excess of labour productivity rises), higher taxes to protect the environment (see **Exhibit 4.6** in **Chapter 4**), or greater unnecessary government regulation might increase production costs and therefore reduce aggregate supply in the economy. This would be represented by a leftward shift in the aggregate supply curve, a shift that is discussed further in the next section.

Applicable concept: applying the AD–AS model

Global perspective

The impact of the GFEC on the US economy

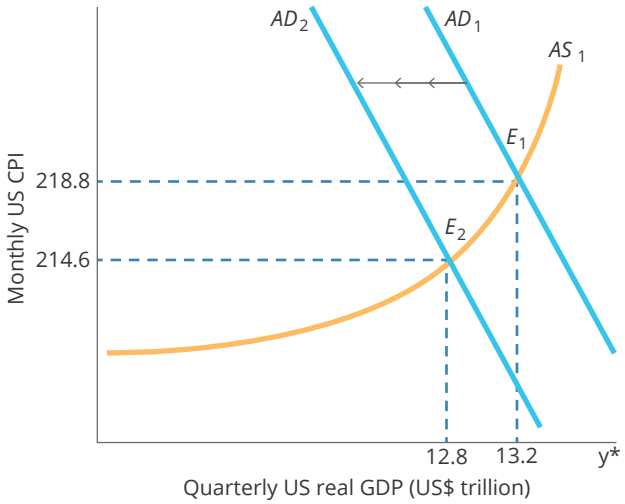
The US economy officially reached its business cycle peak in December 2007. The US share market peaked in October 2007. By September 2008, the S&P 500 (the US share market index) had fallen 25 per cent, and by February 2009 it had fallen a whopping 53 per cent! For those many millions of Americans and others around the world who held significant proportions of their wealth in US shares this amounted to a significant cut in wealth. And, of course, simultaneously, all the world's share markets were experiencing similar dramatic drops. *How might such significant share market drops have impacted on economic activity?*

Earlier it was explained that wealth is a significant non-income determinant of consumption spending and that any substantial reduction in people's wealth, like a share market fall, even without any significant reduction in their income, could be expected to impact negatively on their consumption. In addition, the uncertainty created by such significant and rapid reductions in share markets invariably leads people – consumers and companies alike – to become more worried and pessimistic about the immediate future. This also leads to people curtailing their spending. These first-round spending effects are amplified through the economy by the action of the spending multiplier explained earlier.

Subsequently, with reduced spending becoming evident, firms begin to cut back their production and, to a certain extent, the quantity of labour they employ, either through reduced hours or through lay-offs. This then leads to reduced income in people's pockets, resulting in even less spending in subsequent periods. In this way, private sector spending, income and production can begin a vicious spiral downwards. In our *AD–AS* model all of these impacts would be represented by an inward shift of the *AD* curve.

In **Exhibit 14.9** we represent the US economy at different times during the GFEC and begin with the economy in, say, the December quarter, 2008, being represented by AD_1 and AS_1. In that quarter, US real GDP was US$13.2 trillion (seasonally adjusted) with a price level – as represented by the December month CPI (average of US cities, seasonally adjusted) – of 218.8. This is depicted in the exhibit as E_1.

Exhibit 14.9 **Illustrating the impact of the GFEC on the US economy: applying the *AD–AS* framework**



As argued above, the significant reduction in aggregate demand triggered, *inter alia*, by the dramatically falling share market throughout 2008 and early 2009 is represented in the exhibit by the inward shift in the *AD* curve from AD_1 to AD_2, representing the state of aggregate demand in the US economy in, say, the June quarter of 2009. In this

quarter, as represented by E_2 in the exhibit, US real GDP had fallen to US$12.8 trillion (seasonally adjusted) and the CPI (seasonally adjusted) in the June month of 2009 stood at 214.6.

At E_2, the resulting significant reduction in aggregate demand had led to a significant fall in GDP, along with a drop in the CPI, and the US was well and truly in the grip of a very severe economic downturn! As it turned out, the June quarter, 2009, was the nadir for US GDP in the GFEC, and from the following quarter, the economy began to grow again.

While we have used the US as the focus here, the same arguments as traced above would be applicable to most of the developed. and developing, world during 2008–09. What began as a financial market crisis very rapidly turned into the very severe synchronised international recession of 2008–09. In **Chapter 19** we will delve into the anatomy of the financial crisis that triggered the subsequent economic downturn, as well as explaining – using the *AD–AS* framework of this chapter – the economics of the policy responses employed by authorities around the world in their attempts to ameliorate the impact of the ensuing international recession on their national economies.

Source: Data for this Global Perspectives box came from Economagic, http://www.economagic.com.

What do you think?

In response to the GFEC the US central bank very quickly and dramatically eased monetary policy and reduced short term interest rates to record lows. Depict this response using the *AD–AS* framework developed in this chapter.

Cost-push and demand-pull inflation revisited

We now apply the aggregate demand and aggregate supply model to the two types of inflation introduced in the previous chapter. In terms of our macro model, a cost-push inflation situation would be represented as the rise in the price level resulting from an inward shift in the aggregate supply curve while the aggregate demand curve remains unchanged. For instance, from the mid-1970s to the early 1980s, many developed economies experienced **stagflation**. Stagflation occurs when an economy experiences the twin maladies of high unemployment and rapidly rising prices simultaneously. How could this happen? It has been argued that the dramatic increase in the price of oil in 1973–74 was a significant factor in a *cost-push inflation* scenario in many countries around the world at the time.

Stagflation
The condition that occurs when an economy experiences the twin maladies of increasing unemployment and rising inflation simultaneously.

As a result of cost-push inflation, our *AD–AS* supply model predicts not only that the rate of inflation would increase but also that real output and employment would decrease – this is quite consistent with the mid-1970s experience resulting from the very significant oil price shocks that occurred during that decade.

On the other hand, an outward shift in the aggregate demand curve can be used to understand a situation of *demand-pull inflation*. In terms of our macro model, a demand-pull inflation situation would be represented as the rise in price level resulting from an increase in aggregate demand while aggregate supply in the economy remains unchanged.

In summary, an economy's aggregate supply and aggregate demand will shift in different directions for various reasons over different time periods. These shifts in aggregate supply and aggregate demand – and their representations within the *AD–AS* model developed in this chapter – may be used to understand upswings and downswings in real GDP and the rate of inflation; that is, the business cycle.

A contraction in either aggregate demand or aggregate supply – represented by a leftward shift in the aggregate demand curve or aggregate supply curve, respectively – can cause a recession. On the other hand, increased aggregate demand or more favourable aggregate supply conditions – represented by a rightward shift of the aggregate demand curve or aggregate supply curve, respectively – can cause real GDP and employment to rise and the economy recovers.

Exhibit 14.10 summarises the shift factors of aggregate demand and supply for further study and review. In **Chapter 17** on fiscal policy you will learn in more detail how changes in the taxation and spending policies of

government can affect the economy. The *AD–AS* framework you have learnt here will be used again there. Also, in **Chapter 16** on monetary policy, you will learn how the *AD–AS* model can help you understand how changing the supply of financial liquidity in the economy can affect the business cycle.

A couple of final words are perhaps in order. First, it is important to be aware that controversy continues over the shape of the aggregate supply curve. Without going into the details, modern-day classical economists – called New Classicists – believe the entire short-run aggregate supply curve is quite steep, with the predicted result that policy-originated aggregate demand increases (of the Keynesian type) will not do much to stimulate increased equilibrium output. On the other hand, New Keynesian economists contend that the short-run aggregate supply curve is much flatter, with the predicted result that policy-originated aggregate demand increases can be effective in stimulating increases in production and employment.

Second, in seeking to avert or at least alleviate a recession, a potential result of government-sourced increases in aggregate demand can be higher budget deficits if the government borrows to fund its increased spending, with higher levels of government debt the result. If such deficits continue for many years, the resulting higher levels of government debt – and interest payments to that – *may* become, such as to necessitate longer-term policy adjustments to ensure government debt never reaches 'unsustainable' levels. We will return to this issue again in the ensuing chapters.

Exhibit 14.10 **Summary of the shift factors of aggregate demand and aggregate supply**

Shift factors for aggregate demand (total spending)	Shift factors for aggregate supply (total production)
1 Non-price factors driving Consumption (*C*) (as just one example, changes in financial wealth)	1 Resource prices (domestic and imported)
2 Non-price factors driving Investment (*I*) (as just one example, changes in profit expectations)	2 Taxes/subsidies
3 Autonomous changes in Government spending (*G*)	3 Technological change
4 Non-price factors affecting Net exports (*X* − *M*) (as just one example, changes in the exchange rate)	4 Regulation

You're the economist

What would the reimposition of an economy-wide carbon abatement tax be likely to do to the rate of inflation, real GDP and employment?

In 2020, the International Monetary fund (IMF) has officially prescribed a carbon tax as the best way for governments around the world to meet their targets. However, in 2013, the Coalition was elected to the federal government in Australia with a critical policy promise to repeal the recently introduced (by the previous Labor Party government) economy-wide carbon pollution reduction tax (CPRT). The argument for its removal was that its removal would decrease business costs, put downward pressure on the cost of living, and improve economic activity and employment. Using the *AD–AS* framework, explain how the re-imposition of such a tax might impact on inflation, economic activity and employment.

In summary

- The *AD–AS* model is a powerful tool to understand the business cycle and policy responses to changing macroeconomic circumstances.
- The impact on the economy of changes in **aggregate demand** depends on the degree of excess productive capacity available in the economy, and this can be represented by the three ranges of the *AS* curve.
- Changes in aggregate supply also affect the business cycle. Any changes that bring about economy-wide changes to the costs of production will shift aggregate supply.
- The business cycle then is the result of relative shifts over time in the state of the economy's aggregate demand and aggregate supply as depicted by shifts in the aggregate demand (*AD*) and aggregate supply (*AS*) curves in the *AD–AS* model.

Key concepts

Consumption demand
Marginal propensity to consume (MPC)
Marginal propensity to save (MPS)
Investment demand
Autonomous expenditure
Aggregate demand (*AD*)
Aggregate demand–output model
Spending multiplier
Aggregate demand (*AD*) curve
Real balances or wealth effect
Interest-rate effect
Net exports effect
Aggregate supply (*AS*) curve
Horizontal range
Intermediate range
Vertical range
Cost-push inflation
Stagflation
Demand-pull inflation

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand the economic determinants of the four key expenditure components of aggregate demand in the economy

- **Aggregate demand** is the sum of *C*, *I*, *G* and (*X* – *M*) and their economic determinants are explained in the early sections of the chapter.
- In Keynesian economics, insufficient aggregate demand can pull aggregate output away from full-employment equilibrium in the economy for long periods resulting in long-term involuntary unemployment.
- In stark contrast, the classical economists believed that a continuing recession was impossible because private markets and the operation of the price mechanism could be relied upon to eliminate persistent shortages or surpluses of products and labour.
- The **spending multiplier** refers to the induced rounds of spending that occur in an economy after some initial spending stimulus. If government spending was increased – as Keynes proposed to combat a recession – the full stimulus to aggregate spending would be much greater due to the spending multiplier.

2. Develop an understanding of the aggregate demand–aggregate supply (*AD–AS*) macroeconomic model

- The **aggregate demand (AD) curve** represents the level of real GDP demanded in the economy at different price levels. It slopes downward due to the **real balances or wealth effect**, the **interest rate effect** and the **net exports effect**.
- The **aggregate supply (AS) curve** represents the level of real GDP firms will be willing to supply at different price levels. The *AS* curve can be thought of as having three segments:
  - The horizontal AS curve, which represents an economy significantly below full employment and, in this segment, the only effects of an increase in aggregate demand are increases in real GDP and employment while the price level does not change. Stated simply, in this segment 'demand creates its own supply'.
  - A vertical AS curve, which represents an economy at the full-employment level of GDP and the only effect of any further increase in aggregate demand is a change in the price level with no lasting change in production.
  - An intermediate segment, where the AS curve is upward sloping and in this range any increase in *AD* will result in a combination of increases in real GDP supplied and higher prices.
- *Macroeconomic equilibrium* in the economy is represented in the model at the point of intersection of the *AD–AS* schedules where suppliers neither overestimate nor underestimate the real GDP demanded at the prevailing price level.

3. Appreciate how the *AD–AS* macroeconomic model can be used to understand the business cycle and how macroeconomic policy can affect prices, GDP and employment

- The *AD–AS* model is a powerful tool to understand the business cycle and also to understand and represent policy responses to changing macroeconomic circumstances.
- The impact of changes in aggregate demand depends on the degree of excess productive capacity available in the economy, and this can be represented by the three ranges of the AS curve:
  - As aggregate demand increases in the *horizontal AS range* (very significant excess productive capacity in the economy), the price level remains constant as real GDP expands.
  - In the *intermediate AS range* (reducing excess capacity and supply bottlenecks beginning to impact), increases in aggregate demand increase real GDP and produce moderate increases in the price level.
  - Once the economy reaches full-employment output – as represented by the **vertical range** of the *AS* curve – additional increases in aggregate demand merely cause price inflation, rather than additional real GDP.
- Changes in aggregate supply also affect the business cycle. Any changes that bring about economy-wide changes to the costs of production will shift aggregate supply; examples are listed in **Exhibit 14.10**. The business cycle then is the result of relative shifts over time in the state of the economy's aggregate demand and aggregate supply as depicted by shifts in the aggregate demand (*AD*) and aggregate supply (*AS*) curves in the *AD–AS* model. Macroeconomic policy responses can also be represented in the *AD–AS* framework.

Study questions and problems

1 Explain how each of the following affects consumption demand.
- **a** The expectation is that the economy is heading into a recession next year.
- **b** Stock prices rise sharply.
- **c** The interest rate on consumer loans falls sharply.
- **d** Income taxes increase.

2 Explain how an increase in the value of the local currency relative to the foreign currency affects the net exports.

3 Explain the three reasons why the aggregate demand curve is downward sloping. How is your explanation different from the explanation for a downward-sloping demand curve for an individual product?

4 How would you shift the aggregate demand curve to represent each of the following changes in economic conditions? Explain your answers.
- **a** Business sentiment turns negative.
- **b** A new government is elected on the platform of increased government spending on infrastructure.
- **c** In order to stimulate the economy, which is verging on recession, the federal government increases spending on a range of government payments to families, individuals and businesses.
- **d** Australia's exchange rate depreciates (decreases in value relative to other currencies) significantly for an extended period.

5 Explain how spending multiplier works with an increase in consumption spending in the economy.

6 Identify the three ranges of the aggregate supply curve. Explain the impact of an increase in aggregate demand in each segment.

7 Consider the statement: 'Equilibrium real GDP is the same as full employment real GDP'. Do you agree or disagree? Explain.

8 How would you shift the aggregate supply curve to represent each of the following changes in economic conditions? Explain your answers.
- **a** The price of oil decreases due to global oil supplies being increased by the rapid growth of shale oil fracking in the US.
- **b** Electricity companies switch some generation capacity to more expensive renewable energy sources and the average price of electricity increases.
- **c** Due to less than expected wage demands, businesses find their production costs are lower than expected.
- **d** The federal government moves to introduce a range of 'red tape' and 'green tape' business regulations, which will increase the compliance cost of firms.

9 Assume an economy operates in the intermediate range of its aggregate supply curve. State the direction of effect on aggregate demand for each of the following changes in conditions. What is the effect on the price level? On real GDP?
- **a** Consumer wealth increases due to rises in the share and property markets.
- **b** A new strategic pact with a group of neighbouring nations is struck, which the federal government judges will allow it to spend significantly less on defence for a few years.
- **c** A severe recession occurs in a country that has been a major importer of the nation's exports.
- **d** The federal government reduces income taxes.

10 Illustrate demand-pull inflation and its likely impact on the price level and real GDP using aggregate demand and aggregate supply analysis.

Answers to 'You're the economist'

What would the reimposition of an economy-wide carbon abatement tax be likely to do to the rate of inflation, real GDP and employment?

Other things being equal, the reimposition of such a tax would increase the costs of production, especially for those production processes that emit significant amounts of carbon. Using the *AD–AS* framework, if you said that this would be represented by a leftward shift in the *AS* curve, with the prediction that inflation would be likely to rise, real GDP decrease and employment fall, then YOU ARE THINKING LIKE AN ECONOMIST.

Of course, one must also remember the reason for the impost of the tax in the first place; namely, to reduce the quantum of Australia's carbon emissions, an objective that a great many Australians support. The idea was that the tax would either make relatively heavy emitters pay for their carbon pollution (via the payment of the tax), forcing them to

raise their relative prices, and thereby reducing the demand for their products or forcing them to invest in alternative production techniques that emitted less carbon (but which were most probably more costly).

The Australian government stated that it was also committed to significantly reducing Australia's carbon emissions but by using what it has termed a 'direct action' approach – which entails removing the economy-wide carbon tax, with the government instead of making payments to companies that reduce their emissions. If you appreciate this is undoubtedly a very important, complex and vexing concern facing Australia (along with all other developed and developing nations) then YOU ARE ALSO THINKING LIKE AN ECONOMIST.

What is your MPC?

Early in your career, when your income is relatively low, you are likely to spend your entire income and perhaps even dissave, just to afford necessities. During this stage of your life, your MPC will be close to 1. As your income increases and you have purchased the necessities, additional income can go to luxuries. If you become wealthier, you have a higher marginal propensity to save and consequently a lower marginal propensity to consume. If you said your MPC will probably decrease as your income increases, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

Note: In some questions, more than one alternative may be correct.

1 Marginal propensity to consume can be explained as:
- **a** consumption spending divided by the income.
- **b** income earned divided by the consumption spending.
- **c** change in consumption spending resulting from disposable income.
- **d** change in consumption spending resulting from a change in disposable income.

2 Which of the following changes produces a decrease in investment demand?
- **a** A wave of optimism about future profitability.
- **b** A new wave of technological change.
- **c** An increase in business taxes.
- **d** High levels of plant capacity utilisation.

3 The aggregate demand curve is defined as:
- **a** the horizontal sum of individuals' demand curves for a particular product.
- **b** the real GDP that would be sought to be purchased by the private and public sectors at different possible aggregate price levels.
- **c** the total level of gross national expenditure (or $C + I + G$).
- **d** the aggregate of spending by both the private and public sectors during a period.

4 The impact on total spending (real GDP) caused by the inverse relationship between the price level and the real value of financial assets with fixed nominal value is called the:
- **a** aggregate price-level change effect.
- **b** real balances or wealth effect.
- **c** the financial asset effect.
- **d** the nominal interest-rate effect.

5 Which of the following changes produces an upward (or outward) shift in consumption demand?
- **a** An increase in consumer wealth.
- **b** A decrease in consumer wealth.
- **c** A decrease in disposable income.
- **d** A rise in interest rates.

6 In terms of the slope of the aggregate demand curve, the net exports effect is the inverse relationship between net exports and the _____ of an economy.
- **a** GDP deflator
- **b** price level
- **c** real GDP
- **d** imports

7 Which of the following would be represented by a shift in the aggregate demand curve to the right?
- **a** Consumers become more optimistic about the future.
- **b** Australia's terms of trade reduce.
- **c** Business pessimism increases.
- **d** The Australian dollar appreciates.

8 Along the horizontal range of the aggregate supply curve, an increase in aggregate demand will:
- **a** increase both the price level and real GDP.
- **b** increase real GDP but have no impact on the price level.
- **c** increase the price level only.
- **d** increase interest rates and reduce real GDP.

9 Other factors held constant, an increase in labour productivity will be represented by a shift in the aggregate:
- **a** demand curve leftward.
- **b** supply curve rightward.
- **c** demand curve rightward.
- **d** supply curve leftward.

10 Which of the following factors could lead to a decrease in equilibrium price level and an increase in equilibrium real GDP?
- **a** Increase in oil prices.
- **b** Decrease in real wealth.
- **c** Decrease in input costs.
- **d** Increase in exchange rate.

11 An increase in the price level caused by a set of economic forces represented by an outward shift of the aggregate demand curve is called:
- **a** supply shock inflation.
- **b** demand shock inflation.
- **c** cost-push inflation.
- **d** demand-pull inflation.

12 Assuming an economy can be represented by an unchanging aggregate demand curve, a rise in the costs of production on account of raw materials price rises would be represented by:
- **a** a rightward shift in the aggregate supply curve, causing an increase in the price level and an increase in real GDP.
- **b** a rightward shift in the aggregate supply curve, causing a decrease in the price level and an increase in real GDP.
- **c** a leftward shift in the aggregate supply curve, causing an increase in the price level and a drop in real GDP.
- **d** no change to output and prices since aggregate demand has been assumed to remain unchanged.

The monetary and financial system

15

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Learn about some monetary fundamentals.
2 Understand the determination of interest rates.
3 Develop an appreciation of the key elements of a modern financial system.

Shutterstock.com/Bankoo

Recall the circular flow model presented in **Chapter 11** on GDP. Households exchange *money* for goods and services in the product markets and firms exchange *money* for resources in the factor markets. In short, money critically affects the way an economy works. In the previous chapter on aggregate demand and supply, the *AD–AS* model was developed without really discussing the role of money. In this chapter and the next, money and, more generally, the monetary and financial system, take centre stage.

But precisely what is 'money'? As we will see, 'money' can come in many different forms. We begin our discussion with the three functions that money serves and the four properties that money should possess. We will also talk about what determines the 'demand for money'. Next, we define the components of four different money-supply definitions. The remainder of the chapter describes the organisation and function of Australia's central bank, the Reserve Bank of Australia (RBA), the role of banks, and some other essential aspects of the financial system.

Then, in the next chapter, you will build on the material in this chapter and learn in some detail how the RBA implements monetary policy in Australia by closely influencing the prevailing level of interest rates in the economy, thereby affecting total spending, output, employment and prices. All this will be explained using the *AD–AS* model introduced in the last chapter.

1 Some monetary fundamentals

Before proceeding, we should probably be clear about what we actually mean when we use the word 'money'. In everyday language, you'll probably hear someone being described as 'having a lot of money', meaning they are pretty wealthy, or that they earn a high income. A person's wealth can consist of a whole range of financial and non-financial assets that they own. Non-financial assets might include such things as houses, land, cars, furniture, paintings, and so forth. These are also known as real or physical assets. Financial assets a person might have could

include such things as company shares, government and corporate bonds, superannuation holdings, managed funds, along with bank deposits and the currency they choose to keep on hand.

As you will see in more detail later, when economists talk about 'money' we mean the most liquid of financial assets – like currency and a range of bank and other deposits – that people can use fairly readily to buy goods and services and settle debts. A person might have a great deal of wealth – and income – but choose to hold only a relatively small amount of their total wealth in the form of money. So, from this point on, when you read the word 'money', don't just think about currency (notes and coins), but instead think of a range of highly liquid financial assets that can be pretty readily used to buy goods and services immediately.

Barter
The direct exchange of one good for another good, rather than for money.

Now, back to our core discussion. Money facilitates the exchange of goods and services, but such exchanges can also occur without money using a trading system called **barter**. Barter is the direct exchange of one good for another good, rather than for money. The problem with barter is that it requires a *coincidence of wants*. In other words, a seller of a product who wants to buy another particular product needs to search for a buyer of their product who also just so happens to want to buy what they are selling in exchange. Alternatively, they may have to engage in a series of intermediate exchanges before finally acquiring the product they are after. This can create a considerable amount of wasted time and effort, which could otherwise be used more productively.

Now, it is undoubtedly true that the advent of the internet and mobile phone access to it has probably allowed a barter system of exchange to occur a little more efficiently these days with less time wasted in effecting suitable transactions. Nonetheless, there is a much more efficient means of conducting transactions, and that is by using a widely accepted single means of payment for the exchange of goods and services and settling debts.

Today, we call this widely accepted single means of payment 'money', and for something to be considered money, it needs to be able to serve three essential functions, which we will now discuss.

The functions of money

Money
Anything that serves as a medium of exchange, unit of account and store of value.

Fundamentally, anything, regardless of its innate value, can serve as money if it conforms to the following definition of *money*. **Money** is anything that serves as a medium of exchange, a unit of account and store of value. Anything that meets the three tests is a candidate to serve as money. This explains why money can be (and has been in different places and at different times) precious metals, beaver skins, wampum (shells strung in belts) or cigarettes. Let's discuss each of the three functions that money serves.

Money as a medium of exchange

Medium of exchange
The primary function of money, which is to be widely accepted in exchange for goods and services.

The most important function of money is to serve as a **medium of exchange**. This means that to be 'money', something has to be very widely accepted in exchange for goods and services. In other words, money can be used as a medium to transfer ownership or property rights.

Money removes the problem of coincidence of wants inherent in a barter system because every person is willing to accept money in payment rather than to be paid in goods and/or services. You give up a $100 note in exchange for a ticket to see a concert. This $100 note is accepted because the concert organiser has every confidence that she will be able to trade the $100 for the equivalent quantity of goods and services that she may want. Because money serves as generalised purchasing power, everyone in the society knows that no one will refuse to trade their products for money. In short, money increases trade by providing a much more convenient method of an exchange than a more cumbersome barter system.

Money as a unit of account

Unit of account
The function of money to provide a common measurement of the relative value of goods and services.

How does a wheat farmer know whether a tonne of wheat is worth one, two, five or more pairs of shoes? How does a family compare its sales revenue to its expenses, or a business know whether it is making a profit? Money must therefore be able to serve as a **unit of account** so as to provide a common measurement of the relative value of goods and services. The use of money makes it much simpler to compare the relative value (or cost) of sometimes very different products. For example, if the price of a movie ticket is $15 and the price of a lunchtime sushi serving is $5, then one sushi serving is valued at one-third of the price to see a movie.

In Australia, the monetary unit is the dollar but in other countries the monetary unit may have a different name; thus, in Japan it is the yen, in Malaysia it is the ringgit, in Europe it is the euro, in the UK it is the pound, in China it is the renminbi (or yuan), and so on.

Money as a store of value

Could you save fresh seafood for months and then exchange it for some product? You could, but not without the extra expense of freezing the seafood for an extended period. Money, on the other hand, serves as a cheap **store of value**, which is the ability of money to hold value over time and this durability of money allows us to synchronise our earned income more precisely with our expenditures. In other words, money can transfer purchasing power from the present to the future.

Store of value
The ability of money to hold value over time.

It may also be noted that if something can serve the above three functions, then it can also serve as the means by which deferred or future payments are stated and honoured; that is, money also serves as a *standard of deferred payment*. This is sometimes mentioned as a fourth function of money but one can reasonably argue this function is implicitly subsumed within the other three.[1]

A key property of money is that it is completely *liquid*. This means that money is immediately available to spend in exchange for goods and services without any additional expense. Money is more liquid than real assets, such as real estate, gold or jewellery, or some other types of financial assets, say, stocks or bonds. These assets also serve as stores of value, but liquidating (selling) them often involves expenses, such as brokerage fees, time delays and even losses of nominal value due to a need for a quick, possibly forced sale when their market value is low.

Are credit cards and cheques money?

It is an understandable misconception that credit cards, such as Visa, MasterCard and American Express, are 'plastic money'. But credit cards are nothing more than a more modern way or means to conduct transactions compared with, say, using cheques. A cheque is not money either. It is merely an instruction to a bank to transfer some number of dollars (which are money) from you to someone else in payment for services rendered or goods provided. Credit cards – and cheques – certainly improve the efficiency of our system of conducting transactions, but in themselves, they do not constitute money.

Are debit cards money?

You're the economist

Debit cards are used to pay for purchases, and the funds are automatically deducted from the purchaser's bank account at the time and place of purchase. Are debit cards therefore money?

Other desirable properties of money

Once something satisfies the above-described basic functions to serve as money, there are additional issues to consider. First, an important consideration is *scarcity*. Money must be scarce, but not too scarce. This was one reason why in earlier times gold often served as money.

Counterfeiting can threaten the scarcity of modern-day paper money and advances in computer graphics, scanners and colour copiers have potentially allowed counterfeiters to gain an advantage over the monetary authorities. However, the introduction in Australia of plastic notes (this kind of 'plastic' really is money!) was motivated at least in part to reduce the possibility of counterfeiting.

Money should be *acceptable* by most people and also it should be of *standardised quality*. That is the value of the dollar notes (plastic notes) should not quality or the material used to make the notes.

[1] On this, in 1919, the four functions of money were summarised by Alfred Milnes in a rather nice little couplet, viz., 'Money's a matter of functions four, a Medium, a Measure, a Standard, a Store' (in A. Milnes, *The Economic Foundations of Reconstruction*, Macdonald and Evans, 1919, p. 55).

The other great advantage of plastic notes is that they have increased *durability*. For example, they dry very quickly – and do not shrink – even after you forgetfully leave them in your pants when you do your washing (the authors can vouch for this). Other countries are now adopting this style of note.

Money should also be *portable* and *divisible*; that is, people should be able to reach into their pockets and be able to make change to buy items at different prices. Miniature statues of some sporting greats might be attractive money to some, but they would be challenging to carry and make change with!

Global perspective

Will Bitcoin soon take its place as another of the world's legitimate international currencies?

Shutterstock.com/kaprik

Bitcoin was created in 2009 by a person or persons calling themselves Satoshi Nakamoto[1] and is essentially a software-based digital payments system. It was the first 'digital currency' or 'cryptocurrency', and has now been followed by over a dozen more (e.g., Litecoin and Peercoin). The 'currency' is not provided or regulated by any centralised authority, such as a central bank, and it only exists in virtual form.

Individual holdings of Bitcoins are maintained on a central public online virtual ledger, which also registers transfers of Bitcoins from one individual or company to another. So, in this sense, Bitcoins have a great deal in common with bank deposits in that bank deposits also exist as electronic entries in databases maintained by banks and central banks around the world.

The easiest way to acquire Bitcoins is to exchange your local currency (say Australian dollars) at the prevailing Bitcoin price with someone who has Bitcoins. Your bank deposits are then transferred to the bank account of the seller of the Bitcoins and you are credited with the agreed number of Bitcoins in the Bitcoin central ledger. Another way is to try to 'mine' Bitcoins yourself (this being a rather specious analogy to gold and silver mining) by solving a complex equation/algorithm using a computer. If successful, you will receive some number of newly created Bitcoins.

To ensure Bitcoins remain relatively scarce, it is apparently the case that only 21 million Bitcoins can ever be mined (estimated to occur around the middle of the 22nd century). Furthermore, the computer mining of them has been predesigned within the computer software 'backing' them to get more and more difficult over time.

At the time of writing (late 2020) the number of Bitcoins in 'circulation' was estimated to be around 18.5 million (i.e., about 88% of the 21 million). Estimated 35 000-plus online sellers with no physical premises were accepting Bitcoins as payment, and quite a few businesses with physical premises were also accepting them. As of September 2020, a list could be found at https://www.buybitcoinworldwide.com/who-accepts-bitcoin. Very well-known businesses that accept Bitcoin include Microsoft, Wikipedia, BMW, Rakuten and Starbucks.

The value of Bitcoins from their creation in 2009 through to 2020 had been quite volatile. For example, from January 2013 to July 2014, the value of a Bitcoin varied from US$13 to US$980 and back down again to US$620 by July 2014. November 2017 was a quite astronomical $12678 and in that month, the Brisbane *Courier-Mail* reported on a house for sale on the Gold Coast in Queensland, which was listed by its owner for sale for 500 Bitcoins!.[2] The owner said he was quite genuine but he did admit that Bitcoin's fluctuating value made his decision to accept the digital currency as payment a pretty risky one! By September 2020, one Bitcoin was valued at $15 005.

The Bitcoin phenomenon clearly continues to have many admirers, adherents and backers – the most well-known of whom were probably the internet entrepreneurs (and 2008 US Olympic rowers) the Winklevoss twins, who came to prominence in the movie, *The Social Network*, about the beginnings of Facebook. Time will tell as to the eventual longevity of this interesting new digital innovation.[3]

What do you think?

1 What is a Bitcoin worth at the time you are reading this and how varied has its value been in the previous year?
2 What properties (durability, scarcity, portability, divisibility) of money do Bitcoins possess, and what functions of money (medium of exchange, store of value and unit of account) do Bitcoins satisfy?
3 Can you think of limitations and/or dangers that might militate against the widespread adoption of such a new form of 'money'?
4 Finally, can you think of any ethical considerations relating to the use of cryptocurrencies in the buying and selling of goods, services and assets?

[1] The name is believed by most people to be a pseudonym for either one person or a group of persons.
[2] 'This can be yours for just 500 Bitcoins', *The Courier-Mail*, p. 6, 28 November 2017.
[3] This discussion draws on a number of sources but most heavily from Wikipedia, http://en.wikipedia.org/wiki/Bitcoin, accessed 16 July 2014.

What stands behind our money?

Historically, early forms of money played two roles. If, for example, a ruler declared beans as money, you could 'spend' them in the marketplace in exchange for something else you wanted, or alternatively, you could eat them yourself. Similarly, precious metals, cigarettes and tobacco, cows and other tangible, reasonably durable goods are examples of *commodity money*. **Commodity money** is anything that serves as money while also having market value in other uses. This means that commodity money itself has intrinsic worth (the market value of the material). For example, money can be pure gold or silver, both of which are valuable for non-money uses, such as making jewellery and serving other industrial purposes. However, value of commodity money will not have a standardised quality; for example, 1 gram of gold will not have the same value and the value depends on the purity of the gold.

Commodity money
Anything that serves as money while having market value in other uses.

Initially, many years ago, the world's currencies were backed by gold; that is, the currencies in use could be exchanged for the appropriate defined quantity of gold. This backing bestowed confidence in the currencies and their values relative to each other. During this time, countries were said to be on the so-called 'gold standard'. Today, this is not the case. For instance, Australia's paper money and coins are no longer backed by gold or silver or any other physical commodity. Today, if you go into your bank, or indeed into your country's central bank, and seek to exchange your old, fraying five dollar note or whatever is your currency, what you will get in exchange at best will be a new five dollar note.

There is no intrinsically valuable commodity backing the world's currencies. They are **fiat money**, which is money issued by the central bank that is accepted by law but not because of any redeemability or intrinsic value. A US$5 note contains only about 11 cents' worth of paper, printing inks and other materials (and so does a $10 and $20 note). Pull out a note and look at it closely. Somewhere on it, you should find something to the effect of its being 'legal tender' – it is therefore legally acceptable in the payment of all debts. In Australia, the wording is: 'This Australian Note is Legal Tender Throughout Australia and Its Territories'. Also notice that nowhere on your note does it say there is any promise to redeem it for gold, silver or anything else.

Fiat money
Money accepted by law and not because of redeemability or intrinsic value.

In summary

- The use of money as a medium of exchange facilitates an increase in market transactions and increases economic efficiency and growth.
- The three functions of money are to act as a **medium of exchange**, as a **unit of account** and as a **store of value**.
- Desirable properties of money are liquidity, scarcity, durability, acceptability, standardised quality, portability and divisibility.
- Today, the world's currencies are **fiat money** and considered legal tender simply by government fiat, not because of any intrinsic value they have.

2 The determination of interest rates

As you will soon see, interest rates are determined by the demand and supply of money. Let's begin then by looking at the determinants of the demand for money.

The demand for money

Why do people want (or demand) to hold some of their wealth in the form of money – or, more generally, some quantity of highly liquid financial assets, such as currency and cheque deposits – rather than putting it to work in shares, bonds, real estate or other non-money forms of wealth? Because money yields no (or very little) direct return, people (and businesses) who hold cash or cheque account balances incur an *opportunity cost* of foregone interest or profits on the amount of money held. So what then are the benefits of holding some portion of one's wealth as money, and under what circumstances might the quantity of money people wish to hold change over time? There are considered to be three important motives for doing so: the transactions motive, precautionary motive and speculative motive.

Transactions motive for the demand for money
The explanation for the stock of money people hold to pay everyday predictable expenses.

Precautionary motive for the demand for money
The explanation for the stock of money people hold to pay unpredictable expenses.

Speculative motive for the demand for money
The explanation for the stock of money people hold to take advantage of expected future financial asset price changes.

Demand for money
A downward-sloping schedule – or inverse relationship – representing the overall quantity of money that people wish to hold at different interest rates, *ceteris paribus*.

The **transactions motive for the demand for money** is the wish to hold a stock of money to pay everyday predictable expenses. Without enough cash, people and businesses must suffer inconvenience and possibly withdrawal penalties as a result of converting their shares, bonds or certificates of deposit into currency or cheque account deposits in order to make transactions. There is, however, a cost for holding this money, namely the income you forgo by not holding that quantity of your wealth in some other form that earns a good rate of interest or return.

Closely allied to the transactions motive is the **precautionary motive**, which stems from the desire of people to hold a stock of money to pay unpredictable expenses or emergencies. Again, holding money balances for this reason incurs the opportunity cost of the interest or income you forgo by not holding that amount of your wealth in some other income-yielding form.

The third motive for holding money is the **speculative motive**, which arises from the wish to try to take advantage of expected future asset price changes. As implied in the discussion above, people will, in general, seek to economise on their holdings of money when interest rates are high. Also, when interest rates are high people may believe – or speculate – that they are likely to fall in the future. Therefore, they may choose to 'park' a relatively more considerable amount of their financial wealth in such things as fixed-interest corporate bonds rather than in money. If interest rates do fall in the future, their bonds, which pay a higher interest rate, will experience an increase in market value.

Similar reasoning applies when interest rates are low. The market value of any fixed-interest bonds with a low rate of interest will fall when interest rates rise. Investors may therefore want to hold relatively more of their wealth in the form of money when interest rates are low to avoid such losses of market value in the future.

The three motives for holding money combine to create an overall **demand for money**. As is evident from the above discussion, the important factor driving all three motives is the prevailing level of interest rates. The demand for money is, therefore, an inverse function of the interest rate, *ceteris paribus*.

As an aside, throughout the text we will often just say 'the interest rate'. Of course, there is a vast array of interest rates in the real world. When we say, 'the interest rate', this is really just a convenient shorthand for that array. It is also true that, in the real world, all interest rates will tend to vary over time in fairly systematic ways relative to one another, and you could therefore think of 'the interest rate' as some average or representative rate of all of the available rates.

Other things being equal, people will tend to increase their money balances when interest rates fall as they will not feel it so important to park their wealth in income-yielding assets. Instead, they will be content with holding higher amounts of their wealth as money balances, such as currency and cheque account deposits.

Note that another important determinant of the demand for money not explicitly highlighted here is income. Other things being equal, as individuals and firms experience higher levels of income they will tend to want (or need) to hold larger money balances to accommodate their higher dollar volumes of transactions. Any tendency

for rises in income to increase money demand will, of course, also be affected by movements in interest rates. For simplicity of exposition, we have chosen here to emphasise the role of interest rates in influencing the demand for money.

Four money-supply definitions

After having considered the demand for money, what exactly constitutes the money supply of modern economies? In the real world there is no hard and fast answer to this question, so central banks compile and monitor a range of measures. The following section presents four different definitions of the money supply in Australia:

- **Monetary Base** – the primary liquidity of the financial system. The assets in the Base are the most liquid of all financial assets and consist of all currency in circulation (notes and coins) in the private sector, plus the deposits that banks have with the RBA, plus other liabilities of the RBA to the private non-bank sector. The Base is under the close control of the RBA and is important for understanding how monetary policy adjustments are made.
- **M1** – the sum of currency in the hands of the non-bank public plus the stock of cheque account (demand) deposits of the private non-bank public with banks.
- **M3** – the sum of M1 plus all other bank deposits plus all other ADI deposits by the non-ADI public.[2]
- **Broad Money** – defined as the sum of M3 plus the non-deposit borrowings from the private sector by AFIs.[3]
- **Currency** – includes coins and paper money. The first component of M1 is the cash the public holds for immediate spending. Currency comprises just under 20 per cent of the economy's M1 money supply – and less than 4 per cent of Broad Money (see **Exhibit 15.1**). Thus, the actual currency is indeed 'small change' when it comes to what we regard as 'money'. The purpose of currency these days is just to permit us to make small purchases.

Exhibit 15.1 shows the components of the four money-supply definitions based on data from July 2020.

- **Cheque account deposits** Most 'big ticket' purchases are paid for with cheques, credit or debit cards rather than currency. As you have seen, credit and debit cards are not money. However, the money spent by customers in their use of these cards is often held as deposits in cheque accounts. These deposits – as with all the other forms of deposits mentioned below – are really just bank borrowings from the public. **Exhibit 15.1** shows that the largest share of M1 consists of cheque account (demand) deposits. These are deposits in banks, which – by the writing of a cheque – can be transferred 'on demand' by the bank on behalf of the bank's customer to a third party.[4] In July 2020, just over 90 per cent of M1 – and about 50 per cent of Broad Money – consisted of cheque deposits.
- *Other deposits* – as shown in **Exhibit 15.1**, M1 was around 54 per cent of M3 in July 2020, with 'other deposits' constituting the rest of M3. These deposits consist of other deposits of the public at banks (e.g., savings deposits, fixed-term deposits and other interest-bearing deposits), plus deposits by the public at non-bank deposit-taking institutions (e.g., credit unions and building societies).
- *Non-deposit borrowings by AFIs* – this final category comprises all other non-deposit borrowings from the private sector by financial institutions (including banks, credit unions, building societies and a defined range of other financial institutions). **Exhibit 15.1** shows that this category accounted for only about 0.2 per cent of Broad Money in July 2020.

To simplify the discussion, throughout the remainder of this text, we will be referring to the narrower definitions of money (say, the Monetary Base or M1) when we discuss the money supply.

Monetary Base
The primary liquidity of the financial system, consisting of currency in circulation in the private sector, plus the deposits of banks with the RBA, plus other RBA liabilities to the private non-bank sector.

M1
The narrowest definition of the money supply, consisting of currency and cheque account deposits.

M3
Equals M1 plus all other bank and other ADI deposits.

Broad Money
Equals M3 plus the non-deposit borrowings by AFIs.

Currency
Money, including coins and paper money.

Cheque account deposits
Convertible to currency 'on demand' by writing a cheque.

---

2 ADI stands for 'authorised deposit-taking institution'. ADIs (banks, building societies and credit unions) are supervised by the Australian Prudential Regulation Authority (APRA).

3 AFIs denotes 'all financial intermediaries'. A more precise definition of Broad Money is 'the sum of M3 plus the non-deposit borrowings from the private sector by AFIs, less the currency holdings and bank deposits by registered financial corporations (RFCs) and cash management trusts'.

4 Cheque account deposits are also sometimes referred to as current deposits or demand deposits.

Exhibit 15.1 **Components of Australia's money supply, July 2020 ($ billion)**

	(a) Currency held by the non-bank public	(b) Current deposits at banks	M1 (a + b)	(c) Other deposits at banks and other non-bank ADIs	M3 (M1 + c)	(d) Non-deposit borrowings from private sector by AFIs	Broad Money (M3 + d)	Monetary Base
	89.0	1202.3	1291.3	1078.5	2369.8	4.6	2374.4	178.1
% of Broad Money	3.7	50.1	54.4	45.4	99.8	0.2	100	7.5

Source: Extracted from http://www.rba.gov.au: Statistics > Money and Credit Statistics > D3, Monetary Aggregates.

Interest rate determination

We are now ready – at least conceptually and very stylistically – to consider the determination of interest rates in the economy by putting the demand for money and the supply of money together. Later in this chapter, we will also give you the flavour of how the financial system works in the real world, and in the next chapter, we will outline in detail how monetary policy works. Here we will sketch out the basic principles of interest rate determination. In **Exhibit 15.2**, the money demand curve (*MD*) is a graphical representation of the discussion we had above with respect to the inverse relationship between the quantity demand for money balances and the interest rate (i.e., as the interest rate rises, the quantity demand for money decreases because of the higher opportunity cost of holding it).

The supply of money curve (*MS*) is drawn as a vertical line. Such a representation amounts to an assumption that the quantity of money available in the financial system does not respond to changes in the interest rate. This is done for ease of exposition. However, the assumption of a vertical supply schedule would be a closer approximation to reality the narrower the definition of money one was considering. The quantity of the Monetary Base in supply, for example, is continuously monitored and extremely closely influenced by the RBA, and so, for that very narrow definition, this is not at all an unrealistic assumption.

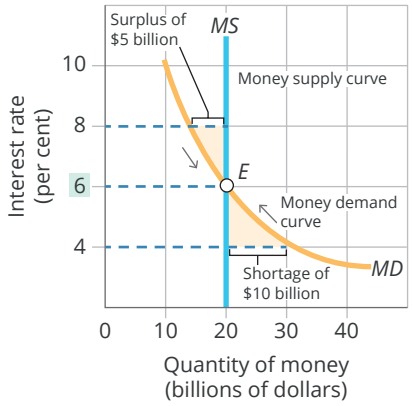
At point *E*, the equilibrium interest rate is 6 per cent at the intersection of the demand for money curve and the vertical supply of money curve. People wish to precisely hold the amount of money in circulation and, therefore, neither upward nor downward pressure on the interest rate exists.

Excess quantity of money demanded

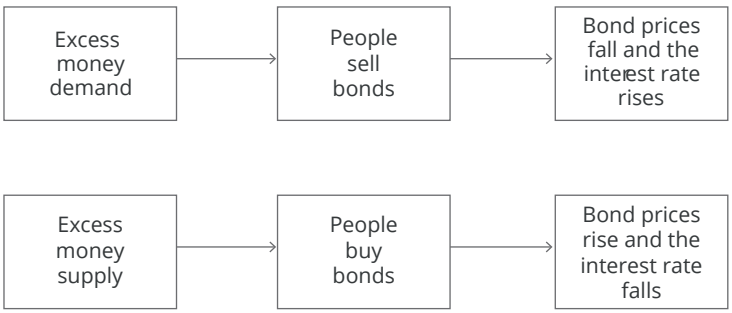
Now, just for argument's sake, suppose the interest rate in **Exhibit 15.2** were 4 per cent instead of 6 per cent. This lower opportunity cost of money would mean that people would desire to hold a greater quantity of money (from the exhibit you can see it is $30 billion) than the quantity available for holding ($20 billion). To try to eliminate this shortage of $10 billion, individuals and businesses adjust their financial asset portfolios. They seek to hold more money by selling their bonds and other non-money financial assets.

Here, you should note our subsequent use of the term 'bonds' in this text. Of course, there are financial assets that explicitly carry this name; for example, Commonwealth government bonds (securities issued by the federal government when it wishes to borrow) and corporate bonds (securities issued by companies when they wish to borrow). By their nature, these financial assets provide relatively less liquidity to their owner than, say, currency or cheque account deposits; that is, money as defined by M1 above.

Exhibit 15.2 The equilibrium interest rate



CAUSATION CHAINS



The money market can be thought of as the demand for and the supply of very highly liquid financial assets – denoted by the term 'money'. The money demand curve represents the quantity of money people are willing to hold at various interest rates. The money supply schedule is drawn as a vertical line at $20 billion, reflecting a simplifying expositional assumption that the quantity of money supplied is unresponsive to changes in the interest rate. The equilibrium interest rate is 6 per cent and occurs at the intersection of the money demand and the money supply curves (point E). At any other interest rate – for example, 8 per cent or 4 per cent – the quantity of money people desire to hold does not equal the quantity available. At 8 per cent, there is excess supply of money and at 4 per cent, there is excess demand for money.

However, in the subsequent discussion we will be using the term 'bonds' in a broader way. We will be using the word 'bonds' to represent the set of all those income-yielding – but relatively less liquid – financial assets that people hold in their wealth portfolios in addition to their holdings of money. Thus, in what follows – and as we have also been trying to stress in earlier discussions – when we talk of 'money' it is best to think in terms of a set of relatively highly liquid, but low or zero-income-yielding, financial assets, and when we talk of 'bonds' we mean the whole set of less liquid, but higher-income-yielding financial assets that people hold in their financial portfolios. In the real world, there is, of course, a whole continuum of financial assets, from the most highly liquid through to the least liquid. However, the use of this artificial two-asset classification into 'bonds' and 'money' is a useful simplifying expositional device.

Now back to **Exhibit 15.2** and the $10 billion excess demand for money at an interest rate of 4 per cent. When many people sell or try to sell their bonds to satisfy their demand for higher money holdings, there is an increase in the supply of bonds relative to the demand for bonds. Consequently, the price of bonds falls, and their implied interest rate (or yield) rises. This rise in the interest rate ceases at the interest rate of 6 per cent because people are then content with their portfolio of money and bonds at point *E*. In other words, at *E* the

money market would, therefore, be in equilibrium and, *ceteris paribus*, there would not be any pressure for the interest rate to move further.

Here we perhaps need to pause and look at an example to understand what happens when market forces change the price of a bond (see **Exhibit 15.3**). Suppose Shell pays 4 per cent on its $1000 two-year corporate bonds. This means it is promising to pay a bondholder $40 in interest each year and to repay the original $1000 (the bond's face value) at the end of two years. However, a holder of these bonds can sell these bonds before maturity at a market-determined price. If Shell bondholders desire to hold more money, they will increase the supply of these bonds for sale in the market. As a result, the increase in the supply of the bonds causes their price to fall to, let's say, $500. At a price of $500 the yield – or implied interest rate – rises to 8 per cent ($40/$500).[5]

Exhibit 15.3 **Illustrating the inverse relationship between the price of a fixed-interest bond and its yield**

Market price of bond ($)	Interest paid annually (4% of face value $1000)	Yield (implied interest rate, %)
2000	40	2
1000	40	4
500	40	8

Excess quantity of money supplied

The story reverses for any rate of interest above 6 per cent. Let's say the interest rate is 8 per cent. In this case, people are holding more money than they wish. In fact, they wish to hold only $15 billion rather than the $20 billion in circulation. To correct this imbalance, people will try to move out of cash and cheque deposits by buying income-yielding/interest-bearing bonds.

This increase in the demand for bonds will drive up the price of bonds and thereby lower the implied interest rate. As the interest rate falls, the quantity of money demanded increases as people become more willing to hold money. Finally, the money market reaches equilibrium at point *E* and people in the aggregate are content with their mix of money and interest-bearing bonds.

Interest rates, aggregate demand, the central bank and monetary policy

Monetary policy Actions taken by the central bank to influence interest rates, aggregate demand, economic activity and prices in the economy to achieve macroeconomic goals.

In the previous section we explained – in admittedly a very simple and highly stylised way – the basic mechanism by which interest rates are determined in the economy's financial markets and, as we saw last chapter, interest rates are a very important determinant of aggregate demand. And, as we will see in more detail next chapter, the nation's central bank can and does have a strong influence over the prevailing level of interest rates in an economy. This then allows the central bank to exercise a significant influence over aggregate demand, and the exercise of this influence is known as *monetary policy*. **Monetary policy** is the term used to describe actions taken by the central bank to influence interest rates, aggregate demand, economic activity and prices in the economy to achieve macroeconomic goals. In Chapter 16 we are going to cover monetary policy in great detail.

[5] The situation is a little more complicated than this because of the time period over which the coupon payments and the face value are paid. Nevertheless, the essence of the inverse relationship between the price of a bond and its yield is reflected in the example.

For the present, in the next section we will look further at the role played by the central bank in a nation's financial system and also include other crucial aspects of a country's financial system. Again, while we will refer to Australia's financial system institutional arrangements, any modern financial system will incorporate very similar components.

In summary

- There is an inverse relationship between the quantity of money demanded and the interest rate.
- Four important definitions of Australia's money supply are: the **Monetary Base**, **M1**, **M3** and **Broad Money**.
- The demand for and supply of money in financial markets determines interest rates.
- The central bank can influence interest rates, aggregate demand, economic activity and inflation.

3 A modern financial system

The financial system is critical to the proper and efficient functioning of any modern developed economy. The **financial system** facilitates the transfer of resources from those with resources excess to their immediate needs (savers) to those with immediate needs for such resources (borrowers). A country's financial system consists of many key players (individuals, organisations, institutions) and a mosaic of interconnected markets.

Financial system Facilitates the transfer of resources from savers to borrowers.

At the very centre of any developed country's financial system stands the nation's central bank and the banking system. Also playing a key role will be the country's non-bank deposit-taking institutions (building societies and credit unions) and a range of financial markets, such as money markets, capital markets (like bond or fixed-interest markets and equity markets), foreign exchange markets, derivatives markets, insurance markets and futures markets. The Australian financial system is an excellent example of a modern financial system, and so in reading about some of its particular institutional features, you will gain an insight into the basic elements of any modern financial system. In this book we will give only its essential flavour.

As we learnt earlier in this chapter, the foundation of Australia's money supply is – as its name infers – its Monetary Base, and the Base is under the close control of the **Reserve Bank of Australia (RBA)**. The RBA is the central banker for the nation and provides banking services to commercial banks, other financial institutions and the federal government. The RBA, through its close control of the Monetary Base and implementing monetary policy, can significantly influence financial market interest rates and thereby significantly influence the economy's performance.

Reserve Bank of Australia (RBA) The central bank of Australia, responsible for monetary policy, the payments system and financial system stability, and providing banking services to the banks, other financial institutions and the federal government.

In addition to a country's central bank, also at the core of a country's financial system is the nation's banking system and a number of major financial markets. Each will be briefly discussed in turn.

Australia's central bank: the Reserve Bank of Australia

The RBA was created by the *Reserve Bank Act 1959*, which outlined its powers and responsibilities and the *Banking Act 1959*; then further elaborated in the *Financial Corporations Act 1974*, and, more recently, its powers and responsibilities have been amended as a result of the 1997 Financial System Inquiry, commonly known as the Wallis Inquiry. As a result, RBA operations are also now further defined in the *Payment Systems (Regulation) Act 1998* and the *Payment Systems and Netting* Act 1998.

The *Reserve Bank Act 1959* outlined the duty that the RBA has to the Australian people: to maintain the stability of the currency (fight inflation); to maintain full employment, and to ensure the economic prosperity and welfare of the Australian people. The *Banking Act 1959* required the RBA to uphold the integrity of the banking system. To this end, it had the power to regulate trading and savings bank lending, interest rates and asset structures. The *Financial Corporations Act 1974* gave the RBA the power to monitor and potentially control Australia's many non-bank financial intermediaries (NBFIs). NBFIs include, among others, building societies

and credit unions. This Act was revoked as a result of recommendations of the Wallis Inquiry. The *Payment Systems (Regulation) Act 1998* and the *Payment Systems and Netting Act 1998* outline the RBA's responsibilities and authority in relation to Australia's financial payments system.

As a result of the Wallis Inquiry, the prudential supervision of all of Australia's deposit-taking institutions – including banks, building societies and credit unions – and the nation's superannuation and insurance companies now falls to the Australian Prudential Regulation Authority (APRA). Also, as a result of the Wallis Inquiry, the Australian Securities and Investments Commission (ASIC) was created to ensure proper financial disclosure, oversee financial market integrity and uphold consumer protection.

Reserve Bank Board
The RBA Board determines monetary policy in Australia. It consists of the Governor, the Deputy Governor, the Secretary to the Commonwealth Treasury and six other members appointed by parliament.

The RBA is governed by the **Reserve Bank Board**, chaired by the Reserve Bank Governor. In 1996, the Governor and the Federal Treasurer co-signed the Statement on the Conduct of Monetary Policy, which commits the government to endorse the RBA's freedom to implement whatever monetary policy it thinks is appropriate to ensure the achievement of the target inflation rate agreed with the government.

The RBA acts as banker to the federal government, the banking system and other major financial organisations, as well as managing the government's debt. In its latter role, on behalf of the Treasury, it sells Commonwealth government bonds (maturities of at least one year) and 13- and 26-week Treasury notes by tender to the public. Together, these are commonly known as **Commonwealth Government Securities (CGS)**. The proceeds of these sales of securities are deposited – along with taxation receipts – into the government's account maintained at the RBA and are subsequently used by the government to pay for its expenditures.

Commonwealth Government Securities (CGS)
Commonwealth government bonds (maturities of more than one year) and Treasury Notes (13- and 26-week).

The banking system

Banks occupy a crucial position in our community. The solvency of a bank is virtually guaranteed because of the careful and comprehensive supervision of a bank's activities and the particular asset structure it is required to maintain. However, a common misperception is that a bank is as 'safe as houses' because it will not be allowed by the authorities to fail.

In fact, this is not true. No bank – or any other financial institution for that matter – is absolutely guaranteed to remain solvent (witness the failures of some of the world's largest financial institutions in the global financial economic crisis of 2008).[6] However, the RBA is obliged to act in such a way as to minimise the possibility of a systemic collapse. In addition, the vigilant supervision of each bank (by APRA), along with the robust liquidity and capital adequacy requirements demanded of them all, ensures that general public confidence, so crucial to the efficient operation of the financial system, is well placed.

Exchange settlement account
Each bank maintains an exchange settlement account with the RBA and they must always have a positive balance by the end of each trading day.

A bank has direct access to the RBA and the nation's payments clearance mechanism. Each bank maintains an **exchange settlement account** with the RBA and any direct business it conducts with the central bank, other banks, the government and specific other financial organisations are settled through deposits to or withdrawals from this account. It is this direct RBA access that enables a bank to offer cheque-writing facilities to its customers. There are now over 50 banks authorised to conduct business in Australia, and, since the freeing up of the financial system in the 1980s, many of these banks are foreign-owned.

The day-to-day business of a bank is to accept deposits and make loans. In addition, banks are very active participants in the capital and money markets and the foreign exchange markets (discussed below) and, through these activities, provide other services to their customers, such as currency conversion. In conducting their business, banks are subject to three forms of regulation:

- A bank's exchange settlement account must have a positive balance at the end of each trading day.
- Each bank must maintain appropriate high standards of liquidity management.
- Each bank must satisfy certain prescribed capital adequacy requirements.

[6] Notwithstanding this, refer also to the upcoming section 'Is it good to allow banking in Australia to be dominated by a few very big banks?'.

Applicable concept: monetary policy

Analyse the issue

Monetary policymaking in action: How does the RBA make its monetary policy decisions?

Getty Images/Bloomberg

Philip Lowe, Governor of the Reserve Bank of Australia

The RBA Board meets 11 times a year. The meeting occurs on the first Tuesday of each month (excluding January) and a consensus decision is reached as to whether to leave interest rates unchanged or, if the decision is to adjust policy, then the quantum of the adjustment. The instrument the RBA uses is the cash rate, and, if an adjustment is decided, it will be to adjust the cash rate by 0.25 per cent, 0.50 per cent or 1 per cent. The most common adjustments are 0.25 per cent.

The Board's decision on interest rates is announced on Tuesday afternoon (1st Tuesday of the month), and if a rate adjustment is to be made it is enacted the next day. The RBA adjusts the cash rate by appropriately adjusting its daily financial operations in the money markets by either *buying* or *selling* Treasury notes (T-notes). The RBA's actions affect liquidity available in financial markets, and this moves the cash rate to the desired level. As a result of these changes in the cash rate, interest rates, in general, are also influenced.

A statement is issued on the Tuesday by the Governor, providing a summary rationale for the Board's decision, while minutes of Board meetings are published two weeks after each meeting.

What do you think?

1 What happened at the most recent Board meeting prior to your reading this?
2 If there was a change in monetary policy, what was the rationale provided by the Governor?
3 If there was no change, when was the last adjustment in monetary policy?

The first requires that exchange settlement accounts maintained with the RBA must always have a positive balance at the end of each day. Exchange settlement accounts are an integral part of the payments system in modern financial systems, such as Australia's. Today, people write cheques, use credit and debit cards, telephone transfer, EFTPOS, direct debit and so on rather than pay with currency. These transactions ultimately involve the electronic transfer of funds held at one financial institution to another, and these settlements occur through appropriate adjustments being made to their exchange settlement accounts at the RBA.

The second regulation requires each bank to maintain appropriate high standards of liquidity management to ensure it has adequate liquidity at all times. As an alternative to this, the financial institution may simply be required to maintain a **minimum liquidity requirement**. This requirement means that a bank must always maintain a certain minimal reserve level of highly liquid assets to cover any short-run demands from its many depositors and other creditors.

The third regulation is the **capital adequacy** requirement, requiring a financial institution to maintain a certain minimum percentage of its risk-weighted assets in the form of shareholders' equity. If a bank needs to pay off creditors (like its depositors), it can liquidate some of its assets, hence the minimum liquidity requirement discussed above. But in an emergency, the value of its assets may be uncertain. Thus, a bank's solvency will ultimately depend on the extent of its financial capital base as represented by its shareholders' equity. This financial capital backing provides further reassurance to creditors and depositors that their funds remain safe.

Minimum liquidity requirement Requirement that a bank must always maintain a certain minimal reserve level of highly liquid assets.

Capital adequacy Requirement to maintain a certain minimum percentage of its risk-weighted assets in the form of shareholders' equity.

The banking system and creating money

In the 1600s, gold was the money of choice in Britain. One of the problems with gold is that it is a heavy commodity, which makes it challenging to use in transactions or to hide from thieves. The medieval solution was to keep it safely deposited with the people who worked with gold – the *goldsmiths*.

This demand for their services inspired goldsmith entrepreneurs to become the founders of modern-day banking. The goldsmiths sat on their benches with ledgers close by and recorded gold placed in their vaults. The word *bank* is derived from the Italian word for bench, which is *banco*.

After assessing the purity of a customer's gold, a goldsmith issued a receipt to the customer for the amount of gold deposited. In return, the goldsmith collected a service charge, just as you pay today for services at your bank. Anyone who possessed the receipt and presented it to the goldsmith could make a withdrawal for the amount of gold written on the receipt. With these gold receipts in circulation, people began using them to pay their debts, rather than actually exchanging gold. Thus, by around 1660, goldsmiths' receipts had become a form of paper money. At first, the goldsmiths were very conservative and issued receipts exactly equal to the amount of gold stored in their vaults. However, some shrewd goldsmiths observed that net withdrawals in any period were only a *fraction* of all the gold 'on reserve'. This observation produced a powerful idea.

Goldsmiths discovered that they could make loans for more gold than the actual gold held in their vaults. As a result, goldsmiths made extra profit from interest on loans and borrowers had more money for spending in their hands. The medieval goldsmiths were the first to practise **fractional reserve banking**. Modern fractional reserve banking is a system in which banks keep only a percentage of their deposits on reserve as vault cash or as deposits at the central bank. Holding less than 100 per cent on reserve allows banks to create money.

Fractional reserve banking
A system in which banks keep only a small percentage of their deposits on reserve as vault cash or as deposits at the central bank.

To see how this happens, suppose that in some mythical economy there is just one single financial institution – say, the BigFriendlyBank. BFB has just had $100 000 in cash deposited with it. It could use the cash to buy some very secure government securities so that it could earn some interest on the money. However, a potentially more profitable possibility would be for it to extend loans to new or existing customers who need money for some business ventures.

Suppose the bank's management has agreed that, in order to satisfy its prudential liquidity management requirements, it needs to maintain a minimum liquidity requirement of at least 10 per cent (a nice round figure used here purely for expositional purposes) of its liabilities in the form of cash, deposits at the central bank, government securities and other highly liquid assets. Given its 10 per cent reserve requirement, it could use the $100 000 cash as its minimum liquidity requirement against it extending loans of $900 000 on which it could earn a market rate of interest. The advances augment its assets by $900 000 ($900 000 of new loans on its books), which are balanced by an increase in its outstanding liabilities as it credits the borrowers' bank accounts with deposits of $900 000. This process is known as **credit creation**. Furthermore, since the bank's deposits are regarded as money by participants in the economy (recall the above definitions), the bank has – literally at the stroke of a pen – created an additional $900 000 of money in the economy! At the end of this process, the bank has on its balance sheet $1 000 000 of additional deposit liabilities – the original $100 000 deposit plus the ensuing created deposits of $900 000 – offset by assets of $100 000 in cash and $900 000 in new loans. This is illustrated in **Exhibit 15.4**.

Credit creation
The process by which money is created by banks extending new loans to their customers in the form of newly created bank deposits.

Exhibit 15.4 **Illustrating the credit (money) creation process**

Changes in the balance sheet of BFB

Assets (A)	Liabilities (L)
$ 100 000 (vault cash)	$ 100 000 (initial deposit)
$ 900 000 (new loans)	$ 900 000 (additional newly created deposits)
$1 000 000	**$1 000 000**

In the above, we illustrate how the balance sheet of BigFriendlyBank changes with a new cash deposit of $100 000. Initially, BFB's assets (vault cash) increase by the amount of the cash deposit, while its liabilities increase by an equal $100 000 of deposits in the name of the depositor. With the reserve ratio set at 10 per cent, BFB is then able to further increase its assets by making new loans of $900 000. At the same time as extending these loans it creates $900 000 of new deposits in the names of those to whom it makes the loans – who can then use the funds extended to them for their various activities. These additional deposits also represent an increase in the money supply and so, 'at the stroke of a pen' as it were, BFB has created money.

In this example, the initial increase in the bank's reserves – the $100 000 – was associated with a tenfold increase of $1 000 000 in bank deposits. The so-called **money multiplier** in this case is 10. In general, the money multiplier may be thought of as the reciprocal of whatever is the required liquidity reserve ratio – here 0.10 (i.e., 10%).

Money multiplier The extent to which banks can multiply an initial increase in their reserves. It may be thought of as the reciprocal of the minimum reserve liquidity ratio that banks maintain to satisfy regulatory requirements; for example, if the minimum reserve liquidity ratio is 10 per cent = 0.1, then the money multiplier would be 10 (= 1/0.1)

Of course, in a real-world economy with many financial institutions, not all of the deposits created by BigFriendlyBank are likely to stay with that bank. The customers receiving the loans will use them to buy goods and services, and some of it will be electronically transferred to other banks and financial institutions. Thus, in the real world, the initial $100 000 cash deposit and the additional $900 000 is created deposits will end up being distributed among many members of the banking system.

Also note that, in the real world, the extent to which the money supply will actually increase as a result of increased reserves at banks (i.e., the $100 000 cash deposit used in the above example) will depend on many factors, among them the risks perceived by banks in extending loans. These risks will change over the course of the business cycle. Nevertheless, the concept of the money multiplier does capture an important aspect of the real-world money-supply process.

Is it good to allow banking in Australia to be dominated by a few very big banks?

The Australian government has accepted the view that the country is well served by allowing an industry structure for the nation's banking system characterised by significant market domination in retail banking[7] by the Big Four (relative to the size of the market) – that is, ANZ, CBA, NAB and Westpac. Those who support this structure argue it gives precious stability to Australia's banking system.

Those opposed to the dominance of the Big Four point to their profit margins and their very large – and, in their view, excessive – profits, and argue greater competition would work to reduce these profit margins. The Big Four banks' rates of return on equity are indeed very high, in fact, higher than some of the most well-known international banks in the world.

Opponents also argue that a bank can grow so big that it gets too big to be allowed to fail by the authorities, and that this creates a potential *moral hazard* in the practices of the bank. A moral hazard arises where an individual's behaviour is influenced by a perception that they will not have to bear the full consequences of their actions. In the case of financial institutions, the presence of moral hazard leads to the so-called 'too big to fail' issue, which goes like this.

Due to its size, and the impact its failure would have on its many depositors and shareholders (i.e., voters) should the institution get into trouble, management might come to believe the government would not allow it to fail should its lending and other investment practices go bad. Instead, they may come to implicitly believe the government would come to its aid and 'bail it out', usually using taxpayer funds. This belief could then lead management to pursue riskier activities than they otherwise would, and therefore actually increase the probability of the need for such a bailout. There is, in fact, considerable evidence from around the world that taxpayer funds have indeed been used – at least in part – to solve many emerging banking crises!

Australian banking and crisis

As we discussed, banks have an important role in facilitating economic activities. However, the nature of banking has changed over time due to financial crisis, developments in technology and competition. Australia's banking sector has learnt lessons from our own responses and international responses to the financial crisis.

[7] As opposed to 'wholesale banking' – also known as 'investment banking'– comprising activities such as wealth management for high-end individuals, facilitating the buying and selling of shares, bonds and other investments, and helping large companies raise funds – often by acting as underwriters. Retail banking usually involves having extensive branch networks. It comprises such activities as taking deposits from individuals and companies and extending loans for consumer and business purposes as well as for residential and commercial building.

According to the RBA and other experts, the Australian banking system performed well through the financial crisis,[8] and its performance was largely due to the thorough prudential supervision and strict management of risks by banks. Banks have reduced their risk taking and expanded the mortgage lending, which has collateral protecting. In addition, major banks started to divest life insurances and wealth management operations and these actions have reduced total profits.

The COVID-19 pandemic brought in enormous health and economic challenges and increased the risks for the financial system, including banks. However, financial systems, including banks, coped well with the shocks as a result of the post-global financial economic crisis reforms. The banking system is expected to have credit losses due to increasing loan defaults due to the global-wide economic contractions. The fiscal and monetary policy measures during the pandemic helped households and businesses in numerous ways. Due to the assistance received households have increased their savings and deferred the loan repayments and businesses received rent relief and loan repayment deferrals.

Banks are faced with increasing risks, lower profits and higher household, business and government debt. The global economic recovery will take some time and risks to financial systems will remain high for some time. However, Australian banks will maintain high capital levels than the minimum requirement and will be well placed to support the economic recovery.

The money markets

Several financial markets exist to allow the lending and borrowing of very large amounts of money for short periods of up to about a year, but often for periods as short as just a few hours or overnight. In some instances, money itself is simply borrowed or lent, and in other instances, a financial instrument is exchanged in return for money. For convenience in this text, we refer collectively to these interrelated markets as 'the money markets'.

As you will see in more detail in the next chapter, the main mechanism that a central bank uses in implementing monetary policy is the buying or selling of short-term government securities with banks and certain other participants in the money markets who have been granted the right to have exchange settlement accounts with it. When the central bank buys government securities, it pays for these securities and increases the Monetary Base (money supply) by creating and injecting new money into the financial system. When the central bank sells such securities, the money given up by market participants to the central bank in exchange for them is withdrawn from the financial system, thereby reducing the Monetary Base (money supply). As we saw earlier in this chapter when discussing interest rate determination, changes in the available supply of Monetary Base liquidity to the financial system impacts on interest rates.

Participants in the money markets consist of banks, investment banks, building societies, finance companies, insurance companies, large commercial and industrial companies, stockbrokers and individuals with large sums to invest or borrow for short periods (the smallest transaction usually is $100 000).[9] As noted above, the commodity being traded on the markets is short-term funds. These funds can be transferred based on a straight loan with an agreed interest rate for a fixed term, or they can be used to buy a marketable security (instrument) with the advantage that it can either be held to maturity or, if necessary, on-sold to someone else to obtain funds.

Treasury note (T-note)
A T-note is sold by the Treasury on behalf of the federal government and is considered riskless.

A couple of the most common instruments traded on the money market are Treasury notes and bank-accepted bills.

- A **Treasury note (T-note)** is sold by the Treasury on behalf of the federal government and is considered riskless. At maturity, the holder presents the note and receives its face value (a dollar amount printed on the note). When it is sold or traded on the market it sells at a discount to the face value, with the difference being the interest to the holder if held to maturity.

[8] J. Kearns, 'Changes in banking: Looking back and looking forward', Reserve Bank of Australia, https://www.rba.gov.au/speeches/2019/sp-so-2019-12-16.html

[9] However, note that not all money market participants have been granted the right to maintain an exchange settlement account at the RBA.

- A **commercial bill** is created when, as evidence of a loan, a borrower draws up a bill that stipulates a face value (say, $100 000) they will repay at maturity and the time to maturity (usually 90 or 180 days). The lender buys the bill from the borrower at a discount to the face value, with the difference representing their interest. To remove the risk of default by the borrower, a bank may, for a fee, guarantee payment at maturity and such a bill is referred to as a *bank-accepted bill*.

Commercial bill Bill created as evidence of a loan, stipulating a face value to be repaid at maturity and the time to maturity (usually 90 or 180 days).

Apart from the money markets, many other financial markets have evolved to satisfy people's financial needs. They all basically have at least one of two important roles: first, to distribute the community's savings to those uses expected to yield the highest returns; and second, to redistribute risk associated with various economic transactions to those most willing to bear them, thereby increasing the overall efficiency with which we use our scarce economic resources.

To go into these other markets in any detail would take us far beyond the purpose of this text. What follows is a brief introduction to some of them.

Capital markets

Whereas money markets exist to allow the exchange of funds for relatively short periods of time, capital markets have been created to allow the transfer of wealth of savers to borrowers who want those resources for the long term; that is, for more than a year, and sometimes for much longer. The exchange occurs through the buying or selling of what are called *instruments*, which are documents evidencing the existence of the transaction: either debt-backed securities (corporate and government bonds) or equity-backed securities (company shares).

As noted earlier in this chapter, a **bond** is a document that evidences an agreement whereby the borrower agrees to pay a fixed annual interest (the coupon) for a defined number of years (the maturity of the bond), and then, on maturity, repay the amount borrowed (the principal) to the holder of the bond. The bond may be held to maturity by the original lender of the funds or subsequently on-sold to someone else. Because the agreed interest rate is fixed for the life of the bond, bond markets are also often called *fixed interest markets*. For example, when the federal government wishes to borrow for long-term purposes, the Treasury will issue Commonwealth government bonds of varying maturities – two, three, five and 10 years. Another example, if a large company needs long-term funds, it can issue (sell) bonds – called corporate bonds – of some maturity. For large companies, this is usually a preferred form of raising debt capital than borrowing from banks, as borrowing from banks is often seen by such companies as involving more restrictions.

Bond A document stipulating a fixed annual interest (the coupon) for a defined number of years (its maturity), and on maturity, the amount to be paid out (its face value).

Another way for publicly listed companies to raise long-term funds is through selling shares. As the name implies, in selling shares, the company is raising funds by selling ownership in itself. As such, the funds raised never have to be paid back to the investor. The buyer of the shares owns a share in the company, has a share in its profits, has a proportional say in its management, and can, if they wish, sell the shares on the share market. Australia's share market is discussed below.

An example of a country's share market: the Australian Securities Exchange

The **Australian Securities Exchange (ASX)** is where shares in publicly listed companies are floated and traded, along with the buying and selling of a wide range of financial derivative products. Unlike the money markets, the ASX has an actual physical location and is among the largest fifteen listed exchanges in the world. Some of the world's biggest share markets are the New York Stock Exchange (NYSE), the Tokyo Stock Exchange (TSE), the NASDAQ Stock Exchange (the NASDAQ), Euronext, the London Stock Exchange (LSE), the Shanghai Stock Exchange (SSE) and the Hong Kong Stock Exchange (HKSE).

Australian Securities Exchange (ASX) Where shares in publicly listed companies are floated and traded, along with the buying and selling of a wide range of financial derivative products.

A company listed on the ASX is called a public company, and is one that is owned by a great many individuals (often thousands) whose financial liability in case of the company becoming insolvent is limited only to the value of their fully paid-up shares in it.

In 'going public', a private company floats its shares on the ASX and in doing so can increase very considerably the amount of financial capital it has to expand its businesses. In addition, in going public, the company has to meet a range of financial and reporting conditions placed upon it by the Corporations Law and the rules of the ASX. The shares of listed companies are traded extensively on a daily basis. Such trading allows individuals to progressively reduce or increase their ownership in a range of companies based on their estimates of the companies' likely future economic performance and their risk appetites.

The foreign exchange market

On any given day, for a multitude of reasons, thousands of individuals and companies want to exchange Australian dollars for foreign currencies or vice versa. For example, exporters want to sell foreign currency for Australian dollars, importers need to buy foreign currency, and companies need to buy foreign currency to make payments to overseas investors or to convert funds borrowed overseas into Australian dollars. The **foreign exchange (Forex) market** has evolved to accommodate such transactions.

Foreign exchange (Forex) market
Where individuals and companies exchange Australian dollars for foreign currencies or vice versa.

As with the money markets, the Forex market is not confined to a physical location. Trading of currencies is carried out by the electronic transfer of bank deposits denominated in different currencies, with participants in communication via the telephone, email and computer trading screen. The major players in the market are the banks but, in total, there are currently over 80 institutions authorised to conduct Forex operations in Australia.

The onset of the COVID-19 pandemic affected the foreign currency exchange industry and forecast to decrease the revenues significantly. International border closures and travel restrictions drastically reduced the demand for the services from the foreign exchange market and smaller institutions in the Forex market are likely to exit the industry. The recovery of the industry largely depends on the anticipated recovery from the COVID-19 pandemic.

Economics and ethics — Costs of inflation

In **Chapter 13** we discussed some of the negative consequences of inflation as well as briefly covering the phenomenon of hyperinflation. Hyperinflation is invariably the result of a government's ill-advised decision to increase a country's money supply dramatically to pay for government expenditure rather than raise taxes.

While hyperinflation is an extreme outcome of governments financing their own spending from increasing the money supply rather than raising taxes, any time a government 'prints money' to finance its spending, the risk exists that its action will produce higher inflation than would otherwise be the case. Furthermore, it can be persuasively argued that, to the extent a government produces higher inflation by so increasing a country's money supply, it has, in effect, levied surreptitious taxes on the community.

This follows because the resulting inflation reduces the real value of the stock of money in circulation held by the public as one component of their financial wealth. Therefore, effectively, there occurs a transfer of wealth from the private sector into the hands of the government. This sounds quite a lot like a tax by stealth, doesn't it? Since this 'tax' hasn't been legislated, the ethics of such a wealth transfer to the government are certainly worth considering.

In summary

- The **financial system** facilitates the transfer of resources from savers to borrowers.
- At the centre of the financial system stands the nation's central bank and its banks.
- The central bank is responsible for a nation's **monetary policy**.
- A bank's solvency is virtually guaranteed by the careful and comprehensive supervision of its activities, and the particular asset structure it is required to maintain.
- Banks can quite literally create money 'at the stroke of a pen'.
- Other important elements of the financial system include the money markets, the capital markets (bond and share markets) and the foreign exchange market.

Key concepts

Barter
Money
Medium of exchange
Unit of account
Store of value
Commodity money
Fiat money
Transactions motive for the demand for money
Precautionary motive for the demand for money
Speculative motive for the demand for money
Demand for money
Monetary Base
M1
M3
Broad Money
Currency
Cheque account deposits
Monetary policy
Financial system
Reserve Bank of Australia (RBA)
Reserve Bank Board
Commonwealth Government Securities (CGS)
Exchange settlement account
Minimum liquidity requirement
Capital adequacy
Fractional reserve banking
Credit creation
Money multiplier
Treasury note (T-note)
Commercial bill
Bond
Australian Securities Exchange (ASX)
Foreign exchange (Forex) market

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Learn about some monetary fundamentals

- **Money** is the most liquid of financial assets and the use of money as a **medium of exchange** simplifies – and therefore facilitates an increase in – market transactions. The use of money – rather than say a **barter** system of exchange – also prevents wasting time that could otherwise be devoted to production, thereby promoting economic efficiency and growth.
- The three functions of money are to act as a medium of exchange, as a **unit of account** and as a **store of value**.
- Desirable properties that anything serving as money should possess are liquidity, scarcity, durability, portability and divisibility.
- Today, the world's currencies are **fiat money** and are not backed by gold, any other precious metal or anything else, and are considered legal tender throughout a country's jurisdiction only by government fiat not because of any intrinsic value they have.

2. Understand the determination of interest rates

- There is an inverse relationship between the quantity of money demanded and the interest rate.
- Four important definitions of Australia's money supply are the **Monetary Base**, **M1**, **M3** and **Broad Money**.
- It is the interaction of the demand for and supply of money in financial markets, which determines interest rates. A central bank can exert a strong influence over interest rates and, by so doing, can influence aggregate demand, economic activity and inflation.

3. Develop an appreciation of the key elements of a modern financial system

- A country's **financial system** facilitates the transfer of resources from those with resources excess to their immediate needs (savers) to those with immediate needs for such resources (borrowers).
- At the centre of the financial system stands the nation's central bank and its banks. The central bank is responsible for a nation's monetary policy, which it implements by appropriate buying and selling of **Commonwealth Government Securities**.

- Banks occupy a crucial position in our community and their solvency is virtually guaranteed by the careful and comprehensive supervision of its activities and the particular asset structure it is required to maintain.
- As bank deposits are considered money, banks can quite literally create money 'at the stroke of a pen'.
- Other important elements of the financial system include the *money markets*, the *capital markets* (bond and share markets) and the *foreign exchange market*.

Study questions and problems

1 What backs the Australian dollar (or euro, yen or pound)? Include the distinction between commodity money and fiat money in your answer.

2 Distinguish between M1 and M3. Which is the broader definition of money?

3 What does the Monetary Base comprise and what role does it play in the financial system?

4 What does the *Reserve Bank of Australia Act (1959)* stipulate are the major purposes of the Reserve Bank of Australia?

5 Consider this statement: 'Banks do not create money because this is the RBA's responsibility'. Do you agree or disagree? Explain.

6 Suppose you find $10 000 under the floorboards in your house and deposit it in your bank. If the bank's required reserve ratio is 6 per cent, what is the amount of highly liquid assets – like currency and CGS – which the bank needs to hold in relation to this deposit? What can be done with the remainder?

7 Explain how money multiplier works. Given everything else held constant, calculate the money multiplier if the reserve ratio is 10 per cent and 20 per cent. What is the relationship between the money multiplier and the reserve ratio?

8 Consider the following conditions:

a In your new job you are paid twice as much as your old job.

b The rate of interest on bonds and other financial assets rises.

c An automatic teller machine (ATM) is installed next door and you have a debit card.

d Bond prices are expected to fall.

Would you decide to increase or decrease your average holding of money (i.e., cash and/or cheque deposit balances)? Which of the three motives for holding money is involved in each case?

9 Suppose a bond pays an annual coupon interest of $3000. Compute the yield per year that a bondholder will earn if the bond is purchased at a market value of $120 000, $100 000 and $75 000. The calculations lead you to what conclusion?

10 Using the demand and supply schedule for money shown below, do the following:

a Graph the demand for and the supply of money curves.

b Determine the equilibrium interest rate.

c Suppose the RBA increases the money supply by $5 billion. Show the effect in your graph and describe the money market adjustment process that is likely to follow. What is the new equilibrium rate of interest?

Interest rate (%)	Demand for money (billions of dollars)	Supply of money (billions of dollars)
8	20	30
7	25	30
6	30	30
5	35	30
4	40	30
3	45	30
2	50	30

Answers to 'You're the Economist'

Are debit cards money?

Debit cards facilitate the exchange of goods and services and debit card statements serve as a useful summary of expenditures. Furthermore, unlike credit cards, debit cards do not represent an extension of credit. However, like cheques (which also facilitate exchange, do not represent an extension of credit and whose bank statement provides a useful summary of expenditures), debit cards are themselves not money. What is money are the deposits in the accounts from which the card facilitates the withdrawal of funds to pay for transactions. If you said that, just like credit cards and cheques, debit cards are also not money because they do not satisfy the three functions required for money, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 The more difficult it is to convert an asset directly into goods and services without loss, the:
- **a** less secure it is.
- **b** more liquid it is.
- **c** less liquid it is.
- **d** more secure it is.

2 Which of the following is a characteristic of money (there may be more than one)?
- **a** It is always backed by something of high intrinsic value, such as gold or silver.
- **b** It is generally acceptable as a medium of exchange.
- **c** It provides a way to measure the relative value of goods and services.
- **d** It allows for saving and borrowing.

3 In the table below, what is the money supply, defined by M1, equal to?

Currency held by the non-bank public ($, billion)	Current deposits at banks ($, billion)	Other deposits at banks and other deposits held by other deposit-taking institutions ($, billion)
60	240	1500

- **a** 60 billion
- **b** 240 billion
- **c** 300 billion
- **d** 1560 billion

4 In the table in question (3) what is the money supply, as defined by M3, equal to?
- **a** 1500 billion
- **b** 1800 billion
- **c** 1740 billion
- **d** 1200 billion

5 In the table below, what is the Monetary Base equal to?

Currency circulating in the private sector ($, billion)	Current deposits at banks ($, billion)	Bank deposits at the RBA ($, billion)
65	240	35

- **a** 305 billion
- **b** 275 billion
- **c** 340 billion
- **d** 100 billion

6 Which definition of the money supply includes balances in your credit and debit cards?
- **a** M1.
- **b** Broad Money.
- **c** M3.
- **d** None of the above.

7 Anne is a university student, she always keeps $50 in her mobile phone cover, just in case she needs cash. Which of the following is her motive for holding money?
- **a** Transactions motive.
- **b** Speculative motive.
- **c** Profit motive.
- **d** Precautionary motive.

8 The action by the central bank to alter interest rates and thereby investment spending, which in turn brings about aggregate demand shifts, affecting prices, real GDP and employment, is known as:
- **a** macroeconomic policy.
- **b** wealth or real balances management.
- **c** the money creation mechanism.
- **d** monetary policy.

9 Which of the following is/are a function of the RBA?
- **a** Promoting the safety and efficiency of the payments system.
- **b** Supervising and regulating banks.
- **c** Implementing monetary policy.
- **d** All of the above.

10 If a bank has total deposits of $100 000, with $8000 in vault cash set aside to meet reserve requirements, and it is not holding any excess reserves, its minimum liquidity reserve ratio is:
- **a** $8000.
- **b** 0.8 per cent.
- **c** 8 per cent.
- **d** 1.08 per cent.

Macroeconomic policy I: monetary policy

16

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand what monetary policy is, what its goal is, and to look at a widely held view of how policymakers believe that changes in monetary policy affect the economy.
2 Understand how central banks implement modern monetary policy.
3 Understand that there are two alternative views as to how monetary policy adjustments actually affect the economy.
4 Understand the 'rules' versus 'discretion' debate in relation to the most appropriate role for monetary policy.

Getty Images/Bloomberg

In this chapter and the next, we will build on the material of the previous five chapters and provide an introduction to the two critical arms of macroeconomic policy. This chapter deals with monetary policy and **Chapter 17** deals with fiscal policy.

You will recall that in **Chapter 15** we briefly introduced the notion of 'monetary policy', which is the term used to describe actions taken by the central bank to influence interest rates, aggregate demand, economic activity and prices in the economy. In this chapter, we will be discussing monetary policy in considerable detail. We will elaborate upon the goals for monetary policy; discuss mechanisms by which changes in monetary policy affect the economy; explain how a central bank implements modern monetary policy; and, finally, on the quite fundamental issue of the most appropriate role for monetary policy, provide an introduction to the 'rules' versus 'discretion' debate.

1 What is monetary policy and how is it thought to affect the economy?

In this section you will learn about the nature of monetary policy, the goal of monetary policy in a typical modern economy, and a quite widespread and widely accepted view as to the mechanism by which monetary policy affects prices, output and employment in the economy.

In modern economies, a nation's central bank has responsibility for the determination and implementation of monetary policy and so any discussion of what monetary policy is and how it is thought to affect the economy must first begin with a description of the goal of monetary policy.

The goal of monetary policy

Goal of monetary policy
In Australia, the goal of monetary policy is to keep consumer price inflation between 2 and 3 per cent, on average, over the medium term.

The **goal of monetary policy** can – and does – vary from country to country. In the case of Australia, it is to keep consumer price inflation between 2 and 3 per cent, on average, over the medium term. 'Inflation-targeting' is not unique to Australia and it is now quite a common approach of central banks around the world – although different central banks define their inflation target in somewhat different ways.

However, as noted in the previous chapter, the *Reserve Bank Act 1959* stipulates that the Reserve Bank of Australia (RBA) has a duty to the Australian people to maintain the stability of the currency (keep inflation low), to maintain full employment, and to ensure the economic prosperity and welfare of the Australian people. This sounds as though the RBA's ongoing goals of monetary policy should be three, not one!

On this, the RBA believes – along with most macroeconomists – that, by keeping inflation low and steady, it can assist with the achievement of full employment and high economic growth over the medium to long term. By doing this it prevents inflation from taking off and thereby avoids all the attendant problems associated with high – and usually highly variable – inflation, which were outlined in **Chapter 13**. With inflation low and stable, the RBA believes that economic growth will be higher – and unemployment lower – over the longer term than if it were to let go of its inflation target to try to promote economic growth to reduce short-term unemployment actively. By achieving its inflation target, the RBA therefore believes – and the government accepts – that it will best achieve all three goals outlined in the original Act.

Given that the goal of monetary policy is the maintenance of annual inflation at between 2 to 3 per cent over the medium term, on average, in the following sections, we will elaborate upon a widely accepted view of how changes in monetary policy affect inflation and the macro economy generally.

The monetary policy transmission mechanism

Monetary policy transmission mechanism
The central bank operates in the financial system to influence interest rates in the desired direction, which in turn influence aggregate demand, thereby leading to changes in prices, real GDP and employment.

One view of the **monetary policy transmission mechanism** is that the central bank operates in the financial system to influence interest rates in the desired direction, which in turn influence aggregate demand, thereby leading to changes in prices, real GDP and employment. We will elaborate on this transmission mechanism in the following sections.

Assume the economy is operating above full employment and inflation has started to take off (from previously low and stable levels), which the central bank detects and wants to act upon. To do this, it would pursue a contractionary monetary policy by operating in the financial system to reduce the available quantity of liquidity and thereby bring about a general increase in interest rates above their current levels. The higher interest rates would act to reduce private-sector demand for investment goods and consumer durables.

In addition, the higher domestic interest rates may increase demand for the currency in international foreign exchange markets – as a result of the higher relative return available on domestic financial assets – thereby possibly increasing the international value of the currency (known as *currency appreciation*). If international prices of products exported are set in one of the world's major currencies like the USD, yen or euro, the exporters have limited market power to influence the world price of their products. The higher value of the currency means that, for unchanged international prices of their products, exporters will receive less for their products (in their currency), thus squeezing their profit margins and impacting negatively on their business. At the same time, those local firms competing for sales with imports will find they become disadvantaged owing to the lower prices – in domestic currency terms – of those imported products. These two forces coming from an appreciated currency will further act to dampen domestic economic activity by retarding net exports ($X - M$).

This reduced demand would also have multiplier effects through the rest of the economy so that overall aggregate demand would reduce. The reduction in aggregate demand would slow down economic activity and reduce the rate of inflation.

Can monetary policy target sectors or sections in the economy?

Economics and ethics

A noteworthy feature of monetary policy is that it may be regarded as quite a 'blunt' policy instrument in the sense that it cannot be used to impact very narrowly on particular sections of the community or sectors of the economy. If the RBA engineers interest rates are up because it is concerned about current and future inflation, the higher borrowing costs will impact on those with existing loans and also those deciding whether to take out new loans.

In particular, for those who have existing loans – and variable interest rate commitments – such interest rate rises can have very substantial impacts on their cash flows, causing some households to experience extreme hardships and some businesses perhaps to even go into bankruptcy. However, it could well be that many of those who experience financial stress may not have contributed directly to whatever may have been the root causes of the inflationary circumstances that led to the central bank's monetary tightening. For example, asset price booms, particularly if the price rises seem out of step with underlying economic fundamentals (and therefore are probably unsustainable), and if they lead to substantial increases in aggregate debt levels in a community, can sometimes factor in a central bank's considerations in setting monetary policy.[1] To the extent that such developments may contribute to a particular monetary tightening, everyone will be affected by the higher interest rates irrespective of whether they were direct participants in the asset price boom.

Another example of this might be where strong economic activity in some economic sectors (e.g., mining) leads to inflation pressure, and the central bank is led to tighten monetary policy. The resulting interest rate rises will impact on everyone, including those operating and working in other sectors where things may not be quite so robust, and put considerable pressure on their household and business budgets.

Thus, while there are no ethical considerations at play, in this case, some may well question the 'social justice' of such outcomes.

[1] **Such asset price booms can become a concern to a central bank for several reasons. If the associated wealth increases lead to increased aggregate demand and price rises for goods and services, eventually general inflation may be affected. Related to this, if the asset price rises turn out to be largely speculative and lead eventually to a correction, this can have very serious damaging effects on economic activity and employment. The central bank may consider it in the best interests of the community to act pre-emptively to stave off such an outcome. In practice, most central banks, including the RBA, monitor such developments very carefully but are equally very careful and reticent about intervening in asset price booms.**

Similarly, if the nation's central bank held the view that aggregate demand needed a stimulus, then it would pursue an expansionary monetary policy by operating in the financial system to increase the available quantity of liquidity, and thereby bring about a general decrease in interest rates below their current levels. The lower interest rates would act to increase private-sector demand for investment goods and consumer durables, as well as potentially providing a stimulus to net exports ($X - M$) if the lower domestic interest rates resulted in a lower-valued or *depreciated currency*. This increased demand would also have multiplier effects through the rest of the economy so that overall aggregate demand would increase.

Let's now have a look at expansionary and contractionary monetary policy in a little more detail.

Expansionary monetary policy: the central bank reducing interest rates

Part (a) of **Exhibit 16.1**, is a stylised depiction of how an expansionary monetary policy by the central bank works to bring about a fall in the equilibrium rate of interest. As with **Chapter 15**, we use the term 'money' to refer to those financial assets in the financial system that are relatively the most liquid and can be used immediately in the exchange of goods and services or in settlement of debts.

Our analysis begins at point E_1, with the supply of money (or liquidity) to the financial system at $20 billion and equal to the quantity of money demanded, and with the equilibrium interest rate at 6 per cent. Now suppose

the central bank operates in financial markets to inject $10 billion additional liquidity into the financial system. While the exact mechanics of how the central bank influences liquidity conditions in the financial system is deferred to the next section, the impact of the central bank's action is to create $10 billion surplus liquidity at the prevailing 6 per cent interest rate.

How will financial market participants react to this additional and excess liquidity in their possession? They will seek to buy interest-bearing bonds (remember that the use of the word 'bonds' here – as in **Chapter 15** – refers to all those financial assets that are income-producing but are less liquid than 'money'). The rush to purchase bonds drives their prices higher and their implied yield – the interest rate – falls. As the interest rate falls, there is a greater *willingness* to hold larger money balances – or, stated differently, there is an increase in the quantity of money demanded until the new equilibrium at E_2 is reached. At the lower interest rate of 4 per cent, the opportunity cost of holding money is also lower, and the imbalance between money demand and money supply disappears.

Contractionary monetary policy – the central bank increasing interest rates

Part (b) of **Exhibit 16.1**, illustrates how the central bank can put upward pressure on the interest rate with contractionary monetary policy. Beginning at point E_1, the money market is in equilibrium at an interest rate of 4 per cent. This time the central bank shrinks the liquidity available to the financial system, depicted in the exhibit by the money supply decreasing from $30 billion to $20 billion. At the initial equilibrium interest rate of 4 per cent, this decrease in the liquidity causes a shortage of $10 billion.

Participants in the financial markets wish to hold a higher quantity of highly liquid money than is available, and to seek to satisfy this wish they move to sell their bonds for money. This selling pressure lowers bond prices, causing the implied rate of interest to rise. Once the interest rate reaches the new equilibrium level of 6 per cent (point E_2) the opportunity cost of holding money balances is such that the $20 billion available liquidity in supply in the financial system is willingly held.

Now let's go through in some detail – and represent graphically – how these changes in interest rates flow on into changes in aggregate demand.

How monetary policy affects prices, output and employment using the *AD–AS* model

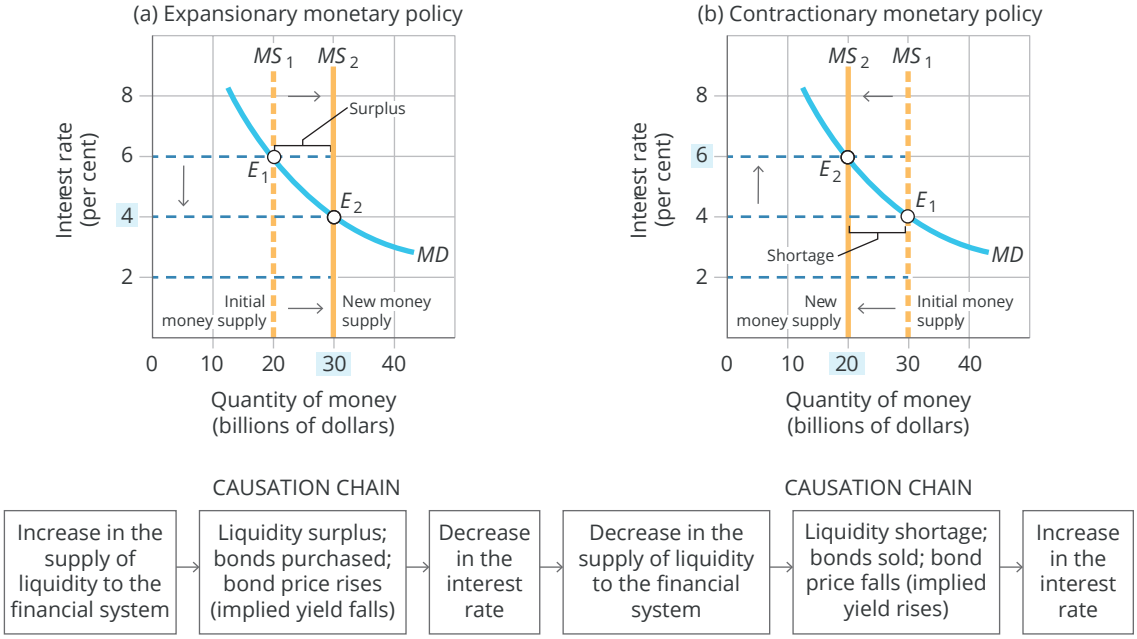
The next step in our journey is to understand how changes in monetary policy impact on the macroeconomy. **Exhibit 16.2** illustrates the causation chain linking monetary policy and economic performance. As mentioned earlier, this is referred to as the monetary policy transmission mechanism, and in what follows below, we elaborate further on the details of this mechanism.

So far, we have sketched out how monetary policy can affect the level of interest rates prevailing in the economy. How do changes in interest rates affect aggregate demand? Begin with part (a) of **Exhibit 16.3**, which is identical to part (a) of **Exhibit 16.1** and represents the financial system.

We assume that the central bank wishes to bring about a reduction in interest rates; that is, it wishes to pursue an expansionary monetary policy. To bring this about it moves to increase the supply of liquidity available to the financial system above what is demanded. In the exhibit, this is represented by a shift in the money supply from $20 billion ($MS_1$) to $30 billion ($MS_2$), and the equilibrium interest rate falls from 6 per cent to 4 per cent.

We know from our discussion in **Chapter 14** that, with lower interest rates, *ceteris paribus*, firms will be more willing to spend on structures, plant and equipment. With lower interest rates, people may also be more willing to purchase newly constructed houses and apartments, which, as you will recall from **Chapter 11**, are also regarded as an investment.

Exhibit 16.1 The effects of monetary policy adjustments on interest rates



CAUSATION CHAIN

Increase in the supply of liquidity to the financial system → Liquidity surplus; bonds purchased; bond price rises (implied yield falls) → Decrease in the interest rate

CAUSATION CHAIN

Decrease in the supply of liquidity to the financial system → Liquidity shortage; bonds sold; bond price falls (implied yield rises) → Increase in the interest rate

In part (a), the central bank increases the supply of liquidity in the financial system from \$20 billion (MS_1) to \$30 billion (MS_2). At the initial interest rate of 6 per cent (point E_1), there is an excess of \$10 billion in liquid financial assets beyond the amount market participants wish to hold. They react by buying bonds, thereby putting upward pressure on the price of bonds. As the price of bonds increases, their implied yield (interest rate) falls until a new lower equilibrium interest rate is reached at 4 per cent (point E_2).

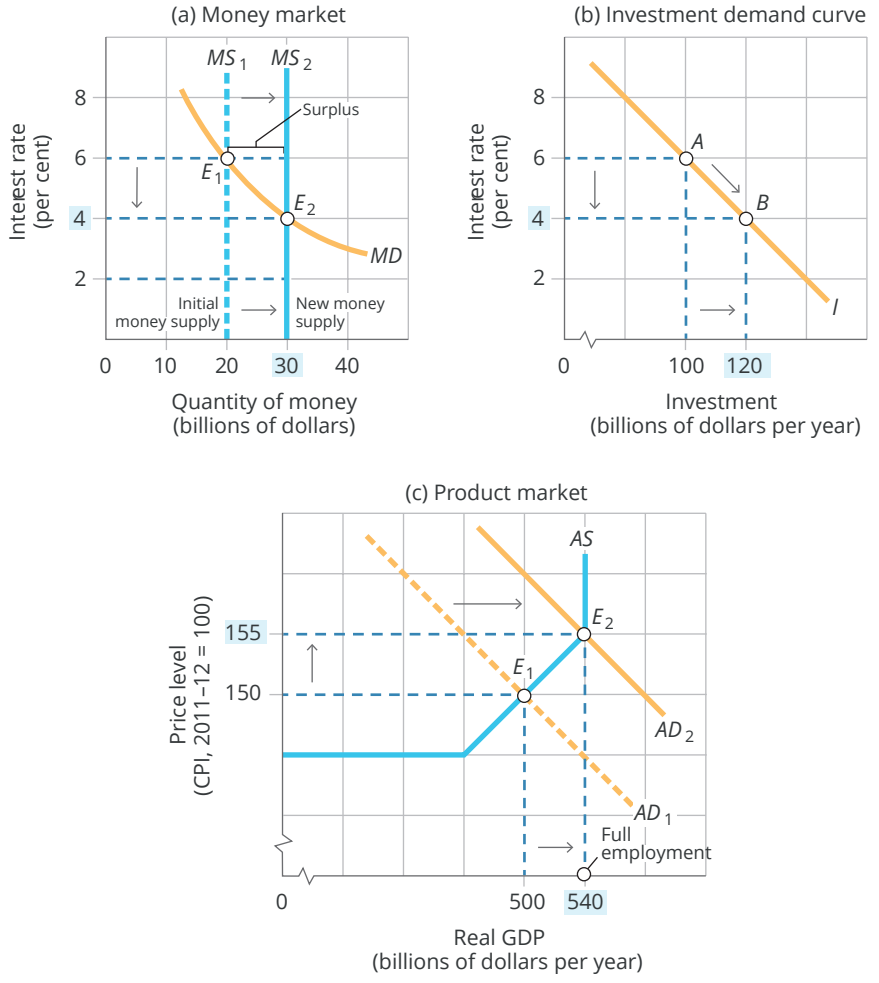
The reverse happens in part (b). The central bank decreases the supply of liquidity to the financial system from \$30 billion (MS_1) to \$20 billion (MS_2). Beginning at 4 per cent (point E_1), market participants wish to hold \$10 billion more than is now available. They react by selling bonds, thereby putting downward pressure on the price of bonds. As the price of bonds decreases, their implied yield (interest rate) rises until a new higher equilibrium interest rate is reached at 6 per cent (point E_2). At the higher interest rate, the reduced supply of liquidity of \$20 billion is now willingly held by market participants, and the initial imbalance disappears.

Exhibit 16.2 The basics of the monetary policy transmission mechanism

Change in monetary policy → Change in the supply of liquidity to the financial system → Change in interest rates → Change in investment, interest-sensitive consumption and net exports → Change in aggregate demand → Change in prices, real GDP and employment

A widely accepted view of the transmission mechanism is that changes in monetary policy affect interest rates, which then impact investment and interest-sensitive consumption spending, as well as net exports. In turn, overall aggregate demand shifts, and this affects prices, real GDP and employment.

Exhibit 16.3 **The effect of expansionary monetary policy on aggregate demand**



In part (a), the existing supply of liquidity available to the financial system is represented by MS_1, and the equilibrium rate of interest is 6 per cent. The equilibrium point in the money market changes from E_1 to E_2 when the central bank increases the supply of liquidity to MS_2. As people attempt to buy income-yielding bonds with the additional highly liquid money, this causes an increase in their price and a commensurate reduction in their yield (i.e., interest rates fall). As interest rates fall to a new equilibrium level of 4 per cent, people are then satisfied to hold the $30 billion in supply rather than seeking to exchange it for income-yielding bonds due to the lower opportunity cost of holding money.

The fall in the rate of interest shown in part (b) causes a movement downward along with the investment–demand curve from point *A* to point *B*. Thus, the quantity of investment spending increases from $100 billion to $120 billion. In addition, interest-sensitive consumption demand will also increase, and possibly net exports to the extent the exchange rate depreciates as a result of the lower domestic interest rate. Furthermore, these increases in spending will, through the spending multiplier, produce a further increase in aggregate demand.

In part (c), this overall expansion in aggregate demand is represented in the exhibit as a shift outward from AD_1 to AD_2. As a result, the aggregate demand and supply equilibrium in the product market changes from E_1 to E_2 and the real GDP gap is eliminated. The price level also changes from 150 to 155.

This is represented graphically in part (b) of **Exhibit 16.3**, where you can see that the falling rate of interest causes an increase in the quantity of investment spending from $100 billion to $120 billion. Stated another way, there is an increase in the quantity of investment goods demanded (I), which, you will recall from **Chapter 14**, is a component of total spending or aggregate demand. The investment demand curve shows the amount that businesses and households spend on investment goods at different possible rates of interest.

Furthermore, as noted earlier, in addition to the increased investment spending, shown in part (b), interest-sensitive components of consumption spending will also increase with lower borrowing costs (not explicitly depicted in the exhibit but nonetheless they would also occur). Also, again as noted earlier, net exports will be stimulated to the extent that lower domestic interest rates depreciate the currency.

In part (c) of **Exhibit 16.3**, we use the aggregate demand and aggregate supply analysis developed earlier in **Chapter 14**. Begin at point E_1, with a real GDP per year of $500 billion and a price level of 150. Now consider the link to the change in the supply of liquidity to the financial system and the subsequent reduction in the interest rate.

The initial increase in spending resulting from the fall in the interest rate described above also then works through the *spending multiplier* to increase aggregate demand by more than the initial increase (if necessary, refer back to the discussion in **Chapter 15** on the multiplier). This overall increase in aggregate demand is represented in part (c) of **Exhibit 16.3** by a shift in the aggregate demand curve from AD_1 to AD_2. At the new equilibrium point, E_2, the level of real GDP rises from $500 billion to $540 billion (a $40 billion increase) and full employment is achieved. In addition, the price level rises from 150 to 155.

Exhibit 16.3 can also be used to understand how a contractionary monetary policy would work. In part (a), imagine the supply of liquidity available to the financial system started at MS_2. In the contractionary case, the supply then shifts inward from MS_2 to MS_1, causing the equilibrium rate of interest to rise from 4 per cent to 6 per cent. The central bank's 'tight' monetary policy causes the level of investment spending to fall from $120 billion to $100 billion on account of the increased level of interest rates. In addition, the higher interest rates will also act to curb interest-sensitive consumer durable spending and, to the extent the currency appreciates due to the higher domestic interest rates, net exports will also be dampened. All of this leads to a decrease in the equilibrium level of real GDP per year from $540 billion to $500 billion with the price level falling from 155 to 150.

Note that, in reality, the price level is unlikely to actually drop as a result of such a monetary tightening. What will happen is that the rate of increase in the price level (the rate of inflation) will fall back to an acceptable level (say, from +3 per cent to +2.5 per cent). The *AD–AS* analysis of **Exhibit 16.3** nevertheless captures quite well the flavour of the economic adjustments involved.

The bottom line, then, is that expansionary monetary policy amounts to the central bank reducing interest rates by increasing the supply of liquidity available to the financial system. The lower interest rates stimulate investment and consumer durable demand – and possibly also net exports if the currency depreciates – and this increases aggregate demand. Thus, expansionary monetary policy is represented in the *AD–AS* framework as an outward shift in the *AD* curve.

Analogously, contractionary monetary policy is represented by an inward shift in the *AD* curve.

In summary

- The central bank implements a change in **monetary policy** by changing the supply of liquidity available to the financial system to bring about a change in domestic interest rates.
- The change in interest rates affects aggregate demand and this can be represented in the *AD–AS* framework by a shift in the *AD* curve.
- Changes in aggregate demand then lead to changes in real GDP, employment and prices.

2 Modern monetary policy implementation: the case of Australia

In this section, you will be developing an understanding of how monetary policy has been conducted in Australia since the early 1990s. While the discussion focuses on Australia, the mechanics of monetary policy adjustments described here is also a good representation of what is done by central banks in most developed countries around the world.

Statement on the Conduct of Monetary Policy
Signed in 1996, it encapsulates the Australian government's agreement that the RBA should have the goal of keeping inflation on average to between 2 and 3 per cent per annum over the course of the business cycle. In 2016, the language was adjusted a little to that of the goal being to keep consumer price inflation between 2 and 3 per cent, on average, over the medium term.

As has been mentioned before, the RBA unofficially adopted an inflation target of 2 to 3 per cent around 1993, and in 1996 this was made official with the co-signing by the Treasurer and the RBA Governor of the **Statement on the Conduct of Monetary Policy**. In pursuing this agreed goal, the RBA enjoys great independence from the government in its operation of monetary policy. Other central banks around the world – but not all – have a similar explicit inflation target as their key monetary policy goal. For example, the European Central Bank (ECB; the central bank for the Eurozone area of Europe) aims to keep inflation rates below, but close, to 2 per cent, over the medium term, while the US Federal Reserve System (the Fed) seeks to maintain inflation at 2 per cent over the medium term.[1]

Overnight cash rate (cash rate)
The interest rate that large-scale borrowers – including banks – in the financial markets must pay to borrow funds overnight.

Modern monetary policy in Australia is implemented by the RBA seeking to maintain the **overnight cash rate** at a pre-announced level. The overnight cash rate (or just 'cash rate' for short) is the interest rate that borrowers in the financial markets pay to borrow funds for as short a period as overnight. It is the RBA's instrument of monetary policy that it uses to achieve its inflation target goal. In particular, this is the interest rate that banks charge one another for borrowing funds to cover their short-term needs. As such, its value influences bank wholesale and retail interest rates and all other interest rates in the financial system. For instance, a change in the cash rate will very quickly flow through to other short-term interest rates in the financial markets (like the interest rate on bank-accepted bills, introduced briefly in the previous chapter). If maintained, changes to short-term financial market interest rates will filter through to other interest rates, such as home mortgage rates, prime lending rates (the rates banks charge their best corporate borrowers), credit card rates, personal loan rates and so on. Changes in these latter rates, if maintained, will affect consumption, investment activity and aggregate demand as has been previously described.

The value of the cash rate, therefore, influences all interest rates throughout the economy. It is by this influence on the interest rates in the broader economy that the RBA seeks to affect aggregate demand and influence the rate of inflation. In passing, it is worth emphasising that the RBA does not literally 'set' the cash rate. Instead, it continuously conducts operations in the financial markets on a daily basis to keep the cash rate as close as possible to its announced target level.

The term 'cash rate' is also commonly used elsewhere in the world; for example, in New Zealand and the UK. In the US, it is called the 'federal funds rate' (FFR), in Japan, the 'overnight call rate' and in Canada the 'overnight rate'. For the ECB, the closest equivalent is probably the 'Repo rate'.

Open-market operations
The buying and selling of government securities by the central bank to the private sector financial markets to impact on the Monetary Base and the 'cash rate'.

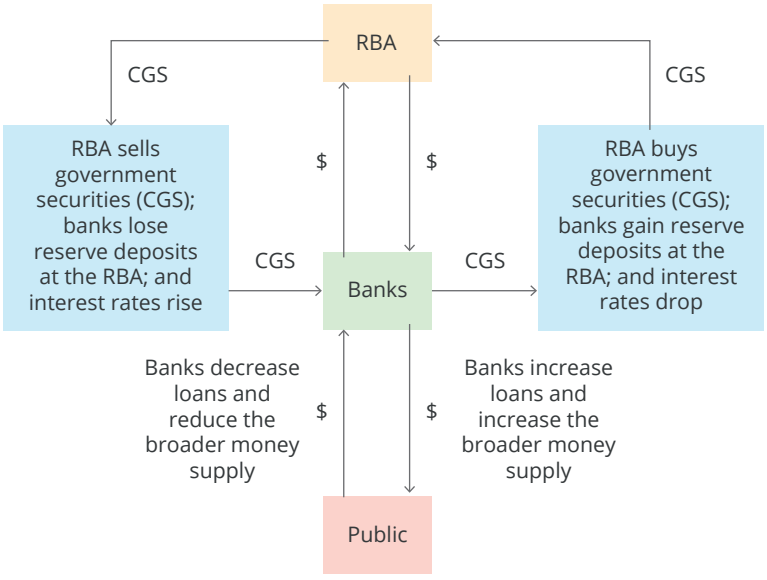
To influence the cash rate, the RBA, like all central banks around the world, carries out **open-market operations** in the financial markets to alter the supply of primary liquidity available to the financial system (the Monetary Base). Open-market operations consist of the central bank buying and selling Commonwealth government securities (CGS – refer to **Chapter 15** for a discussion of CGS) from the private sector – usually banks – for 'cash'.

In settling such transactions between itself and a buyer or seller of securities, the RBA simply electronically credits or debits their exchange settlement account (refer to the previous chapter). Deposits in these accounts are called 'exchange settlement money' and comprise an important component of the Monetary Base. Also, it may be noted that, operationally, the RBA usually carries out its open-market operations by the buying/selling of short-term CGS on a repurchase (Repo) basis, whereby it agrees to sell/buy back the security at a later date, with the difference in price representing the rate of interest earned/paid.

[1] Refer to the Fed website for further details, https://www.federalreserve.gov/faqs/economy_14400.htm, accessed 28 May 2018.

The schematic in **Exhibit 16.4** depicts how open-market operations by the central bank affect the Monetary Base, broader definitions of money supply and interest rates.

Exhibit 16.4 Open-market operations



When the RBA buys Commonwealth government securities (CGS) from banks, it credits the exchange settlement accounts of those banks that sell them and this increases their reserve deposits. This increase in Monetary Base puts downward pressure on interest rates. Banks are also likely to use these extra reserve deposits to make loans, which, through the credit/money multiplier, will work to expand the broader money supply. This is all depicted in the right-hand half of the figure. When the RBA sells CGS to banks, it debits the exchange settlement accounts of those banks that bought them and this decreases their reserve deposits. This decrease in Monetary Base puts upward pressure on interest rates. The reduction in the reserve deposits of banks will also diminish their capacity to lend and, as a consequence, the broader money supply will tend to decrease (all depicted in the left-hand half of the figure).

To illustrate an open-market operation, suppose the RBA has evidence that the underlying inflation rate is – or soon will be – creeping up towards the top end of its acceptable range (3%). What will it do?

If the RBA Board feels that this higher rate of inflation is likely to persist, it will most likely decide that it is necessary to tighten financial conditions. To do this the RBA will operate in the financial markets to ensure the quantity of Monetary Base available in the financial system is less than what is needed and/or demanded by financial market participants. This will exert upward pressure on the cash rate and hence put upward pressure on interest rates generally. The extent to which it will seek to raise the cash rate is a matter of judgement. However, in reality, it usually does so in a series of small monthly steps of 0.25 per cent, as more and more information on current and likely future economic conditions comes to hand.

Operationally, the RBA undertakes the desired monetary policy tightening by offering to sell short-term government securities from its portfolio to banks and other financial market participants at prices implying a higher rate of interest. Suppose a bank buys some of the securities on offer. In payment for them, the RBA will immediately debit the bank's exchange settlement account and, by doing this, will immediately reduce the Monetary Base in the financial system.

Analyse the issue **Applicable concept: monetary policy implementation**

Actual monetary policy adjustments

In January 2020, the Monetary Base was around $116.8 billion. Over the period February 2020 to July 2020, the overnight cash rate was reduced by the RBA from 0.75 per cent down to 0.25 per cent. In July 2020, the Monetary Base had expanded to $178.1 billion (Case 1).

But, during an earlier period from December 2009 to May 2010, the RBA managed the cash rate up from 3.75 per cent to 4.5 per cent and during this time the Base decreased from $57.1 billion to $53.1 billion (Case 2).

Interpret each of these cases in terms of changes in the supply of and demand for Monetary Base in the financial system, and illustrate each with an appropriate demand and supply diagram.

By way of a specific example, suppose the funds needed to pay for the securities amount to $100 million. As soon as the RBA debits the $100 million from the exchange settlement account of the purchasing bank(s), these funds have been taken out of the private sector financial system and are no longer part of its liquidity base. Since the banks' deposits with the RBA are part of the Monetary Base and since the deposits are extinguished in payment for the securities, the Monetary Base of the economy is reduced by $100 million. Since the Base has been reduced, we can also expect that, through the action of the credit multiplier (refer to **Chapter 15**), M1 and the other broader monetary aggregates may also reduce over time as banks adjust their lending.

In this monetary policy tightening example, as the overnight cash rate rises by the open market operations of the RBA, other short-term deposit rates in the financial markets will also rise. The higher short-term interest rates prevailing in the financial system will induce investors to start shifting their funds out of longer-maturity securities with the result that their interest rates will – other things being equal – also tend to rise along with the rises in the short-term rates. Thus, the entire spectrum of interest rates in the economy is affected by the RBA's actions.

Very interestingly, this was the scenario facing the RBA during the second half of 2007 when it made a now famous and historic adjustment to monetary policy – an increase in the cash rate by 0.25 per cent – in the middle of the federal election campaign of that year. Given the tremendous political sensitivity of earlier rises in interest rates in the lead-up to the 2007 election, in raising interest rates in the middle of the campaign itself, the RBA indeed demonstrated – in a quite dramatic fashion – its political independence.

In **Chapter 19** we will outline the details of the very dramatic expansionary monetary policy actions taken by the RBA – and other central banks around the world – as a response to the unfolding global financial and economic crisis (GFEC) of 2008–09 and during the COVID-19 pandemic. There we will also outline the continuing monetary policy actions taken by central banks in the aftermath of the economic crisis and COVID-19 pandemic.

You're the economist **The money supply curve when the RBA targets an interest rate**

In the previous chapter, you learnt that it is the interaction of the demand and supply of money that determines interest rates. As you learnt above, in implementing monetary policy today, the overnight cash rate is the interest rate that the RBA focuses on maintaining at some desired target level. Just as with any market interest rate, the value of the cash rate is determined by the interaction of the demand and supply of the relevant funds, namely the demand for and supply of the Monetary Base available to the financial system.

Assume a downward-sloping money demand curve and imagine that money demand in financial markets is fluctuating, as it would be from period to period. You would represent this shifting demand by shifts (inward and outward) in the demand curve. In these circumstances, if the RBA is seeking to control the cash rate at some target level, what would be the effective shape of the money supply curve?

In summary

- Australia's RBA has an inflation target as the goal of monetary policy.
- In pursuing this goal, the RBA seeks to maintain a key short-term interest rate, the **overnight cash rate**, at a desired policy level.
- Changes in the cash rate are affected by the RBA carrying out **open-market operations** – buying or selling Commonwealth Government Securities (CGS) – in financial markets.
- These actions impact on the financial system's Monetary Base, and, by so doing, the RBA maintains the cash rate at the desired level.

3 An alternative view of the monetary policy transmission mechanism

As outlined in the above discussion, monetary policy acts through changing interest rates first before affecting aggregate demand, and then consequently affecting prices, real GDP and employment. In that characterisation, changes in interest rates are critical.

Another school of thought, known as **Monetarism**, sees monetary policy operating more widely than this, and argues that changes in the money supply much more directly determine changes in prices, real GDP and employment. While acknowledging that interest rates certainly play a role in the monetary policy transmission mechanism, Monetarism, to an extent, de-emphasises the interest rate–investment linkage.

It is important to present some discussion of this view of the monetary policy transmission mechanism since the **quantitative easing** actions of the world's major central banks – the US Fed, the Bank of Japan, the Bank of England and the ECB – can be interpreted through this prism.

As will be discussed in more detail in **Chapter 19**, as a policy response to the GFEC and the COVID-19 pandemic, most of the world's central banks dramatically reduced their respective interest rate instruments to historically low levels – some to effectively zero or near zero – and despite this, domestic demand in their economies remained stubbornly weak and their economies did not convincingly recover from the crisis. It seemed that the widely held characterisation of the monetary policy transmission mechanism operating through interest rates was not working – or at least not working as intensely as had been hoped or anticipated.

Thus, when this occurred, these monetary authorities resorted to vigorously expanding the Monetary Base of their financial systems. They did this by buying large quantities of long-dated government and corporate bonds from financial institutions in the financial markets to provide stimulus to their economies, to avoid the possibility of falling back into a recession.

To understand the intellectual underpinning of this approach, we start with the so-called *equation of exchange*.

Monetarism
The theory that changes in the growth of the money supply directly determine changes in prices, real GDP and employment.

Quantitative easing
The process whereby the central bank expands its balance sheet by buying long-dated government and corporate bonds from the financial sector. In paying for these purchases the central bank expands the Monetary Base. The primary goal of the strategy is to significantly expand the available liquidity in the financial system.

The equation of exchange

Monetarists put the spotlight firmly on the money supply. They argue that the condition of the economy is related directly to monetary growth. If the money supply is expanding too much, higher rates of inflation will be likely. If it grows too slowly, prices may fall, or the unemployment rate may increase.

The easiest way to understand Monetarism is, to begin with, the so-called **equation of exchange**, which was developed in the 19th century. The equation of exchange is an accounting identity that states that the money supply times the velocity of money equals total spending. Expressed as a simple formula, the equation of exchange is written as:

$$MV = PQ$$

Equation of exchange
An accounting identity that states that the money supply times the velocity of money equals total spending.

Velocity of money
The average number of times per period a dollar of the money supply is spent on final goods and services.

Let's begin with the left-hand side of the equation (MV). M is the money supply (for argument's sake, say, M1) in circulation, and V represents the **velocity of money**. The velocity of money is the average number of times per time period (say, a year) a dollar of the money supply is spent on final goods and services.

To illustrate, assume an ultra-simple economy in which the money supply consists of just one crisp $20 note. Suppose this $20 is spent on a pizza and a drink at the local pizza shop. The owner of the pizza shop, in turn, takes the $20 and buys an economics book from the local bookstore to learn about quantitative easing.

At this point, a single $20 bill has financed $40 worth of total final spending. Moreover, as long as this $20 bill passes from hand to hand during, say, one week, the value of final sales will increase. For example, assume the $20 note in question travels from hand to hand five times. This means the velocity of money is five, and the left-hand side of the equation of exchange is expressed as:

$$\$20 \times 5 = \$100$$

The equation of exchange is an *identity* – true by definition – that expresses the fact that the value of what people spend is equal to, or exchanged for, what they buy. What people buy is nominal GDP or PQ. Nominal GDP is equal to the average selling price during the year (P) multiplied by the quantity of actual output of final goods and services (Q).

In the example of the simple economy, total spending, or PQ, equals $100. Note that the identity between MV and PQ only indicates what happens to the product of P and Q if MV increases. Although we know total spending (PQ) increases, we do not know whether the price level (P), the quantity of output (Q), or both, increase.

Consider a more realistic example. Suppose nominal GDP is $400 billion during some time period and M1 was $100 billion. How many times on average did each dollar of the M1 money supply have to be spent on final goods and services to generate this level of total final spending in the economy? Using the equation of exchange:

$$MV = PQ$$

$$\$100 \text{ billion} \times V = \$400 \text{ billion}$$

$$V = 4$$

Thus, each dollar (in M1) is spent an average of four times during the period in the purchase of final goods and services.

The quantity theory of money

The equation of exchange is converted from an *identity* to a *theory* by making certain assumptions. Suppose it can be argued that the velocity of money (V) and real output (Q) are fairly constant.

Quantity theory of money
The theory that changes in the price level (inflation) are directly related to changes in the growth of the money supply.

With V and Q being considered approximately constant, we have one of the oldest theories of inflation, called the **quantity theory of money**. The quantity theory of money states that changes in the price level (inflation) are directly related to changes in the supply of money.

Monetary policy based on the simple quantity theory of money is, therefore, founded on the view that changes in the money supply directly affect the price level. To illustrate, the equation of exchange is modified into the quantity theory of money by putting a bar (–) over V and over Q to indicate that they are assumed fixed or constant in value:

$$M \times \overline{V} = P \times \overline{Q}$$

Under these circumstances, what happens if the central bank increases the money supply by 10 per cent? The price level must also increase by 10 per cent! In short, in the context of the quantity theory of money, the cause of inflation is sometimes described as 'too much money chasing too few goods'.

Of course, the quantity theory of money is a very simple theory, which may have been a good approximation to the world of the 19th century. It may, however, be a less useful approximation to our present-day 21st-century world. For example, it abstracts from non-monetary factors – such as supply shocks from, for example, a hike in oil prices – that cause cost-push inflation. Moreover, its predictions about the direct relationship between money growth and prices will tend to unravel if V does not remain reasonably stable over time.

As far as the constancy of V is concerned, as noted above, the velocity of money will depend on the community's demand for money and, in times of very rapid financial innovation – such as, for example, began occurring in the 1980s and 1990s – the demand for money could be changing very substantially. As just one example, the widespread introduction of electronic funds transfer at point of sale (EFTPOS) facilities in Australia and elsewhere in the 1990s changed very dramatically the public's need and demand for 'walking around' currency.

Finally, we know that economies may, at times, be very far from full employment. (Recall discussions in previous chapters about recessions and the Great Depression.) In such circumstances, increases in money supply may mostly stimulate increases in output rather than price increases. Nevertheless, for economies at full employment and/or for economies experiencing extremely rapid money supply growth, the inflation predictions of the simple quantity theory may be quite realistic.

Of course, modern monetarists recognise the limitations of the overly simple original classical quantity theory of money for most economies in most circumstances in the world today. The empirical evidence indicates that the velocity of money is not constant over time and that the economy does not always operate at full employment. Therefore, while they would argue that M and P would not be expected to change exactly proportionally, they would nonetheless expect them to be reasonably correlated – especially over the longer term.

In sum, *monetarists argue that whilst velocity may not be unchanging; it is nonetheless reasonably predictable* in the short run. Suppose that the *predicted* velocity of money ($\bar{V}$) in the next year is five and the money supply increases by $10 billion this year. Monetarists would predict that nominal GDP would increase next year by about $50 billion ($M\bar{V}$).

Furthermore, while modern monetarists would predict that if the economy were near full employment, much of the increase in nominal GDP from some increase in M would simply be in rising prices, they would certainly argue that if the economy is far below full employment (say in a recession), most of the rise in total nominal GDP would probably be in real output rather than in rises in prices.

On this, in the case of the recent *quantitative easing* exercise by the world's major central banks post-2008, it most certainly would have been hoped that those institutions that sold bonds to the central banks, thereby finding themselves with significant additional liquidity, would have found alternative active uses for that liquidity. In particular, the powerful hope was that banks who had sold bonds would put their newly acquired liquidity to active economic purposes – such as new lending to businesses and households to stimulate economic growth and employment.

Monetarism and the monetary policy transmission mechanism

Although recognising that changing availability of money is likely to affect interest rates, monetarists do not place the same critical emphasis on the role of interest rate changes in the transmission mechanism. Instead, monetarists argue that when people find themselves with larger quantities of money on their hands than they had intended or expected, they will spend the money on a wide range of things and not just buy interest-bearing securities.

This buying pressure will manifest not only in increased prices of bonds (i.e., lower interest rates) but also in increased spending on a wide range of goods and services. Instead of working just through the interest rate to affect aggregate demand and the economy, changes in the money supply much more directly determine economic activity.

There is a famous thought experiment that illustrates the monetarist position. Suppose one morning that a community's inhabitants wake up to find $50 notes lying all over their yards – perhaps as a result of a midnight helicopter drop from some mysterious benefactor. What do they do?

Some will certainly, rush out to buy some financial securities, thereby exerting downward pressure on interest rates with consequential effects on investment demand. However, others will also quite likely want to buy cars, washing machines, clothes, food and so on. This generally increased buying pressure – as well as the lower nominal interest rates – could be expected to raise nominal GDP (and, in this parable, probably almost entirely as a result of rising prices rather than output!).

Conditional projection The implementation of monetary policy in Australia during the period 1976–85 consisted of a form of monetary targeting. Each year a 'conditional projection' for the growth of the monetary aggregate, M3, was announced in the federal budget by the Treasurer.

A case study in monetary targeting: Australia, 1976–85

The monetarist view had its initial policy impact in the mid to late 1970s. Australia was just one of many countries that experimented with policy based on the monetarist policy prescription of monetary targeting. Because of the historically very high rates of inflation in the early 1970s in Australia, and the worldwide intellectual influence of Monetarism, the implementation of monetary policy in Australia during 1976–85 consisted of a form of monetary targeting.

Each year a **conditional projection** for the growth of the monetary aggregate, M3, was announced in the federal budget by the Treasurer. These projections were in the nature of bands within which the RBA was expected to keep the growth of M3. The projections were conditional in the sense that, if the world or domestic economic conditions changed unexpectedly, the government allowed itself the room to vary the targets over the year. This era of monetary targeting continued from late 1976 until early 1985.

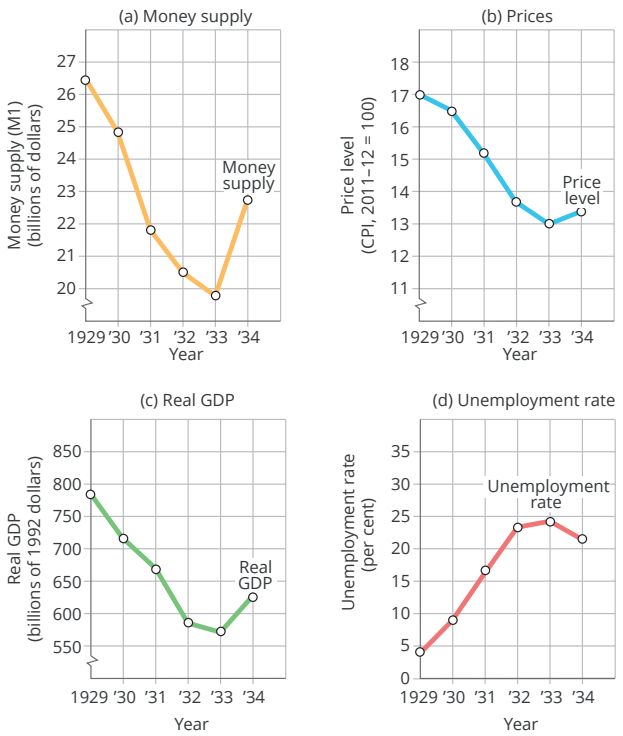
Global perspective

Applicable concept: the role of quantity of money in supply in the economy

Monetary policy during the Great Depression

Milton Friedman and Anna Schwartz, in their book *A Monetary History of the United States*,[1] argued that the Great Depression was in most part caused by the decline in the US money supply, as shown in part (a) of **Exhibit 16.5**. The accompanying parts (b), (c) and (d) present changes in the price level, real GDP and unemployment rate.

Exhibit 16.5 **The Great Depression economic data, 1929–34**



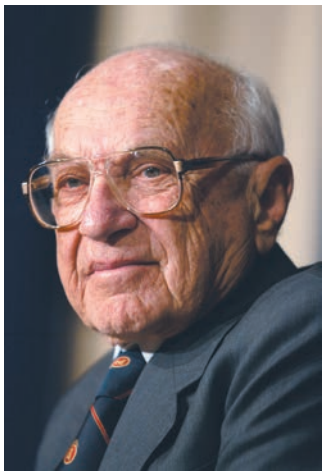
During the 1920s, the money supply expanded steadily and prices were generally stable. In response to the great stock market crash of 1929, bank failures, falling real GDP and rising unemployment, the US central bank, the Fed, elected to allow banks to fail, and since bank deposits (which were lost when the banks collapsed) were destroyed in the process, thereby allowed the US money supply to contract.

Through the Great Depression years from 1929 to 1933, M1 declined by an extraordinary 27 per cent. Assuming velocity is relatively constant, the sharp reduction in the quantity of money in circulation would be predicted to result in a sharp reduction in prices, output and employment. As part (b) of **Exhibit 16.5** shows, the price level indeed declined by 24 per cent between 1929 and 1933. In addition to deflation, part (c) and (d) of **Exhibits 16.5** show that real GDP was a quite extraordinary 30 per cent lower in 1933 than in 1929, and unemployment rose from 3.2 per cent in 1929 to 24.9 per cent in 1933!

Friedman and Schwartz argued that it was the ineptness of the Fed's monetary policy during the Great Depression that caused the trough in the business cycle to be so severe and sustained. As further evidence, they pointed to what happened in the period after 1933. The money supply grew and was followed closely by an increase in prices, real GDP and employment.

What should the Fed have done? Friedman and Schwartz argued that the Fed should not have allowed banks to fail and should not have waited until 1933 to use open-market operations to increase the money supply. Thus, their conclusion was that the Fed was to blame for the severity of the Great Depression by not pursuing an appropriately expansionary policy early enough in the crisis. This conclusion was a significant factor influencing subsequent views on the appropriate role for central banks to play in such a financial and economic crisis, and which in turn led to very different responses (compared with the Great Depression) by the world's central banks to the GFEC in 2008–09 and the COVID-19 pandemic.

Finally, although the emphasis here is on monetary policy, it should also be noted that government fiscal policies at the time of the Great Depression may also have worsened the situation; for example, US President Hoover was attempting to balance the budget, rather than using expansionary fiscal policy. In **Chapter 19**, when we discuss the roles played by monetary and fiscal policy in lessening the severity of the GFEC-induced recessions around the world in 2008–09, it will be beneficial to refer back to this 'Global perspective' box. Also, there will be discussions on the COVID-19 pandemic and the policy responses from around the world.



Getty Images/Corbis Historical/Brooks Kraft

Milton Friedman

Getty Images/Bloomberg

Anna Schwartz

What do you think?

1 Explain why monetarists believe the Fed should have expanded the money supply during the Great Depression. (*Hint:* Use the *AD–AS* model to represent the impact of the stock market crash on *AD* and what the impact would have been of an expansionary monetary policy by the Fed on *AD*.)

2 How might monetarists reconcile this position on maintaining the growth of the money supply during the years of the Great Depression with also generally tending to take a non-interventionist position to monetary policy?

[1] M. Friedman and A. J. Schwartz, *A Monetary History of the United States, 1867–1960*, Princeton University Press, Princeton, NJ, 1963.

The RBA had varying degrees of success in achieving these targets, mainly because of the financial institutional framework in place at the time severely hampered its ability to control money growth. Nevertheless, the policy presumably had some impact on inflation. From a peak of a very high 19 per cent in 1974–75, inflation had dropped to about 7 or 8 per cent by the middle of the 1980s. At the same time, unemployment, unfortunately, had also commensurately increased from about 5 per cent to about 8 per cent by the mid-1980s.

The financial institutional issues that hampered the RBA in meeting its announced monetary targets were rectified by 1985, giving the RBA greater ability than it had had during the previous eight years to achieve the agreed annual targets for M3. Nevertheless, it was precisely at this time that the government and the RBA decided to abandon monetary targeting as a means of conducting monetary policy.

The justification for abandoning the targets stemmed from the correct perception by the authorities that the demand for M3 – and for the other monetary aggregates – was shifting around unpredictably due to the substantial financial deregulation and rapid financial innovations that were occurring at the time.

In other words, it was precisely at this time that the velocity of money (V) had become very unstable and unpredictable both in the short run and in the longer term. Thus, the critical condition for a monetarist approach being a reasonable one – namely that the demand for money balances and, therefore, the velocity of money be stable – ceased to apply.

It is important to elaborate on this point a little more. As has been explained, it is the interaction of the demand and supply of money that determines interest rates. Controlling the growth of the supply of money (in this case M3) would result in interest rates becoming unnecessarily volatile simply because the supply of money is being controlled while the demand is fluctuating. Such interest rate volatility could prove detrimental to economic growth. In such circumstances, it would be better to directly target interest rates at whatever level is desired to maintain the appropriate level of aggregate demand rather than to target money growth. Money growth should simply increase or decrease as necessary to curb any market pressure for interest rates to change from their desired levels.

It is this idea that led to the approach the RBA now uses in its operation of modern monetary policy in Australia as was outlined previously.

Monetary policy measures during the COVID-19 crisis

The Reserve Bank of Australia has implemented comprehensive monetary policy measures to support the Australian economy to recover from the economic crisis due to the COVID-19 pandemic.[2] The RBA lowered the cash rate twice in March 2020 to 0.25 per cent and 0.1 per cent in November 2020. **Exhibit 16.6** shows the cash rate targets over the period from 1990 till end of 2020. During the COVID-19 pandemic, the RBA decided to set the cash rate target for the lowest rate in history. The aim of reducing the cash rate was to boost the cash flow of households and businesses and to support the trade-exposed industries through the lower exchange rates. The RBA has added extra liquidity to the financial systems through daily market operations. The RBA announced in the November 2020 meeting that they would purchase $100 billion worth of bonds issues by the Australian government, states and the territories. The RBA also implemented 'term funding facility' to lower the cost of credit to households and businesses and to provide incentives for lenders to support credit to small- and medium-sized businesses.

[2] Reserve Bank of Australia, 'Supporting the economy and financial system in response to COVID-19', https://www.rba.gov.au/covid-19/#:~:text=The%20Reserve%20Bank%20Board%20reduced,industries%20through%20the%20exchange%20rate

Exhibit 16.6 **Monetary policy decisions, cash rate target 1990–2020**

Effective date	Change % points	Cash rate target %	Effective date	Change % points	Cash rate target %
4-Nov-20	-0.15	0.1	2-Mar-05	0.25	5.5
20-Mar-20	-0.25	0.25	3-Dec-03	0.25	5.25
4-Mar-20	-0.25	0.5	5-Nov-03	0.25	5
2-Oct-19	-0.25	0.75	5-Jun-02	0.25	4.75
3-Jul-19	-0.25	1	8-May-02	0.25	4.5
5-Jun-19	-0.25	1.25	5-Dec-01	-0.25	4.25
3-Aug-16	-0.25	1.5	3-Oct-01	-0.25	4.5
4-May-16	-0.25	1.75	5-Sep-01	-0.25	4.75
6-May-15	-0.25	2	4-Apr-01	-0.5	5
4-Feb-15	-0.25	2.25	7-Mar-01	-0.25	5.5
7-Aug-13	-0.25	2.5	7-Feb-01	-0.5	5.75
8-May-13	-0.25	2.75	2-Aug-00	0.25	6.25
5-Dec-12	-0.25	3	3-May-00	0.25	6
3-Oct-12	-0.25	3.25	5-Apr-00	0.25	5.75
6-Jun-12	-0.25	3.5	2-Feb-00	0.5	5.5
2-May-12	-0.5	3.75	3-Nov-99	0.25	5
7-Dec-11	-0.25	4.25	2-Dec-98	-0.25	4.75
2-Nov-11	-0.25	4.5	30-Jul-97	-0.5	5
3-Nov-10	0.25	4.75	23-May-97	-0.5	5.5
5-May-10	0.25	4.5	11-Dec-96	-0.5	6
7-Apr-10	0.25	4.25	6-Nov-96	-0.5	6.5
3-Mar-10	0.25	4	31-Jul-96	-0.5	7
2-Dec-09	0.25	3.75	14-Dec-94	1	7.5
4-Nov-09	0.25	3.5	24-Oct-94	1	6.5
7-Oct-09	0.25	3.25	17-Aug-94	0.75	5.5
8-Apr-09	-0.25	3	30-Jul-93	-0.5	4.75
4-Feb-09	-1	3.25	23-Mar-93	-0.5	5.25
3-Dec-08	-1	4.25	8-Jul-92	-0.75	5.75
5-Nov-08	-0.75	5.25	6-May-92	-1	6.5
8-Oct-08	-1	6	8-Jan-92	-1	7.5
3-Sep-08	-0.25	7	6-Nov-91	-1	8.5
5-Mar-08	0.25	7.25	3-Sep-91	-1	9.5
6-Feb-08	0.25	7	16-May-91	-1	10.5
7-Nov-07	0.25	6.75	4-Apr-91	-0.5	11.5
8-Aug-07	0.25	6.5	18-Dec-90	-1	12
8-Nov-06	0.25	6.25	15-Oct-90	-1	13
2-Aug-06	0.25	6	2-Aug-90	-1	14
3-May-06	0.25	5.75			

Source: https://www.rba.gov.au/statistics/cash-rate/

In summary

- Modern monetarists argue that the growth rate of the money supply is an important influence on the growth of nominal GDP.
- They also maintain that the channel of that influence is broader than the interest rate channel.
- Their policy prescription to avoid high inflation and/or unemployment is to maintain the money supply growth rate at a prescribed, predetermined level.

4 The appropriate role for monetary policy: the rules versus discretion debate

All economists agree that monetary policy has powerful effects on the macro economy. They even mostly agree that the short-run effects – say, from three to nine months – of a monetary change are likely to manifest mostly in changes in demand, production and employment, with changes in inflation taking somewhat longer, usually from about nine to 18 months. However, there are nonetheless some divergent views as to the proper role of monetary policy.

Many economists favour the central bank retaining considerable discretion to adjust monetary policy as needed and as frequently as required to achieve its stated goals. Others, on the other hand, are more likely to prescribe a policy whereby the central bank is required to adhere fairly closely to pre-announced and widely known monetary policy rules and not to have the discretion to adjust its policies as economic circumstances appear to change. The most common type of rules-based approach is one involving pre-set targets for money supply growth.

While the 'stick to your rules, sit on your hands and do nothing' approach in the face of continually changing macroeconomic circumstances may at first sound counterintuitive, those who advocate such an approach to monetary policy do so with good reasons, which those with a more interventionist bent no doubt also appreciate. In what follows, we will try to provide the flavour of some of the core issues at stake in this debate.

One core tenet of Monetarism is a non-interventionist approach to monetary policy. Monetarism gained intellectual credibility among the academic economic community in the late 1950s and 1960s, led by Nobel Laureate Professor Milton Friedman, at the University of Chicago. It also consequently had a significant impact on the conduct of monetary policy around the world in the 1970s and early 1980s.

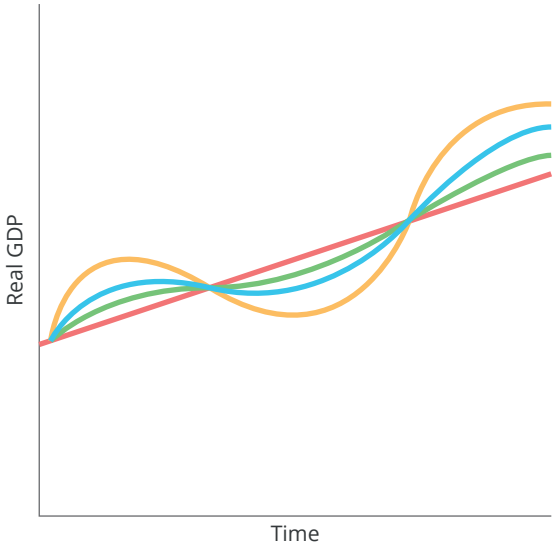
Consider **Exhibit 16.7**. Suppose the yellow line represents the course the economy would follow over time if left to its own accord without any policy intervention. In other words, the yellow line represents the economy's business cycle. **Countercyclical macroeconomic policy** has the aim of smoothing out the fluctuations in the business cycle.

Countercyclical macroeconomic policy
Policy that aims to smooth out the fluctuations in the business cycle.

In terms of the exhibit, successful countercyclical monetary policy would have the central bank adjusting monetary policy in a timely way so that the impact of monetary policy adjustments would be felt in the macro economy at just the right time to dampen down the swings in the business cycle. The path followed by the economy with such successful policy intervention might follow something like the green line in the exhibit.

Friedman argued very persuasively that, due to information lags (recognition lags), policy determination lags (implementation lags) and policy effectiveness lags (impact lags), such discretionary countercyclical monetary policy can have the opposite effects on the macro economy to those intended by policymakers. These lags may well result in the monetary policy intervention impacting on the economy at precisely the wrong time, exacerbating the business cycle rather than dampening it. Such an undesirable result is represented in the exhibit by the red line. In what follows, we spend a little time discussing these lags.

Exhibit 16.7 **A stylised depiction of the business cycle and countercyclical macroeconomic fine-tuning policy**



The yellow line represents the path of real GDP without policy intervention. The green line represents the path of real GDP under successful countercyclical macroeconomic fine-tuning policy. The red line represents the path of real GDP under the effects of inappropriate policy that exacerbates the business cycle rather than dampening it.

Information about the current phase of economic activity in the real world is imperfect and partial, and sometimes becomes available only after a lag of some months. This information is crucial to identify the phase of the business cycle or economic issues. This time delay constitutes the **information lag**.

It then takes more time for the authorities to decide on the appropriate policy response. This can involve protracted deliberations, particularly in light of the partial and sometimes conflicting messages being distilled from the various monthly and quarterly macroeconomic indicators as they become available. The time taken to decide on an apparently appropriate policy response may amount to several months and is known as the **policy determination lag**.

Once a policy is initiated there will be a further lag before it begins to take effect on economic activity. This lag is referred to as the **policy effectiveness lag** and is likely to be quite long for monetary policy, mainly if one accepts the version of the transmission mechanism, which emphasises the cost of borrowing. In this case, a change in monetary policy can take a long time to have a discernible impact on the macro economy. Most economists agree that, from the moment any given monetary policy change is implemented, this time is most likely to be at least three months and could be up to over a year.

The reality of these lags and the fact they are also likely to be quite highly variable over time makes active countercyclical monetary policy fine-tuning extremely difficult in practice. To illustrate the difficulty, a motoring metaphor may be helpful.

The situation facing the monetary policymaker would be akin to trying to drive a car along a winding road (the bends represent the effects of unexpected shocks to the economy, sending it away from its long-term growth path) with vision out of the front windscreen almost entirely obscured and only an imperfect (fuzzy) rear-vision mirror for guidance. So far, this represents the very uncertain future, and the information and policy determination lags described above. Furthermore, to keep the car on the road, there is an accelerator, brake and steering wheel that react only after a variable delay and with differential effect from one application to another. This represents the uncertain response of the economy to any given change in monetary policy – that is, the policy effectiveness lag. Under such adverse circumstances, it would not be at all surprising for the car to often end up in a ditch on the side of the road!

Information lag
The delay (months) before information about the current phase of economic activity in the real world becomes available.

Policy determination lag
The time taken to decide on an apparently appropriate policy response; may be several months.

Policy effectiveness lag
The delay between the time when a policy is initiated and the time when it begins to take effect on economic activity.

Monetarists have an answer to the above problems to ensure the economy grows at about the right rate over the longer term. Instead of running the risks of policy errors, the answer is to forget about policy activism and instead follow steady, predictable monetary policy. In particular, given the significant role they ascribe to money supply growth, monetarists argue that the central bank should announce a target for money growth (e.g., in the Monetary Base, M1 or M3) for, say, the year ahead and the reasons for that target in the context of its overall goals (as agreed with the government).

The bank should then use its various tools to try to meet the announced money supply growth target, with the target itself subjected to annual review. The central bank would continue to seek to meet the pre-announced money supply target, irrespective of what happens in the economy over the course of the year in question. Monetarists argue that such an approach would deliver more certainty to economic decision-makers and, in the longer term, deliver better macroeconomic outcomes – for inflation, growth and employment – than following a discretionary approach.

Monetarists would, for example, advise that the central bank should expand the money supply at a rate equal to what is considered to be the potential growth rate in real GDP (say 3%), plus some desired inflation rate (say 2 to 3%); that is, it should increase the money supply somewhere between 5 and 6 per cent per year. The central bank should pick a money supply growth rate and stick to it even if unexpected changes in velocity cause short periods of inflation or low-output growth.

Monetarists argue that their 'straitjacket' approach would reduce the average intensity and duration of unemployment and inflation by eliminating the central bank's discretion to change the growth rate of the money supply. Needless to say, this view is not endorsed by all economists. A proponent of the discretionary approach might sardonically note that the rules-based money supply growth approach is best summed up as: 'In the event of an accident, don't do something, just stand there!'

You're the economist A horse of which colour?

A famous Nobel Prize-winning economist once proposed replacing the US Fed – the central bank of the US – with an intelligent horse. Each New Year's Day, the horse would stand in front of Fed headquarters to answer monetary policy questions. Reporters would ask, 'What is going to happen to the money supply this year?' The horse would tap its hoof four times and the next day headlines would read 'Fed to Once Again Increase the Money Supply 4 Per Cent'.

Is this famous economist in favour of a discretionary approach to monetary policy or a rules-based approach (i.e., a monetarist approach)?

Foreign exchange market operations

During the era when Australia maintained a fixed exchange rate system, any discrepancy between the demand and supply of Australian dollar (AUD) for sale in the foreign exchange (Forex) markets had to be met by the RBA in order to maintain the foreign exchange value of the AUD fixed. For example, if there was an excess supply of AUD for sale, the RBA would have to purchase the excess AUD, selling some of its holdings of foreign exchange in return. This action would reduce the liquidity of the financial system since the AUD purchases by the RBA would reduce the Monetary Base. This tighter liquidity would have the further undesirable effect of tending to push up domestic interest rates and hinder economic growth. The RBA also engage in the foreign exchange market when managing its foreign currency reserves.

Floating exchange rate In a floating exchange rate, market demand and supply determine the foreign price of the currency.

Such action by the RBA is no longer necessary since Australia adopted a floating exchange rate system in 1983. Instead, with a **floating exchange rate**, if sellers of AUD cannot find buyers, they simply have to reduce the price of the AUD until buyers are forthcoming. Floating exchange market in Australia is more favourable to face external shocks like the COVID-19 pandemic and to insulate the economy from such shocks.

Of course, the RBA still regularly operates in the Forex markets if it considers the AUD is coming under undue short-run speculative pressure or if market conditions are excessively volatile. The latter activity is termed 'smoothing' while the former is referred to as 'testing'. **Smoothing** operations by the RBA are intended to be only very short term, sometimes over the course of a day and rarely for more than a few days. **Testing** by the RBA, also referred to as 'leaning against the wind', is intended to test the market's resolve in moving away from what the RBA might consider the current equilibrium value for the AUD. However, given sustained market pressure to shift the exchange rate, the RBA will (by necessity) follow the market.

The AUD is now one of the most highly traded currencies in international foreign exchange (Forex) markets and fluctuates quite widely. For example, in the 10-year period from 2001 to 2011 the AUD varied between US$0.49 and US$1.10. In other words, at its highest value, the AUD was some two and a quarter times above its lowest value during this period! This issue is discussed further in **Chapters 18 and 19**.

Smoothing and testing
The RBA regularly operates in the market if it considers the AUD is coming under undue short-run speculative pressure ('testing'), or if Forex market conditions are excessively volatile 'smoothing').

In summary

- **Monetarists** advocate that the central bank should follow a publicly known rules-based approach to setting the growth in the money supply each year.
- Most central banks around the world these days do not do this. Instead, they use some appropriate short-term interest rate as their monetary policy instrument, and vary it to try to bring about a desirable path for inflation and economic growth.
- This is a discretionary approach to monetary policy rather than a rules-based approach.

Key concepts

Monetary policy
Goal of monetary policy
Monetary policy transmission mechanism
Statement on the Conduct of Monetary Policy
Overnight cash rate (cash rate)
Open-market operations
Monetarism
Quantitative easing
Equation of exchange
Velocity of money
Quantity theory of money
Conditional projection
Countercyclical macroeconomic policy
Information lag
Policy determination lag
Policy effectiveness lag
Floating exchange rate
Smoothing and testing

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand what monetary policy is, what its goal is, and to look at a widely held view of how policymakers believe that changes in monetary policy affect the economy

- **Monetary policy** can be defined as the set of actions taken by the central bank to influence interest rates, aggregate demand, economic activity and prices in the economy.

- Many central banks, including Australia's RBA, operate monetary policy with a proximate goal of maintaining some appropriately defined inflation target.
- In seeking to achieve its goal, the central bank changes the supply of liquidity available to the financial system to bring about a change in domestic interest rates. The change in interest rates, in turn, affects investment and consumer durable demand and possibly also net exports if the value of the currency changes – and this changes aggregate demand.
- These effects can be represented in the *AD–AS* framework by either an outward shift (an **expansionary monetary policy**) or an inward shift (a contractionary monetary policy) in the *AD* curve.
- Changes in aggregate demand then lead to changes in real GDP, employment and prices.

2. Understand how central banks implement modern monetary policy

- The way Australia's central bank, the RBA, implements monetary policy is a good exemplar of how modern monetary policy is implemented by central banks. In pursuing its inflation target goal, the RBA – like other central banks – seeks to maintain some key short-term interest rate at the desired policy level. In Australia's case, it is the **overnight cash rate**.
- Changes in the cash rate also influence changes in other interest rates throughout the economy, and the general change in interest rates impacts on aggregate demand.
- In bringing about changes in the cash rate, the RBA carries out **open-market operations** in financial markets where it transacts with financial institutions like banks in buying or selling Commonwealth government securities (CGS). These actions impact the short-term liquidity available in the financial markets (i.e., the Monetary Base) and, by so doing, the RBA maintains the cash rate at the desired level.

3. Understand that there are two alternative views as to how monetary policy adjustments actually affect the economy

- The channel through which monetary policy impacts on the economy is through changing interest rates. This, in turn, affects aggregate demand, and then consequently affects real GDP, employment and prices. In that characterisation, changes in interest rates are critical.
- The **Monetarism** school of thought sees monetary policy operating more widely than this, and argues that changes in the money supply much more directly determine changes in prices, real GDP and employment.
- Modern monetarists, drawing on the simple **quantity theory of money** (where changes in the money supply lead directly to proportional changes in the price level) argue that the growth rate of the money supply is an important influence on the growth of nominal GDP.
- Modern monetarists also maintain that the channel of the influence of the growth rate of the money supply is broader than the interest rate impact of changes in money supply/liquidity and that to avoid high inflation and/or unemployment, their policy prescription is to maintain the money supply growth rate at a prescribed, predetermined level.

4. Understand the 'rules' versus 'discretion' debate in relation to the most appropriate role for monetary policy

- **Monetarists** argue that the presence of long and variable lags (the **information lag**, the **policy determination lag** and the **policy implementation lag**) are such that activist **countercyclical monetary policy** is more likely in practice to exacerbate the business cycle rather than dampen it.
- Monetarists advocate that the central bank should follow a publicly known rules-based approach to setting the growth in the money supply each year. However, most central banks around the world these days use some appropriate short-term interest rate as their monetary policy instrument, and vary it as they believe is necessary to bring about a desirable path for inflation and economic growth. In other words, they follow a discretionary approach to monetary policy rather than a rules-based approach, and their respective governments support this. It is nonetheless essential to understand the arguments involved in the debate.

Study questions and problems

1 Using diagrams analogous to those in **Exhibit 16.3**, trace out the likely effects of a decision by the RBA to sell Commonwealth government securities to participants in the financial markets. (*Hint:* What happens to the quantity of Monetary Base in supply in the financial system when the RBA sells CGS to the private sector?)

2 Suppose that, during a year, a country's nominal GDP was measured at $1400 billion and the volume of money (say, defined by M1) circulating in the economy during the year was, on average, equal to $350 billion. Using the equation of exchange determine the velocity of money during the year.

3 Now suppose that in question (2) the velocity of money remained unchanged during the following year and the volume of money – as defined by M1 – increased by 5.5 per cent. Using the quantity theory of money, predict the value of annual nominal GDP during the next year.

4 Now suppose that in the situation described in question (3) the economy in question was operating at full employment. What would you predict would have happened to the price level from the first year to the second?

5 With the aid of money supply and money demand, explain how expansionary monetary policy could be implemented by the central bank and explain the effect on the interest rate.

6 Explain the monetarist view of the monetary policy transmission mechanism and how **Exhibit 16.2** might be amended to represent this mechanism.

7 What would be the impact of a monetary policy easing if aggregate demand was inelastic with respect to interest rates (i.e., it did not respond to changes in interest rates)? Represent this situation by appropriately adjusting **Exhibit 16.3**.

8 Explain quantitative easing and how it works as a monetary policy instrument to influence aggregate demand and economic activity. Contrast this to the more normal monetary policy implementation approach followed by central banks.

9 What, if anything, happens to the economy's Monetary Base if the RBA sells $50 million in foreign currency from its holdings of foreign currency reserves to the ANZ Bank? Explain.

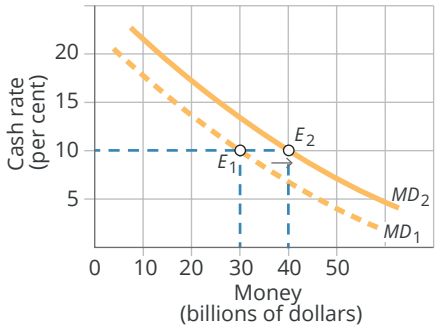
10 After the onset of the GFEC, if the world's central banks had then allowed large-scale bank failure to occur, this would have resulted in the destruction of huge volumes of bank deposits and thereby very significant reductions in the money supplies in their countries. Represent this money supply reduction impact on the economy using the *AD–AS* framework and explain the outcome predicted by the framework.

Answers to 'You're the economist'

The money supply curve when the RBA targets an interest rate

When the RBA targets a particular value for the cash rate, it must adjust the supply of liquidity to the money markets in such a way as to equate the supply to whatever is the particular demand for liquidity at the desired target value for the cash rate. Consider **Exhibit 16.8**. Suppose the RBA wants to keep the cash rate at 10 per cent and demand is represented by MD_1. The RBA would therefore need to ensure it supplied $30 billion of funds to the money markets to bring about a cash rate of 10 per cent. Now suppose that for some reason money market demand increased, represented in the exhibit as a shift from MD_1 to MD_2. If there were no increase in the supply, there would be pressure for the cash rate to increase. To offset this pressure, the RBA would therefore need to respond to this increase in demand by increasing the available supply of liquidity to the money markets from $30 to $40 billion. Effectively, then, the RBA has to supply whatever funds are demanded by the markets at the desired cash rate of 10 per cent.

Exhibit 16.8 **The money supply curve when the cash rate is targeted**



In this situation short-run equilibrium in the money market is represented by points E_1 and E_2 and both are points on an effectively horizontal money supply curve. If you said that when the RBA targets the cash rate the money supply curve must be horizontal at the targeted rate or perfectly elastic supply, YOU ARE THINKING LIKE AN ECONOMIST.

A horse of which colour?

The famous economist was Milton Friedman, who favoured a monetary rule for the Fed. The story of the horse is a sarcastic way of rejecting activist policies that Friedman believed only destabilised the economy. Friedman has even argued that the Board of Governors of the Federal Reserve System should announce the growth rate for the money supply each year and should resign if the target is missed. If you said this famous economist is a monetarist, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 With an unchanging money demand schedule, an increase in the interest rate would, other things equal, be associated with:
- **a** an unchanging money supply and an upward movement along the demand curve for money.
- **b** an increase in money supply and a downward movement along the demand curve for money.
- **c** a reduction in money supply and an upward movement along the money demand schedule.
- **d** an unchanging money supply and a downward movement along the demand for money curve.

2 Assume the schedule representing money market demand for liquidity is unchanged and the RBA increases the quantity of liquidity in supply to financial markets. The result is that:
- **a** the price of bonds rises.
- **b** the price of bonds remains unchanged.
- **c** the price of bonds falls.
- **d** none of the above occurs.

3 Using the aggregate supply and demand model, assume the economy is in equilibrium on the vertical portion of the aggregate supply curve. Other things being equal, an increase in the quantity of money in supply in the financial system will:
- **a** lower interest rates, and increase aggregate demand, real GDP and the average price level.
- **b** raise interest rates, and lower aggregate demand, real GDP and the average price level.
- **c** lower interest rates, and raise aggregate demand and the average price level.
- **d** raise interest rates, and lower aggregate demand and the average price level.

4 The monetary policy transmission mechanism:
- **a** uses government spending to influence output and prices.
- **b** is based on the belief that the *AS* curve is always horizontal.
- **c** places a key role on interest rates in the transmission mechanism.
- **d** emphasises that policy activism is to be avoided due to long and variable lags in the transmission mechanism.

5 The goal of monetary policy in Australia is:
- **a** to keep the overnight cash rate at the level desired by the RBA.
- **b** to minimise the deviation of economic growth from its long-term sustainable level.
- **c** to maintain unemployment at as low a rate as possible.
- **d** to keep inflation between 2 and 3 per cent per annum on average over the medium term.

6 Which of the expressions below correctly represents the equation of exchange?
- **a** $M = PV/Q$.
- **b** $MP = QV$.
- **c** $M = PQ/V$.
- **d** $MQ = PV$.

7 The velocity of money, *V*, is:
- **a** the speed with which people spend their money on average after receiving it in salary.
- **b** the volume of money stock divided by nominal GDP.
- **c** the average number of times per period a dollar of the money supply is used to purchase final goods and services.
- **d** the speed with which the Monetary Base moves around the financial markets.

8 From 1976–85, Australia employed a monetary policy that had a monetarist flavour. Which is not true about the monetary policy during 1976–85 in Australia?
- **a** Implementation of monetary policy in Australia during 1976–85 in Australia consisted of a form of monetary targeting.
- **b** During 1976–85 in Australia, the growth of the monetary aggregates was announced in the federal budget by the treasurer.
- **c** Implementation of monetary policy in Australia during 1976–85 in Australia consisted of a form of interest rate targeting.
- **d** During 1976–85 in Australia, RBA was expected to keep the growth of the monetary aggregates at the projected levels.

9 The information lag is:
- **a** the time it takes to determine the appropriate policy to implement.
- **b** the time between when a policy response is determined by the authorities and when this is communicated to the public.
- **c** the time it takes to get the required information upon which to base policy decisions.
- **d** now no longer an issue due to the widespread use of the internet in communications.

10 Quantitative easing occurs when:
- **a** central banks ease back on the quantity of government and corporate securities they purchase from the private sector.
- **b** central banks expand their balance sheets – and thereby the Monetary Base – by buying long-dated government and corporate securities.
- **c** banks ease back on the quantity of funds they are prepared to lend per period.
- **d** the quantity of goods and services produced in an economy eases from one period to another.

11 In a fixed exchange rate regime:

a the central bank never has to involve itself in foreign exchange transactions since the value of the currency is fixed by the government.

b a country's monetary base is not affected at all as the country's central bank intervenes into foreign exchange markets to hold the currency at some desired level.

c since the value of a country's currency is fixed by the government, the forces of supply and demand for the country's currency do not operate at all.

d the country's central bank fixes the Monetary Base at the required level in order to keep the exchange rate fixed.

Macroeconomic policy II: fiscal policy

17

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Develop an understanding of discretionary demand-side fiscal policy.
2 Understand the concept and rationale of supply-side fiscal policy, and understand how it differs from demand-side fiscal policy.
3 Appreciate the macroeconomic significance of the federal budget.
4 Develop an awareness of the implications of the federal budget outcome for government debt levels.

AAP Image/LUKAS COCH

In Australia in the early 1990s, the federal government increased its expenditures relative to revenue in a bid to expand aggregate demand. Its objective was to boost national output and employment in order to assist in ending the recession, which began in 1990–91. In addition, throughout 2009 the federal government took quite dramatic action in using its own spending to counter the anticipated debilitating economic impacts of the global financial and economic crisis (GFEC) in order to try to avert a recession occurring in Australia. More recently, during the COVID-19 pandemic, the Australian government changed its tax structure and spending patterns to support the weaker economy.

During the Australian federal election of 2019, both sides of politics made promises about assisting Australia's near-term and long-term economic growth prospects by appropriate adjustments to government spending and various taxes. The debate included the most appropriate levels of government expenditure in such areas as education, health, welfare and the nation's physical infrastructure (e.g., roads, rail, ports, bridges and telecommunications networks). Debate on taxation measures centred the coalition government's budget proposal. It was aimed at reducing income taxes to low and middle-income earners to decrease the tax rates in the middle-income tax brackets in the coming years. The idea was to make the income tax rates highly progressive where top income earners will pay slightly more taxes than previous years. Also, the focus was on company tax and the tax rates are expected to decrease from 27.5 per cent to 25 per cent, first for small- and medium-sized companies, and then later for larger companies progressively over a 10-year period.

Fiscal policy The use of government spending and taxes and their influence on the nation's economic growth, employment and price level.

A major issue, then, that touches everyone's life is **fiscal policy**, which is the use of government spending and taxes and their influence on the nation's economic growth, employment and price level. Federal government spending policies potentially affect all areas of the economy and the community more generally. Government spending initiatives directly affect specific sectors of the economy but also have indirect flow-on effects on other sectors. Changes to social welfare provisions directly affect welfare recipients, but also again have flow-on effects to others in the community via changes to the spending levels of such recipients. Changes in tax policies affect the amount of income retained by individuals and companies, and therefore directly influence spending, saving and investment decisions.

Using fiscal policy to influence the performance of the economy has been an important idea since the very influential writings of John Maynard Keynes in the aftermath of the Great Depression of the 1930s. This chapter looks at fiscal policy from the perspectives of two economic viewpoints. First, we will discuss the demand-side fiscal policies that attempt to 'fine-tune' aggregate demand in the economy in order to achieve full employment economic growth. Second, we will discuss supply-side fiscal policy, which gained prominence during the early 1980s. Supply-siders view aggregate supply as far more important than aggregate demand. Their fiscal policy prescription is to adjust policies to endeavour to increase aggregate supply in the economy in order to achieve full employment economic growth.

In addition, discussion will focus on how and why the government might run a budget deficit, the macroeconomic importance of a deficit, how it can be financed, the impact it may have on a country's national debt, and whether and when such increases in national debt can prove a problem.

1 Discretionary demand-side fiscal policy

Discretionary fiscal policy The deliberate use of changes in government spending and/or taxes to alter aggregate demand and stabilise the economy's business cycle.

We begin by discussing the use of discretionary demand-side fiscal policy – or more simply, **discretionary fiscal policy** – as Keynes advocated in the 1930s in his *General Theory*, to influence the economy's performance. Discretionary fiscal policy is defined as the deliberate use of changes in government spending and/or taxes to alter aggregate demand in an attempt to stabilise the economy in some desirable way. As will be seen later, the national government's approach to fiscal policy is enunciated each year in its annual budget.

Exhibit 17.1 lists two basic types of discretionary fiscal policies and some of the corresponding ways in which the government can pursue each of these options. The first column of the table shows that the government can choose to increase aggregate demand by following an *expansionary* fiscal policy. The second column lists *contractionary* fiscal policy options the government can use to restrain aggregate demand.

Exhibit 17.1 **Discretionary fiscal policies to influence aggregate demand**

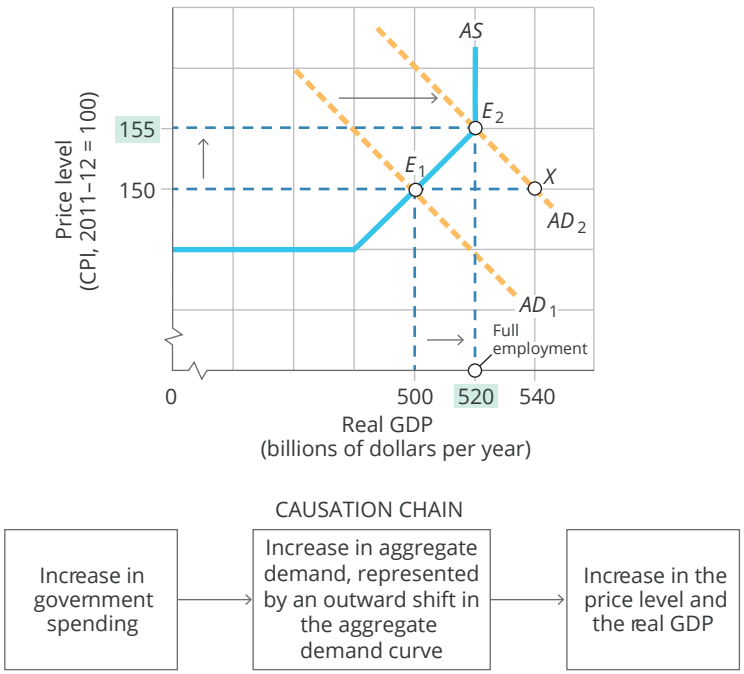
Expansionary fiscal policy	Contractionary fiscal policy
Increase government spending	Decrease government spending
Decrease taxes	Increase taxes
Increase government spending and also decrease taxes	Decrease government spending and also increase taxes
Increase government spending and increase taxes equally*	Decrease government spending and decrease taxes equally*

* This relates to the 'balanced budget multiplier' discussion in the text.

Increasing government spending to combat a recession

Suppose the economy represented in **Exhibit 17.2** has entered a period of slower than normal economic growth represented by equilibrium point E_1, where aggregate demand curve AD_1 intersects the aggregate supply curve AS in the near full-employment range. (Note that for simplicity, the aggregate demand and aggregate supply curves are drawn here as straight lines.)

Exhibit 17.2 **Using government spending to combat a recession**



The economy represented in this exhibit has recently entered a relatively low-growth phase represented by equilibrium point E_1 on the intermediate range of the aggregate supply curve, *AS*. The price level is 150, with an output level of $500 billion real GDP. In using discretionary demand-side fiscal policy, to reach the full-employment output of $520 billion in real GDP, aggregate demand needs to be stimulated. This is represented in the figure by an outward shift in the *AD* curve to the right by $40 billion real GDP, measured by the horizontal distance between point E_1 on curve AD_1 and point *X* on curve AD_2. The necessary increase in aggregate demand from AD_1 to AD_2 can be accomplished by increased government spending. Given a spending multiplier of 4, a $10 billion increase in government spending brings about the required $40 billion increase in aggregate demand, represented in the exhibit by the rightward shift in the aggregate demand curve, and equilibrium in the economy changes from E_1 to E_2. Note that the equilibrium real GDP changes by $20 billion and not the full amount by which the aggregate demand curve shifts horizontally, as some of the additional aggregate demand results in a rise in the price level.

The price level measured by the consumer price index (CPI) is 150 and the real GDP gap happens to be $20 billion below the full-employment output of $520 billion real GDP. In this situation policymakers may elect to follow a Keynesian-type policy prescription and act to try to increase aggregate demand, represented by a shift of the aggregate demand curve rightward from AD_1 to AD_2, thereby bringing about an end to the slow growth period but in the process also producing a higher price level.

How can the government do this? Remember that, in principle, any increase in private-sector demand, namely consumption (C), investment (I) or net exports ($X - M$), can spur aggregate demand. But these spending boosts are not directly under the government's control, whereas government spending (G) is. After all, there is always a long wish list of spending proposals for national, state and local roads, health care, education, other

transport infrastructure, environmental programs and so on. Therefore, rather than waiting for the private sector to increase its demand for, and production of, consumption and investment goods and services, suppose the government chooses to increase government spending to boost employment in the short term.

But just how much new government spending is required? Suppose the government increases its spending on transport infrastructure by $10 billion. How much will aggregate demand increase as a result? The answer is by considerably more than the initial $10 billion!

This is through the action of the spending multiplier (if you need to, refer back to **Chapter 14** for details). In brief, the initial spending by the government is amplified as some of it is in turn spent by those receiving the money as income, some of that money is spent again by those who received it as income from the second induced spending round, and so on.

The original spending causes a ripple effect throughout the entire economy. The extent of the ripples depends on the marginal propensity to consume (MPC) of the recipients of the income. In **Exhibit 17.2** it is assumed that the MPC is 0.75, which, as will be shown below, implies a spending multiplier of 4. Thus, the *AD* curve is depicted as shifting out horizontally by $40 billion (and not by the initial increase in spending of $10 billion by the government).

Why does an MPC of 0.75 imply a multiplier of 4? Consider $1 of initial spending. This will give rise to an induced second round of spending of $0.75 (with the other $0.25 saved). This second round of spending will give rise to a third round of spending of $0.56 (0.75 × 0.75), a fourth round of spending of $0.42 (0.56 × 0.75), a fifth round of spending of $0.32 (0.42 × 0.75) and so on. Each round, the induced spending gets smaller.

Total spending will eventually be 1 + 0.75 + 0.56 + 0.42 + 0.32 + … After a very large number of spending rounds, the induced spending in the next round will be almost zero and subsequent spending could be ignored. After such a large number of spending rounds, the resulting sum will be approximately 4, meaning the MPC of 0.75 results in a spending multiplier of 4.[11] However, the spending multiplier that was discussed above is very simple and real-world spending multipliers are more complicated with the inclusion of taxes, interest rate, inflation and imports.

Rather than write down all the spending rounds until further induced spending gets very small, a much quicker way of calculating a spending multiplier from a given MPC is to recognise that the above sum converges in the limit to 1/(1 − MPC). Thus, for an MPC of 0.75 the multiplier would be 1/(1 − 0.75), which equals 4. With a larger MPC – meaning a greater proportion of any income received is spent – each round of induced spending will be greater and so the multiplier will be larger. For example, for an MPC of 0.8, the multiplier would be 5, while for an MPC of 0.50 the multiplier would be 2.

Now, as described in **Chapter 14**, bottlenecks, labour shortages and diminishing returns to labour productivity occur throughout the upward-sloping range of the *AS* curve. This means that costs of production rise as production increases in response to the greater aggregate demand. In times of buoyant demand, these cost increases are likely to be passed on as higher prices.

Returning to **Exhibit 17.2**, you can see that $10 billion worth of new government spending increases aggregate demand by $40 billion, represented by an outward shift of the aggregate demand curve from AD_1 to AD_2. As a result, the equilibrium in the economy changes from point E_1 to point E_2 and full employment is achieved. In the process, the economy experiences *demand-pull inflation* (refer back to **Chapter 14**) and the CPI rises from 150 to 155.

Note that although the aggregate demand curve shifted to the right by $40 billion, total output rose by only $20 billion between points E_1 and E_2. The other effect of the extra aggregate demand was to raise prices in the economy.

---

[1] The spending multiplier discussion in this chapter abstracts from the issue of income taxes and so the calculated multipliers make no allowance for income tax rates.

Calculating the size of the spending multiplier

You're the economist

Assume a simple economy where MPC for a community is 0.6 and suppose there is an initial stimulus to spending of $15 billion. Calculate the total stimulus to spending after three rounds of induced spending. Now calculate the total spending stimulus after six, then eight, and then 10 rounds of induced spending.

What does the total spending stimulus appear to be approaching? What do you conclude about the size of the spending multiplier in this case?

Cutting taxes to combat a recession

Another type of expansionary fiscal policy intended to increase aggregate demand and restore full employment calls for the government to cut taxes. Let's return to point E_1 in **Exhibit 17.2**.

As before, the task is to increase aggregate demand. But this time, instead of a $10 billion increase in government spending, assume parliament votes for a $10 billion personal income tax cut. How does this cut in personal income taxes affect aggregate demand? First, *personal disposable income* (take-home pay) increases by $10 billion – the amount of the tax reduction. Second, once again, assuming the MPC is 0.75, the increase in personal disposable income induces new consumption spending of $7.5 billion (0.75 × $10 billion).

Notice that the initial stimulus of $7.5 billion to aggregate demand here is less than the $10 billion cut in income taxes. The reason is that some of the tax cut will be saved rather than spent. After this initial round of new consumption spending of $7.5 billion, the spending multiplier process as described above will again produce further rounds of spending.

However, because the initial stimulus is less than before ($7.5 billion versus $10 billion), an important thing to note here is that the tax cut produces a smaller overall increase in aggregate demand than does the impact of the same-sized increase in government spending. In this case, with an MPC of 0.75, the total stimulus from the $10 billion tax cut will be $30 billion (4 × $7.5 billion). This would be represented by an outward horizontal shift of the *AD* schedule by $30 billion, which, from the exhibit, would be insufficient to fully move the economy to full employment output (an increase in aggregate demand of $40 billion being required). In the following 'You're the economist' you are asked to calculate the quantum of the necessary income tax cut required to move the economy to full employment.

In passing, it is worth pointing out that, as with the above analysis of a tax cut, the same is true in principle of an increase in welfare payments since the recipients can be expected to save some portion of their payment. Again, say their MPC is 0.75 and the government increases welfare payments by $10 billion. Their spending would thus increase by $7.5 billion only and the final stimulus to overall aggregate demand would therefore again be $30 billion, rather than the $40 billion resulting from increased government spending of $10 billion.

Some warnings about the above analysis: first, in reality, different income groups will have different MPCs and so the actual spending impact of cuts in income taxes will depend very much on where they are targeted. It is also highly likely that, as far as welfare payments are concerned, their recipients are almost certain to have a higher MPC than others in the community.

Also, it is likely that the MPC for any group over time may change (e.g., over the business cycle as business and consumer confidence waxes and wanes), and so the spending impact of any given change in tax policy may be quite difficult to predict. If people – for one reason or another – decide to save a lot of any given income tax cut rather than spending it, then the hoped-for impact on aggregate demand may be very much reduced.

Using fiscal policy to combat inflation

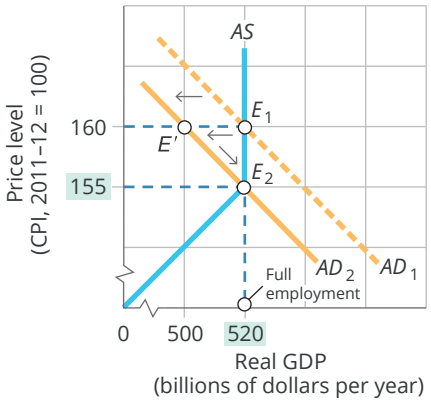
So far, discretionary expansionary fiscal policy has been presented as the cure for an economic downturn. Contractionary fiscal policy, on the other hand, can assist in the fight against inflation, particularly demand-pull inflation.

You're the economist — Calculating the required tax cut

Suppose you are an economic adviser in the federal Treasury and, similar to the situation depicted in **Exhibit 17.2**, it is believed that the economy needs an increase in aggregate demand of $40 billion to reach full-employment equilibrium. If the marginal propensity to consume (MPC) in the economy was 0.75, by how much would you advise the government to cut taxes in order to restore the economy to full employment?

Exhibit 17.3 shows an economy operating at point E_1 on the vertical range of the aggregate supply curve, *AS*. This economy is producing the full-employment output of $520 billion real GDP and the price level is 160. In this situation, any increase in aggregate demand only causes inflation, while real GDP remains unchanged.

Exhibit 17.3 Using fiscal policy to combat inflation



CAUSATION CHAIN

Decrease in government spending or increase in taxes → Decrease in aggregate demand, represented by an inward shift of the aggregate demand curve → Decrease in the price level

The economy in this exhibit is in equilibrium at point E_1 on the vertical range of the aggregate supply curve, *AS*. The price level is 160 and the economy is operating at the full-employment output of $520 billion real GDP. To reduce the price level to 155, aggregate demand must be reduced by $20 billion. This is represented by an inward shift in the aggregate demand curve to the left by $20 billion, measured by the horizontal distance between point E_1 on curve AD_1 and point E' on curve AD_2. One way this can be done is by decreasing government spending. With MPC equal to 0.75 and therefore a spending multiplier of 4, a $5 billion decrease in government spending results in the needed $20 billion leftward shift in the aggregate demand curve. As a result, equilibrium in the economy reaches point E_2 and the price level falls from 160 to 155, while real output remains unchanged at full capacity.

An identical decrease in aggregate demand can be obtained by a hike in taxes. A $6.67 billion tax increase would reduce disposable income by this amount and, with an MPC of 0.75, would result in an initial reduction in spending of $5 billion and thereby, through the spending multiplier (of 4 with an MPC of 0.75), work to bring about the needed $20 billion decrease in aggregate demand. This, again, would be represented by an inward shift in the aggregate demand curve from AD_1 to AD_2.

Suppose the government decides to use fiscal policy to reduce the CPI from 160 to 155 because it is concerned about the rate of inflation in the economy. One possible way of doing this is by cutting government spending.

Given a marginal propensity to consume of 0.75, the spending multiplier is 4. Cutting government spending by $5 billion will therefore reduce aggregate demand by $20 billion (4 × $5 billion). As shown in **Exhibit 17.3**,

such a reduction in AD (note that the horizontal distance between point E_1 on AD_1 and point E' on AD_2 is \$20 billion) from AD_1 to AD_2 is enough to establish equilibrium at E_2, with a price level of 155.

What is happening here is that the \$5 billion cut in government spending produces a total decrease in aggregate demand of \$20 billion, represented in the exhibit by an inward shift of the aggregate demand curve (from AD_1 to AD_2). The result is a temporary excess aggregate supply of \$20 billion in the economy, measured by the distance from E' to E_1. This excess supply creates pressure on firms to reduce prices and so inflation consequently cools, with the final new equilibrium being E_2.

It should be emphasised that the level of the CPI would rarely drop in real-world modern economies.[2] In a real-life situation, the government might want to reduce the CPI rate of growth – that is, reduce the inflation rate from, say, 3 per cent back down to, say, 2.5 per cent.

As explained in the previous chapter, one way to try to reduce the rate of price inflation – and by far the more usual approach in the real world – would be for the country's central bank to tighten monetary policy. An alternative approach would be to use fiscal policy as just described. The essential mechanism is as above, but the outcome, if successful, would be a reduction in the rate of inflation rather than an absolute reduction in the CPI level.

Another approach to the inflation problem would be for the government to raise taxes. Although this approach is often considered political suicide, let's calculate the amount of the tax hike required to reduce aggregate demand by \$20 billion.

To see this, think about how much the government would need to cut taxes if it were trying to increase aggregate demand by \$20 billion. The amount we calculate will also be the same amount by which the government would need to raise taxes to reduce demand by the same amount.

With a spending multiplier of 4, we know that an initial increase in spending of \$5 billion will produce an increase in aggregate demand of \$20 billion. In order to bring about an initial spending increase of \$5 billion, the government would need to reduce taxes by \$6.67 billion if the MPC was 0.75 (\$5 billion = 0.75 × \$6.67 billion). In other words, if disposable income were to rise by \$6.67 billion, people would only spend 75 per cent of this amount, namely \$5 billion. Alternatively, if the government wanted to *reduce AD* by \$20 billion, taxes would need to be *raised* by \$6.67 billion.

Finally, as already mentioned above, we need to emphasise that, although theoretically fiscal policy could be used to try to curb an inflationary situation as illustrated in **Exhibit 17.3**, in reality the job of trying to control inflation these days falls to the country's central bank and its appropriate use of monetary policy.

The balanced budget multiplier

One approach to fiscal policy that first gained support during the 1980s and 1990s – and still receives strong support from many analysts, commentators and policymakers today – was the idea that the government should be required to match or 'balance' any new spending with new taxes; in other words, a requirement that new spending initiatives should have – in the first instance – a neutral effect on government finances. The proponents of this type of approach argue that it represents a more prudent way of financing government.

Understanding the impact on the economy of this type of fiscal policy requires consideration of the so-called **balanced budget multiplier**. What would be the size of the multiplier if a spending increase by the government was, at the same time, exactly offset by an increase in taxes? It turns out that the multiplier in this instance – the so-called balanced budget multiplier – is 1! In other words, despite the offsetting tax, aggregate demand will

Balanced budget multiplier An equal change in government spending and taxes that changes aggregate demand by the amount of the change in government spending.

[2] However, even today and in modern economies, deflation (a falling price level) is not completely unknown. For example, Japan experienced recurring bouts of deflation throughout the 1990s and 2000s and deflation was a very real concern of US policymakers in the years immediately after the GFEC. Australia experienced annual deflation during July 2020 due to the COVID-19 pandemic for the first time since March 1998.

increase by the amount of the increase in government spending. Furthermore, theoretically, this will be the case irrespective of the MPC of the community.

To see how the balanced budget multiplier works, suppose a government proposes to increase petrol excise tax by $1 billion with the proceeds hypothecated and spent on a $1 billion increase in government spending for national highways – this situation was essentially precisely what the Abbott government proposed in its 2014–15 federal Budget. Let's look at the macroeconomic impact of this.

With an MPC of 0.75, the $1 billion increase in direct government spending will, with a spending multiplier of 4, generate an overall increase in aggregate demand of $4 billion. The increase in petrol tax will reduce disposable income by $1 billion. However, given that some (25 per cent) of this income would have been saved had it not been taxed away, this means that the direct reduction in consumption spending will only be $0.75 billion. Again, this initial spending reduction of $0.75 billion will have a multiplied impact on overall demand and, given the same spending multiplier of 4, will result in a reduction in aggregate demand of $3 billion. Thus, the net addition to total aggregate demand of this balanced budget increase in government spending is just $1 billion ($4 billion – $3 billion).

Notice also that the same outcome would have occurred for any value of the MPC. In other words, the balanced budget multiplier is always equal to 1, and the net change in aggregate demand is the amount of the initial change in government spending.

Automatic stabilisers

Automatic stabilisers
Federal expenditures and tax revenues that automatically change over the course of the business cycle in such a way as to help stabilise an economic expansion or contraction.

Unlike discretionary fiscal policy, **automatic stabilisers** are built-in mechanisms that automatically help fight unemployment and inflation without there being any need to explicitly alter spending initiatives or tax laws. Automatic stabilisers are federal expenditures and tax revenues that automatically change over the course of the business cycle in such a way as to help stabilise an economic expansion or contraction.

Exhibit 17.4 illustrates the influence of automatic stabilisers on the economy. The downward-sloping line *E* represents federal government expenditures, including such *transfer payments* as unemployment benefits and other forms of social welfare payments. This line falls as real GDP rises.

To understand this, consider that when the economy expands, unemployment, for example, falls and government spending on unemployment benefits, therefore, also falls. During a downturn, people lose their jobs and government spending automatically increases because unemployed individuals become eligible for unemployment benefits and other transfer payments.

On the other hand, the direct relationship between tax revenues and GDP is shown by the upward-sloping line *T*. During an expansion, jobs are created, unemployment falls, and workers and companies earn more income and therefore pay more taxes. Thus, income tax collections automatically vary directly with the growth in real GDP.

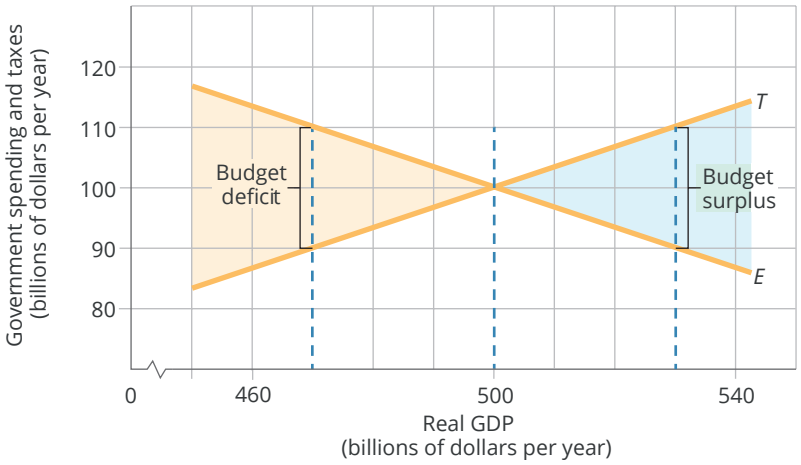
In **Exhibit 17.4**, suppose the federal budget is initially in balance. Federal spending, *E*, is equal to taxes collected, *T*, and the economy is in equilibrium at $500 billion GDP. Now assume consumer optimism soars and a spending spree increases the consumption component (*C*) of total spending. As a result, the economy moves to a new equilibrium at $530 billion GDP.

The rise in GDP creates more jobs and generates higher tax collections. Consequently, taxes rise to $110 billion on line *T*. At the same time, with increased economic activity, employment grows and the government's expenditure on unemployment benefits falls to, say, $90 billion on line *E*. The vertical distance between lines *T* and *E* represents a federal **budget surplus** of $20 billion. A budget surplus occurs when government revenues exceed government expenditures in a given time period.

Budget surplus
A budget situation in which government revenues exceed government expenditures in a given time period.

Now begin again with the economy at $500 billion in **Exhibit 17.4** and assume that business managers lower their profit expectations, so they cut investment spending (*I*), causing aggregate demand to decline. The corresponding decline in GDP from $500 billion to $470 billion causes tax revenues to fall from $100 billion to $90

Exhibit 17.4 **Automatic stabilisers**



CAUSATION CHAIN

Increase in real GDP → Tax collections rise and government transfer payments fall → Increasing budget balance assists in mitigating inflationary pressure

CAUSATION CHAIN

Decrease in real GDP → Tax collections fall and government transfer payments rise → Decreasing budget balance assists in mitigating the recession

Federal government expenditure is represented by the downward-sloping line *E*. This means that government spending for welfare and other transfer payments declines as real GDP rises. Taxes, on the other hand, vary directly with real GDP. Thus, if the real GDP falls below $500 billion, the budget deficit rises automatically. The size of the budget deficit is shown by the vertical distance between lines *E* and *T*. This budget deficit assists in mitigating a recession because it stimulates aggregate demand. Conversely, when the real GDP rises above $500 billion, a federal budget surplus increases automatically and assists in mitigating inflationary pressure.

billion on line *T*. At the same time, the reduced rate of economic activity increases unemployment and therefore increases the government's expenditure on unemployment benefits to $110 billion on line *E*.

The combined effect of the fall in taxes and increased government expenditure creates a **budget deficit**, which occurs when government expenditures exceed government revenues in a given time period. The vertical distance between lines *E* and *T* at $470 billion real GDP illustrates a federal budget deficit of $20 billion.

Budget deficit A budget situation in which government expenditures exceed government revenues in a given time period.

The key feature of automatic stabilisation is that it 'leans against the prevailing wind'. In short, changes in federal spending and taxes moderate changes in aggregate demand. When the economy expands, the fall in government spending for transfer payments and the rise in the level of taxes result in a budget surplus. As the budget surplus grows, people are paying more money to the government than it is spending, which applies braking power against further increases in real GDP.

When the economy contracts, the rise in government spending for transfer payments and the fall in the level of taxes yield a budget deficit. As the budget deficit grows, people are receiving more money from the government to spend than the government is receiving in taxes, which helps to ameliorate the decline in real GDP.

In summary

- **Discretionary demand-side fiscal policy** is the deliberate use of changes in government spending and/or taxes to alter aggregate demand.
- The initial action by the government is multiplied through the action of the spending multiplier, which in turn is influenced by the marginal propensity to consume (MPC).
- If any spending change is offset by a tax change of the same size – a balanced budget situation – the **balanced budget multiplier** is one.
- **Automatic stabilisers** are built-in mechanisms that automatically help fight unemployment and inflation without there being any need for the government to explicitly alter spending initiatives or tax laws.

2 Supply-side fiscal policy

Supply-side fiscal policy
A fiscal policy that emphasises government policies that increase aggregate supply in order to achieve long-run growth in real output and full employment (along with a possibly lower price level).

The focus so far has been on fiscal policy that affects the macro economy solely through the impact of government spending and taxation on aggregate demand. On the other hand, **supply-side fiscal policy** emphasises government economic policies that increase aggregate supply in order to achieve long-run growth in real output and full employment (along with a possibly lower price level). In the next two sections, you will gain an understanding of the concept of, and rationale behind, supply-side fiscal policy, and appreciate how it differs from discretionary demand-side fiscal policy.

Supply-side policies first became an active economic idea in the early 1980s. As discussed in **Chapter 13**, many economies in the 1970s experienced high rates of both inflation and unemployment – such a combination being difficult to explain solely in terms of inadequate aggregate demand. This set the stage for a new approach to macroeconomic policy.

Suppose the economy is initially at E_1 in part (a) of **Exhibit 17.5**, with a CPI of 150 and an output of $480 billion real GDP. The economy is experiencing high unemployment, so the goal is to achieve full employment by increasing real GDP to $500 billion. As described earlier in this chapter, the federal government might follow an aggregate demand oriented expansionary fiscal policy and move to increase aggregate demand, represented by a shift outward in the aggregate demand curve from AD_1 to AD_2.

Higher government spending or lower taxes operate through the multiplier effect and cause this increase in aggregate demand. The good news from such a demand-side fiscal policy prescription is that the economy moves towards full employment, but the bad news is that the price level rises. In this case, *demand-pull inflation* would cause the price level to rise from 150 to 200.

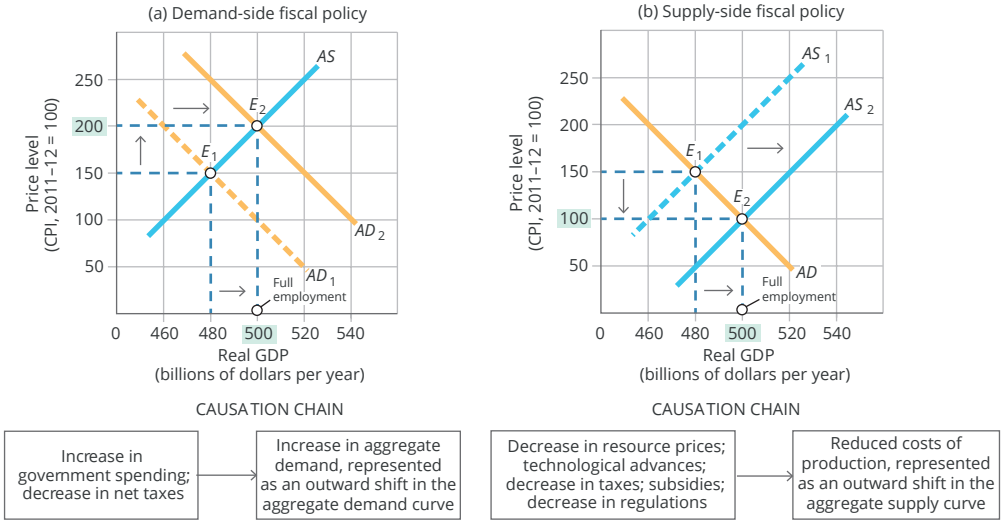
Part (b) of **Exhibit 17.5** represents a supply-side fiscal policy alternative. Again, suppose the economy is initially in equilibrium at E_1. Supply-side economists would argue that the federal government should adopt policies that encourage greater aggregate supply in the economy, represented by a shift rightward in the aggregate supply curve from AS_1 to AS_2.

An increase in aggregate supply would move the economy to E_2 and achieve the full-employment level of real GDP. Under supply-side theory, there is an additional bonus to full employment. Instead of the price level rising, as in part (a) of **Exhibit 17.5**, the price level in part (b) falls from 150 to 100. Comparing the

two graphs in **Exhibit 17.5** it seems evident that supply-siders seem to provide a better policy prescription than proponents of demand-side fiscal policy when both inflation and unemployment are concerns. The demand-side fiscal policy options in **Exhibit 17.5** listed under the top graph are drawn from column 1 of **Exhibit 17.1** and the supply-side policy alternatives listed under the bottom graph are drawn from column 2 of **Exhibit 14.10** in **Chapter 14**.

For supply-side economics to be effective, the government must implement policies that increase the total output that firms want to produce at any price level, depicted in **Exhibit 17.5** by an outward shift in the *AS* curve. A shift in aggregate supply can be accomplished by some combination of: cuts in resource prices (e.g., by enacting various economic reform policies aimed at increasing the efficiency with which such resources are produced); policies enacted to encourage (the pace of) technological advances; increased investment allowances for companies and/or reductions in a range of company taxes; and reductions in government regulation (thereby reducing compliance costs).

Exhibit 17.5 Demand-side versus supply-side effects



In part (a), assume an economy begins in equilibrium at point E_1, with a price level of 150 and a real GDP of $480 billion. To boost real output and employment, one policy approach would be aimed at increasing aggregate demand and would prescribe the federal government raising government spending or cutting taxes. Such demand-side policies would be represented in the exhibit as a shift in the aggregate demand curve from AD_1 to AD_2. As a result, the equilibrium changes to E_2, where the real GDP rises to $500 billion, but the price level also rises to 200. Hence, full employment has been achieved at the expense of higher inflation.

The initial situation for the economy at point E_1 in part (b) is identical to that shown in part (a). However, a supply-side approach would offer a different fiscal policy prescription. Using some combination of policy-induced cuts in resource prices, technological advances, tax cuts, subsidies and regulation reduction, supply-side fiscal policy reduces the cost of production, represented as a shift in the aggregate supply curve from AS_1 to AS_2. As a result, the equilibrium in the economy changes to E_2 and the real GDP increases to $500 billion, just as in part (a). However, with this approach, the price level falls to 100, rather than rising to 200.

An example of supply-side economic policy initiatives: microeconomic reform in Australia

Much of the microeconomic policy reforms pursued in Australia from the early 1980s into the 1990s and 2000s may be viewed as having a strong supply-side flavour to them, and most would agree they represented a vital component of the foundations for a sustained period of economic prosperity and higher productivity in Australia from the end of the 1990–92 recession right through to well into the 2010s. The Australian economy has been resilient to significant economic shocks largely due to the flexibility of the Australian economy and a series of macroeconomic and microeconomic reforms over the past 40 years or so. Major policy reforms over the period in question included:[3]

- the reduction of tariffs – the effective rate of assistance to manufacturing fell from 35 per cent in the early 1980s to under 5 per cent by the early 2000s (see **Chapter 18** for further discussion); this means that a wide range of manufacturing businesses needed to find more efficient ways of operating in order to compete with imported products. The reduction in tariffs decreased the manufacturing's share of the output at first but increased the real output in the sector
- reforms to Australia's financial system including the removal of interest rate controls, the floating of the Australian dollar in March 1983 along with the removal of foreign exchange controls and foreign bank entry being allowed so that by the early 2018s, 45 foreign banks were operating in Australia as either subsidiaries (eight) or as branches (37) of foreign-owned banks, all of which significantly increased the efficiency with which Australia's financial system operated
- the introduction of a range of competition policy reforms – from 1995, under a coordinated national program called 'National Competition Policy', broad-ranging reforms to the provision of major business inputs like energy and road transport, reforms to the way government businesses were required to operate, and anti-competitive regulation was introduced across all Australian jurisdictions, all producing very significant increases in production efficiency
- beginning with the first 'Prices and Incomes Accord' in 1983, a range of labour market reforms – such as the move away from centralised wage fixing to enterprise awards and enterprise agreements and the introduction of individual employment contracts – occurred to allow greater flexibility in work practices and wages across Australia, which again acted to reduce business costs
- taxation reform, including the introduction of capital gains tax and dividend imputation in the 1980s, subsequent reductions in the company tax rate, the introduction of the GST (Australia's version of a broad-based consumption tax), along with the lowering of personal income tax rates – all to increase incentives to work, save and invest in productive capital.

As stated above, most of the reforms had the aim of increasing the efficiency of production processes in Australia, thereby reducing the costs associated with any given level of supply of goods and services. In other words, in the context of the *AD–AS* framework, such policies were supply-side in nature and aimed at increasing aggregate supply in the economy, represented by shifting out of the aggregate supply curve.

[3] This section draws on 'Structural reform Australian-style: lessons for others?', 2005, by Gary Banks, then Chairperson of the Productivity Commission, see http://www.oecd.org/australia/39218531.pdf, accessed on 22 June 2017 and 'Australia's experience with economic reform', 2018, by Laura Berger-Thompson, John Breusch and Louise Lilley, see https://treasury.gov.au/sites/default/files/2019-03/p2018-t332486-economic-reform-v2.pdf, accessed on 13 October 2020.

The idea of using various government policies, such as more favourable taxation allowances to individuals and/or businesses, to increase aggregate supply is not without controversy. Despite its logic, many economists argue that the magnitude of any increase in aggregate supply is likely to be small or might occur only in the longer run. For example, they argue that it may take many years before more favourable taxation arrangements for business generate any substantial change in actual plant and equipment or technological advances. Thus, as a short-term countercyclical macroeconomic policy tool, it may not be as useful as policies directed more at shifting aggregate demand.

Applicable concept: supply-side fiscal policy

Analyse the issue

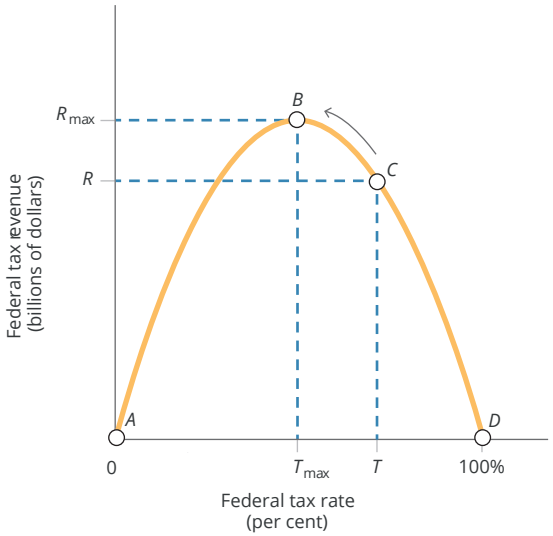
The Laffer curve

Supply-side economics became popular during the early 1980s. The fiscal policy prescription of income tax cuts came to be associated with the American supply-side economist Arthur Laffer. The basic idea is easily explained using the so-called **Laffer curve**, which is a graph depicting the relationship between tax rates and total tax revenues.

Laffer curve
Representation of the relationship between the income tax rate and the amount of income tax revenue collected by the government.

As shown in the graph (**Exhibit 17.6**), the hypothetical Laffer curve can be drawn with the income tax rate on the horizontal axis and tax revenue on the vertical axis. The idea behind this curve is that the income tax rate affects the incentive for people to work, save, invest and produce, which in turn influences tax revenue. As the tax rate climbs, Laffer and other supply-siders argued that the erosion of incentives shrinks national income and total tax collections.

Exhibit 17.6 **The Laffer curve**



Suppose the federal government sets the federal income tax rate at zero (point *A*). Here people have the maximum incentive to produce. But there is zero income tax revenue for the government. Now assume the federal government sets the income tax rate at the opposite extreme of 100 per cent (point *D*). At a 100 per cent income tax rate, people have no reason to work, produce and earn income. People seek ways to reduce their tax liabilities by engaging in unreported or underground transactions or by not working at all. As a result, again, no income tax revenue is collected by the authorities.

Because the federal government wishes to collect some income tax revenue, it sets the income tax rate between zero and 100 per cent. Assuming there is the tax-rate/tax-revenue relationship depicted in the graph, maximum tax revenue, R_{max}, is collected at a tax rate of T_{max} (point *B*). Laffer argued that, in the case of the US, the federal income tax rate of *T* (point *C*) in 1981 exceeded T_{max} and the resulting tax revenue of *R* was below R_{max}.

In Laffer's view, reduction of the income tax rate would lead to an increase in tax revenue because people would increase their work effort, saving and investment, and would reduce their attempts to avoid paying taxes. Thus, the Laffer curve policy argued that a cut in federal income tax rates would unleash reported economic activity and boost tax revenues to such an extent that the federal government would actually raise more revenue with the lower tax rates than with the higher rates.

The Laffer curve remains a controversial part of supply-side economics. There is still considerable uncertainty about the shape of the Laffer curve and at what point – *B*, *C* or otherwise – the economy might be operating at any given time. Thus, the practical usefulness of the Laffer curve is a matter of debate.

What do you think?

Based on what you have just read, compare the common perception of how a tax rate cut affects tax revenues with Laffer's theory.

Notwithstanding this, supply-siders may well argue that the long 10-year US economic expansion of the 1990s (from March 1991 to March 2001) was at least partially the result of the earlier supply-side type tax inducements of the 1980s implemented by the Reagan administration in that country to encourage firms to invest in research and development. It is certainly widely acknowledged that the very rapid growth in productivity in the second half of the 1990s in the US was driven by technological advances, particularly in computing and communications, and it could well be argued that these ICT advances were significantly stimulated by earlier R&D investment tax incentives.

In summary

- **Supply-side fiscal policy** emphasises government policies that increase aggregate supply in order to achieve long-run growth in real output and full employment.
- A shift in aggregate supply can be accomplished by some combination of policies that cut energy and other production costs, reduce government regulation, encourage technological advances, increase investment allowances, and/or reduce a range of taxes.
- Beginning in the 1980s Australia has implemented quite a comprehensive set of microeconomic reform policies to increase the productive efficiency of Australian industry – depicted in the *AD–AS* framework as a shifting out of aggregate supply – and which therefore have had a strong **supply-side** flavour to them.

3 The macroeconomic significance of the federal budget

The federal budget is the principal fiscal policy statement by the federal government each year; in Australia it is delivered by the federal Treasurer every May. As well as presenting estimated final figures for government revenues and expenditures for the current financial year (ending, in Australia, in June), it also contains all-important forecasts of the government's financial situation for the coming four years.

Budget speech night is a big media event and the budget is closely scrutinised by economic analysts. Any changes to federal taxation policies and/or expenditure policies are announced in the budget. Also tabled are the federal treasury department's forecasts for the coming year for such macroeconomic variables as inflation, wages growth, unemployment, GDP growth, the current account deficit and so on.

At the time of writing, the most recent federal budget was the 2020–21 Budget, brought down in September 2020. The budget in 2020–21 was different from the normal budgets and was crucial due to the COVID-19 crisis. It slashed the aim of slowly moving Australia towards a budget surplus after a number of years of deficits pre-pandemic. Many years of deficits, together with the huge budget deficit due to COVID-19 resulted in the accumulation of a considerable level of government debt. The cumulated effect of the adjustments to federal government spending proposed in the 2020–21 Budget was to leave the federal budget in a deficit right through to and including 2023–24.

The core components of the 2020–21 Budget was to implement the government's COVID-19 economic recovery plan by supporting Australians with additional COVID-19 response measures and driving job creation. The key measures included additional spending on JobKeeper payments, payments to pensioners and eligible recipients, and to secure access to potential vaccine doses. Also, there was funding to create jobs, tax relief to households and businesses, increases to the government's infrastructure investments and support for new entrants to the job market and women in the workforce. Moreover, there was additional funding for the National Disability Insurance Scheme (NDIS) and to regional Australia.

Of particular interest to analysts is the size of the budget deficit/surplus, which is, as noted above, the difference between total federal government outlays and receipts. Federal budget outcomes (as a percentage of GDP) for the past 30 plus years are provided in **Exhibit 17.7**. The budget outcome has been both positive and negative over this period, but note, in particular, the projected 12-year run of deficits (negative outcomes) from 2008–09 through until 2023–24 inclusive.

Exhibit 17.7 **Australian general government sector cash revenue, outlays and surplus**

	Revenue (% of GDP)	Outlays (% of GDP)	Underlying cash surplus (% of GDP)
1981–82	23.6	23.4	0.2
1982–83	24.0	25.7	−1.8
1983–84	23.4	26.6	−3.3
1984–85	24.9	27.5	−2.6
1985–86	25.4	27.3	−2.0
1986–87	26.1	26.9	−0.8
1987–88	25.5	25.1	0.4
1988–89	24.5	23.1	1.5
1989–90	24.2	22.7	1.5
1990–91	23.9	24.0	−0.1
1991–92	22.4	25.4	−3.0
1992–93	21.7	25.8	−4.0
1993–94	22.0	25.8	−3.9
1994–95	22.7	25.5	−2.8
1995–96	23.4	25.5	−2.1

1996–97	23.9	25.0	−1.1
1997–98	23.8	23.8	0.0
1998–99	24.4	23.8	0.6
1999–2000	25.0	23.1	2.0
2000–01	25.8	25.0	0.8
2001–02	24.7	24.8	−0.1
2002–03	25.4	24.5	0.9
2003–04	25.2	24.3	0.9
2004–05	25.5	24.0	1.5
2005–06	25.6	24.0	1.6
2006–07	25.0	23.2	1.6
2007–08	25.0	23.0	1.7
2008–09	23.2	25.1	−2.1
2009–10	22.0	26.0	−4.2
2010–11	21.5	24.6	−3.4
2011–12	22.2	25.0	−2.9
2012–13	23.1	24.1	−1.2
2013–14	22.7	25.6	−3.0
2014–15	23.4	25.5	−2.3
2015–16	23.4	25.6	−2.4
2016–17	23.2	24.9	−1.9
2017–18	24.2	24.5	−0.5
2018–19	24.9	24.5	0.0
2019–20	23.7	27.7	−4.3
2020–21 (e)	23.8	34.8	−11.0
2021–22 (e)	22.5	28.0	−5.6
2022–23 (e)	23.0	27.2	−4.2
2023–24 (e)	23.9	26.9	−3.0

(e) estimates

Author's note: From 2005–06 through to 2019–20 inclusive, the underlying cash balance per cent of GDP does not equal the difference between revenue and expenditure due to the actual and expected earnings of the federal government's 'Future Fund'.

Based on Commonwealth of Australia data. Source: 2020–21 Australian Government Budget – Budget Paper No. 1, Statement 11, Table 1.

These deficits (actual and estimated) were both the result of the impact of the GFEC and COVID-19 on the Australian economy – the country experienced a period of reduced economic growth, leading to the impact of the automatic stabilisers mentioned earlier coming into play – as well as the result of the government's use of active discretionary countercyclical fiscal policy to try to reduce the impact of the GFEC and COVID-19 and avert a serious recession. Whilst more will be said on the role of fiscal policy in combating the GFEC and COVID-19 in **Chapter 19**, many would argue that, notwithstanding the GFEC crisis, earlier and stronger action should

have been taken to return the federal budget to surplus. However, COVID-19 impacts will see the Australian government's budget in deficit at least for another ten years to come.

The overall budget outcome is of interest for a couple of reasons. First, federal outlays as a percentage of GDP represent the size and influence of the federal government in the economy. Those in favour of a reduced role for government look for reductions in this percentage over time. On this, notice that this percentage had declined slightly from 2000–01 to 2007–08 – from 25 per cent down to 23 per cent. It then quite quickly increased to a peak of 26 per cent in 2009–10 – at the time of the government's dramatic fiscal response to the GFEC. The situation got worse with the COVID-19 pandemic in 2019–20 and the government spending increased to 34.8 per cent in 2020–21 and was projected to remain at 27 per cent even in 2023–24, a situation those in favour of a reduced government role and influence in the economy would be lamenting.

Second, the size of the deficit/surplus is interpreted in many quarters as a summary overall measure of the government's **fiscal stance** – that is, the extent to which the authorities are actively trying to expand or retard the activity in the national economy through the settings of government revenue-raising and spending policies. A larger deficit, because it means the government is spending more than it is receiving, is interpreted as expansionary, with the prospect of there being a stimulus to economic activity with a concomitant increased risk of higher inflation and higher interest rates. On the other hand, a larger surplus is interpreted as a brake on economic activity along with the expectation of lower future inflation and interest rates.

Fiscal stance
The term used to represent the extent to which the authorities are actively trying to expand or retard the growth of the economy through the settings of government revenue-raising and spending policies.

On this, as far as the 2020–21 Budget was concerned, as noted previously, the government increased its outlays in 2019–20 to 34.8 per cent from a recent 24.5 per cent in 2017–18, this in part to counteract the COVID-19 induced recession in Australia. In dollar terms, this represented a very substantial $225 billion increase in federal government outlays in 2020–21 compared with 2017–18. A similar increase in federal government spending was observed as a response to avoid GFEC induced recessions. In hindsight, many have argued that it was unnecessary for the Australian government fiscal response to the GFEC to be of this magnitude (indeed, some economists argued at the time that it was significantly excessive), and we will look at this issue again in more detail in **Chapter 19**.

Putting this to one side, for now, irrespective of debate about the appropriateness of the size of the Australian fiscal response in 2009–10, it is of considerable concern that even eight years later in 2017–18 this expenditure percentage was still at 24.5 per cent. Unfortunately, the goal of the federal government reaching a budget surplus will not be achieved soon due to the impact of COVID-19. In fact, the federal government outlays during 2020–21 (as a response to COVID-19) was $340 billion more than the outlays in 2009–10 (as a response to the GFEC). Furthermore, as noted earlier, the federal budget was projected to remain in a deficit right through to at least 2023–24, meaning that the government was projected to spend more than it receives in revenue in each year through till then. For this reason, some analysts argued the government needed to go much further in cutting back its outlays than it was intending so it can return the budget to surplus sooner.

The countercyclical role of fiscal policy

As described earlier in this chapter, many economists argue that it is an appropriate role for authorities to purposely manipulate aggregate demand – by varying government taxation and spending – in order to stabilise the business cycle. This is termed 'discretionary fiscal policy activism' and means, for example, that government should respond to unacceptably high unemployment levels by temporarily supplementing private-sector demand with increases in government expenditure to reflate the economy. Such a countercyclical role for fiscal policy implies large deficits in recessionary times offset by surpluses (reduced government spending and increased tax revenue from the higher levels of activity) in growth periods.

While there has undoubtedly long been a vital element of such discretionary activism in Australia's approach to the conduct of fiscal policy, the 1997 budget speech by the then Treasurer (Peter Costello, AC) indicated that there would be a greater fiscal discipline exercised by the government in future years to ensure that the federal budget would remain in balance, on average, throughout the business cycle. This approach would still allow

fiscal policy to be appropriately stimulatory (i.e., to run deficits if necessary) during times of economic slowdown or recession. However, it would then require it to be tightened again once economic recovery was underway.

Consider again **Exhibit 17.7** in the context of the 1990–91 recession. This recession was well and truly over by the end of 1992, yet the federal budget remained in deficit right through until 1996–97. Given this, many economic commentators argued that fiscal policy had been allowed to be inappropriately stimulatory for too many years into the ensuing long economic expansion and that too much government debt had been accumulated unnecessarily over those years. Indeed, the continuation of budget deficits for so long into the expansion provided the motivation for the incoming government in 1996 to embark on a period of significant fiscal consolidation.

Even more starkly, the projected 10-year run of deficits from 2020–21 through to 2030–31 would certainly be regarded by many as very much at variance with the notion of proper prudent fiscal policy discipline as enunciated above. Given that Australia never actually went into a GFEC-induced recession – and even if this were in part the result of the very significant discretionary fiscal stimulus by the government in 2009–10 – it is easy to understand why so many economic commentators believe the government had not done enough to get the budget back into surplus much sooner. However, the impact of COVID-19 and the huge discretionary fiscal stimulus made by the government during 2020–21 made it unrealistic for the government (as well as for the economic commentators) to bring the economy to a budget surplus much sooner than 2030.

Other roles of fiscal policy

Apart from its short-term business cycle stabilisation role, fiscal policy is also used more generally to try to achieve desirable medium- to long-term economic and social goals. The introduction of the GST in Australia in July 2000 is an example of a fiscal policy initiative with medium- to long-term economic objectives. Briefly, these objectives were to:

- broaden the range of goods and services attracting tax, thereby removing a lot of hitherto tax-induced distortions to consumption patterns and allowing the burden of taxation to be spread more widely and evenly across the economy
- reduce the reliance on income taxation in government revenue-raising
- reduce the ability of individuals to avoid paying their legally obliged taxes
- encourage increased rates of community saving by taxing income when it was spent rather than when it was earned.

Another example of recent fiscal policy changes in Australia with a clear medium-term focus has been a range of policy initiatives focusing on reducing Australia's unemployment rate. Of course, the best and most sustainable way of reducing unemployment is to employ appropriate macroeconomic policies to keep the nation's rate of economic growth as strong as possible without precipitating an outbreak in inflation. However, it is also important to ensure there are the right micro-incentives in place to encourage the appropriate responses from the unemployed and from prospective employers. Relatively recent policy initiatives in these areas include:

- the creation of a publicly funded private-sector job network
- mutual obligation initiatives, whereby in return for unemployment benefits the unemployed are required to actively seek employment, to either engage in training and/or education to improve their job-readiness, or otherwise to work in a wide range of part-time, volunteer or community-oriented projects
- limited employment subsidies provided to employers to encourage the hiring of long-term unemployed
- the removal of financial disincentives that discourage people from accepting work offers. These disincentives arise from the sometimes-unhelpful interactions between the welfare and taxation systems, which result in some people on welfare support facing very high effective marginal tax rates on small amounts of earned income.

Some words of caution

As noted above, while the overall size of the deficit/surplus is widely interpreted as a summary measure of the extent to which government fiscal policy is meant to be expansionary/contractionary, several important caveats are worth noting.

First, in determining the macroeconomic impact of a given budget, we need also to consider the actual pattern of expenditure and revenue in addition to the bottom-line deficit/surplus figure. A simple example will clarify the importance of this. Imagine that the government wants to provide an expansionary stimulus to the economy. There are many alternatives open to it to try to accomplish this.

On the revenue side it could reduce taxes, but which taxes? It can choose from company taxes, payroll taxes (a state government tax which a state government might well decide to vary in its annual state Budget), import taxes, excise taxes or personal income taxes, to name just a few. Since these taxes initially affect different sections of the community, we can expect them to have varying macroeconomic outcomes. For instance, cutting the personal marginal income tax rate of high-income earners could be expected to have a different impact on consumption than cutting the taxes of low-income earners because of their different marginal propensities to save.

Alternatively, the government can alter the extent of income tax concessions accruing to various classes of expenditures. For example, it could alter the tax treatment of business investment – such as changing depreciation allowances – which would have its direct impact on aggregate investment rather than consumption.

On the expenditure side, the government can choose broadly from either increasing its purchases of goods and services or increasing its transfer payments. The purchases decision can be further broken down into a choice between spending to increase the nation's capital stock (e.g., building more roads) and spending on current consumption (e.g., holding more Parliament House functions). On the other hand, the impact on the economy of increased transfer payments will depend on the socioeconomic characteristics of the recipients of the increased welfare payments.

A second caveat involves the importance of distinguishing between the domestic budget deficit and the overall deficit. While the federal government's receipts are almost entirely derived from domestic sources, a significant percentage of its spending is on foreign production (e.g., spending on defence equipment).

The **domestic deficit** is the difference between domestically derived revenue and domestic government expenditure. Hence, the overall deficit may increase due to additional government spending on imports without any increase in the domestic component at all, and this spending will not have any expansionary impact on the economy whatsoever. Therefore, it is important to look at the domestic deficit as well as the overall deficit in analysing the likely expansionary impact of the budget.

Domestic deficit
The difference between domestically sourced government revenue and domestic government expenditure.

A third caveat involves the so-called **cyclically adjusted budget deficit**. It is important to recognise that as an economy moves through a business cycle over time, the government's finances are automatically affected by the varying levels of economic activity occurring at the different stages of the cycle. The cyclically adjusted budget deficit is the actual recorded budget deficit for some period adjusted for the stage of the economy's business cycle during the period.

Cyclically adjusted budget deficit
The actual recorded budget deficit adjusted for the stage of the economy's business cycle.

As mentioned earlier in the discussion of automatic stabilisers, during periods of slow economic activity, government expenditure on unemployment benefits will increase while government receipts will decline due to lower taxation revenue. Thus, even without any conscious government fiscal policy decision to try to reflate the economy, the budget deficit will increase as a result of the cyclical impact of the slowdown. Similarly, the budget deficit will automatically reduce and move into surplus as economic activity accelerates during the cyclical upturn.

Although such changes in the budget outcome provide some useful information about the overall macroeconomic effects of government activities, they occur passively. They, therefore, give little indication of whether the government is actively trying to use fiscal policy to expand or retard economic activity; that is, they give little indication of the government's active fiscal policy stance.

To gauge the active stance of fiscal policy more precisely, economic analysts often adjust the actual deficit to account for such passive cyclical effects. For example, consider two periods, the first of which was characterised by cyclically high unemployment and the second by cyclically low unemployment. The actual measured budget outcomes would be converted into cyclically adjusted budget outcomes by appropriately reducing the size of the actual budget outcome in the first period and increasing it appropriately in the second. The resulting adjusted budget outcomes may then give a more precise indication of what the government's active fiscal stance was in the two different business cycle periods.

Economics and ethics — Is fiscal policy fair?

This is an appropriate point to refer back to our discussion in **Chapter 16** about the 'bluntness' of monetary policy as a policy instrument. It is certainly the case – as shown in our discussion about the range of options available to the government when it adjusts fiscal policy should illustrate – that fiscal policy may be regarded as an instrument that can be used to impact more precisely on some sections of the community and/or sectors of the economy than on others.

Of course, through the spending multiplier effects, any change in taxation or spending policies by the government is likely eventually to spread throughout the economy and community. Nonetheless, it will undoubtedly be true that some will feel the effects much more directly and significantly than others. This fact inevitably leads to considerations that reach far beyond the realm of straight economics. It may be, for example, that most analysts believe that the government's discretionary expansionary fiscal stance is expected to result in economic activity is 4.5 per cent higher by 2021–22 and lower the unemployment rate around 5 percentage points. This is to be achieved by lowering the tax rates and expanding government spending. This was a fairly widespread view held by policymakers, analysts and commentators.

This fact inevitably leads to considerations that reach far beyond the realm of straight economics. It may be, for example, that most analysts agree that a fiscal tightening (increased spending or/and decreased taxation) is in order – in the lead up to the 2020–21 Budget. However, there may well be – and usually is – very great disagreement as to which group(s) should have their taxes decreased or, alternatively, whether some welfare payments should be increased (again, for which groups?); whether increased spending should be in public housing, or health and medical services; or whether it should instead be on public economic infrastructure and where in the country should these infrastructure spending rises occur.

All of these issues will inevitably involve ideological and political dimensions with the result that some in the community may question the 'fairness' and appropriateness of the enacted or proposed policy changes. This was very much in evidence in the quite vociferous public commentary and there were going to be immediate winners and losers resulting from the 2020–21 fiscal initiatives.

In summary

The **federal budget** is the principal annual fiscal policy statement by the federal government.

- If government expenditure exceeds revenue the budget is in **deficit**.
- Persistent budget deficits have serious consequences for the level of **government debt**.
- The economic impact of the budget outcome will depend on the **domestic deficit/surplus** and the actual make-up of the proposed spending/taxation changes.
- The budget outcome needs to be adjusted for the stage of the business cycle to better determine the **fiscal stance** of the government.

4 Implications of the budget outcome for government debt levels

Every time the government runs a budget deficit, the **national debt** is increased by the government issuing securities to pay for its expenditures in excess of its tax receipts. One useful statistical definition of Australia's national debt, therefore, is the total face value of outstanding Commonwealth Government securities (CGS) on issue; it represents the extent of previous borrowing by the federal government.

National debt The total face value of outstanding Commonwealth Government securities (CGS) on issue.

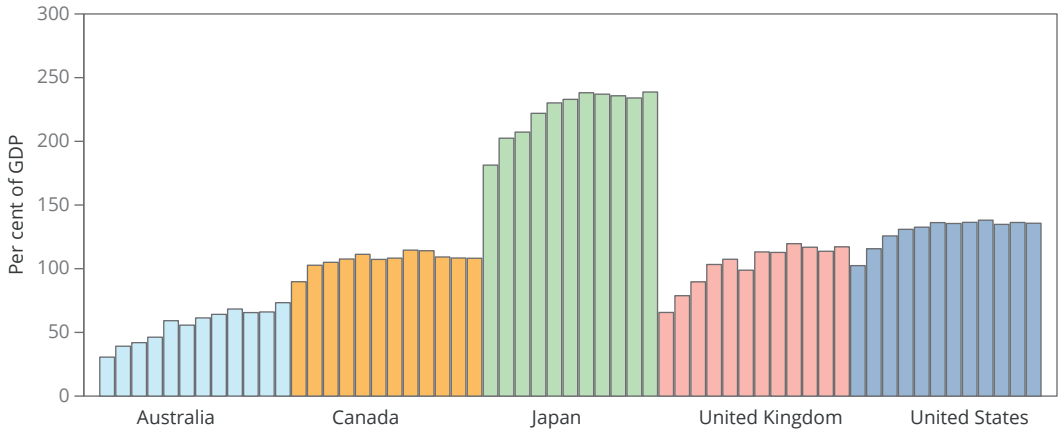
Budget deficits and Australia's national debt

Using this perspective, in recent decades the national debt (defined as outstanding CGS on issue) had fallen fairly steadily to 12 per cent of annual nominal GDP in 1990–91, at which time Australia experienced a deep recession. Debt then rose quite rapidly to around 21 per cent of annual GDP by 1995–96, at which time it again began to fall, reaching just 5 per cent of GDP by 2007–08. It stood at around $53 billion at that time.[4]

Subsequently – and certainly at least to some extent as a result of government actions in response to the GFEC – after a string of budget deficits from 2008–09, it had risen to an estimated $320 billion in 2013–14, or 20 per cent of GDP, representing a quite extraordinary quadrupling of the relative size of the debt in just six years! While still low by international standards, Australia's national debt was significantly higher than it had been in the years immediately before the GFEC. Furthermore, in June 2017, Australia's gross national debt ticked over $500 billion for the first time in history! The debt was projected in the 2017–18 Budget to peak at 31 per cent of GDP in 2018–19, and then to fall slightly to 29 per cent of GDP in 2020–21. At that time, Australia will be paying $20.4 billion annually in interest on its government debt!

Exhibit 17.8 provides an international comparison of various countries' general government debt levels post GFEC and leading up to COVID-19 crisis. As is evident from the exhibit, Australia's debt levels were modest

Exhibit 17.8 **General government debt levels as a percentage of GDP in selected countries (from 2008 to 2019)**



Data source: OECD (2021), General government debt (indicator). doi: 10.1787/a0528cc2-en

[4] In fact, it is more correct to refer to this as the gross debt of the Australian government. Net debt (which subtracts from the gross debt of the Australian government the financial assets held by it) was eliminated in 2005–06, and in 2007–08 was estimated to be running at about – 3 per cent of GDP.

when compared internationally, and allowed the Australian federal government a great deal of manoeuvring room to use discretionary expansionary fiscal policy to combat the effects of the GFEC and COVID-19 crisis.

A portion of a country's national debt will be held by the central bank and corresponds to that part of the debt that has been monetised. This is because the bank, in purchasing the government securities, pays for them by simply creating money. Therefore, this part of the government's outstanding debt is transformed into increases in the country's money stock. The rest of the debt will be held by the general public in the form of interest-bearing liabilities of the government.

Thus, to finance a budget deficit, the government must sell securities. If any of these securities are purchased by the central bank, then part of the deficit is monetised; financing a deficit in this way is known as **money financing**. The government pays for its expenditures by using this newly created money. The rest of the deficit is financed by the government selling the securities to the general public; that is, by the government borrowing from the public. Financing a deficit in this way is known as **debt financing**.

Through to the 1970s, it was quite common in Australia for budget deficits to be financed at least partially through money financing. However, since the early 1980s, budget deficits in Australia have been financed virtually entirely through debt financing.

Money financing of the deficit
When a budget deficit is financed by the central bank purchasing the securities issued by the government to fund its deficit.

Debt financing of the deficit
When a budget deficit is financed by selling government securities to the general public–that is, by the government borrowing from the public to fund its expenditure.

The burden of the debt

This borrowing by the government from the private sector leads to the question of the so-called **burden of the debt**. When the government borrows, the lenders can either be residents or foreigners. If we view the national debt as the collective debt of the Australian people, in 2020–21 it was around $34 200 per person (in gross debt terms). This represented 63 per cent increase in per capita debt from the estimated $21 600 per person in 2017 when the sixth edition of this text was written. To what extent was the $34 200 a burden, and had the debt burden on every man woman and child in Australia also increased by 63 per cent over those three years?[5]

For the moment let's just consider the portion of the national debt that is domestically held. In this case, the answer is almost certainly no, it is not a burden. It is an internal debt that we owe to ourselves. It's like one family member borrowing from another to finance some kind of spending. Although one family member now has a debt obligation to another, the net increase in debt of the family – to everyone else – is zero.

Just as with the family, as far as the nation is concerned, there may certainly be future redistributive effects as the debt matures and the holders of the securities (who may be the original lenders, their heirs or someone else who has subsequently bought the securities) are paid from government tax revenue or from new government borrowing. Of course, this redistribution is perfectly reasonable, since those who previously had lent to the government sacrificed some of their current consumption, which enabled the wider community to benefit from the subsequent government expenditure.

It's also very important to keep in mind that the national debt is only one side of the nation's balance sheet. On the other side is the nation's stock of publicly owned capital assets. If the government borrows to fund capital spending on the construction of additional public economic infrastructure, such spending will increase the nation's productive capacity in a similar way to increases in the capital stock of firms. However, if the government borrows to finance its own current consumption spending rather than to augment the country's national capital stock, it is possible that such debt may become a burden on future generations.

This is an appropriate point at which to introduce the notion of crowding out.

Burden of the debt
The possibility that existing national debt may represent a burden on current and future generations, namely, to pay the future interest payments on it and to eventually repay the principal.

[5] The figure of $34 200 was calculated by taking the 2020–21 stock of outstanding CGS on issue – estimated at around $872 billion (source: *2020–21 Federal Budget Paper No 1*, Statement 11, Table 5) – and dividing this by Australia's estimated population in 2020–21 of 25.5 million, to give $34 200.

The crowding-out issue

Critics of discretionary expansionary fiscal policy believe that increased government spending (designed to boost aggregate demand) that gives rise to federal budget deficits may have little if any effect on real GDP. The reason they give for this is that crowding-out effects may dampen down – or even completely eliminate – the stimulus to aggregate demand arising from the increased federal government spending. The **crowding-out effect** is a reduction in private-sector spending as a result of federal budget deficits financed by government borrowing. How does this occur?

Crowding-out effect
A reduction in private-sector spending as a result of federal budget deficits financed by government borrowing from the private sector.

The reasoning is as follows. Since the federal government competes with other borrowers for available community saving, interest rates experience upward pressure as the government seeks to borrow funds from the financial markets. The result of this rise in interest rates may be lower interest-sensitive consumption and business investment, which would act to offset the boost in aggregate demand from the increased government spending. If so, some private-sector demand would be said to be crowded out by the public-sector spending.

If significant, the crowding-out effect provides one possible avenue through which future generations may bear a burden for the current generation's accumulation of national debt. Recall from **Chapters 2 and 12** that present investment spending increases living standards in the future by increasing the country's capital stock, thereby shifting the production possibilities curve outward. (See **Exhibit 2.4** in **Chapter 2**.) If federal deficits crowd out private investment in plant and equipment, future generations will not have this additional private-sector productive capacity available to them.

The crowding-out theory is not uncontroversial. Some economists would argue that any private-sector crowding-out effects may be unimportant when viewed in comparison to the increases in public-sector capital stock arising from the government spending. Of course, this argument relies on the implicit assumption that the government spending in question is of a capital nature. For example, government capital spending for highways, dams, rail, telecommunications and ports infrastructure financed by borrowing may well offset any possible decline in private-sector investment. On the other hand, this argument does not have as much force if the government spending is of a current consumption nature (say funding pay rises for public servants) rather than capital in nature.

Another argument against the crowding-out idea is the view that consumers and businesspeople may believe the federal deficit spending might be 'just what the doctor ordered' for an ailing economy. For instance, federal budget deficits may be incurred to increase aggregate demand during a period of severely subdued private-sector demand. Spending by the government in this instance, amplified by spending multiplier effects, may in fact stimulate much-needed increases in private-sector consumption and business investment as private sector sentiment improves. This in turn may also raise the profit expectations of businesses, thereby stimulating additional investment spending even in the presence of higher interest rates. The effect of such increased private-sector spending could nullify some or all of the crowding-out effect that would otherwise have offset the boost in aggregate demand from the increased government spending.

Regardless of these arguments about the existence and significance of private-sector investment crowding out, one thing remains reasonably uncontroversial. While borrowing by the government reduces the immediate pool of saving available to finance private investment, and even if this does put upward pressure on interest rates, it should not prove to be a burden on future generations if spending is directed into the provision of public infrastructure. Such expenditure is likely to increase the nation's future income-generating potential and, as such, should not be a problem.

Again, it is important to remember that, when considering the national debt, we should never forget the publicly owned national capital stock that this debt will have financed.

The external component of national debt

The portion of a country's national debt that is owed to foreigners is a different matter. In Australia, at around the time of writing, this proportion was apparently running at around 53 per cent of total outstanding CGS on issue. This percentage had risen sharply from 35 per cent in 2003 but was down significantly from the most recent peak of 76 per cent in 2012. This is an external debt that nonetheless should – if borrowed for the purposes of public capital investment – increase the quantity of output from future production, but will also involve repayment with interest to foreigners.

Thus, this type of debt is not a debt 'owed to ourselves'. As such, in this case, it is important to carefully weigh the potential future income-generating benefits of the capital accumulation financed by the debt against the future obligations the increased debt places on the nation. Indeed, it was estimated that, in 2020–21, around $17.2 billion of interest would be paid on Australia's outstanding CGS. If around 55 per cent of this accrued to foreigners, this would amount to over $9.1 billion in debt-servicing interest payments that would need to be paid to non-residents in 2020–21 by the Australian government.[6] However, the government's interest expense is estimated to be $19.8 billion, of which is expected to be paid $10.5 billion for non-residents in 2020–21.

Of course, foreign borrowing by governments to finance current government consumption is a very dubious practice. Since this borrowing would not finance the creation of new capital, there would be no prospect of increased future national income from which the extra debt could be serviced.

In this regard, a recent relevant case study on externally owed sovereign debt is provided in **Chapter 19**.

[6] The figures quoted in the text were obtained from the *2020–21 Budget Paper No 1*, Statement 7, 'Debt Statement, Assets and Liabilities'.

Global perspective

Applicable concepts: national debt, real versus nominal, and capital budgeting

How real is the national debt?

Perhaps federal budget deficits and the national debt are really not as large as we think. At least, this is the case presented by the late Professor Robert Eisner, past president of the American Economic Association. He said:

> With all the deficits, the general trend of real federal debt – the debt adjusted for inflation – has been downward. On a per capita basis it has indeed gone down very sharply over most of the last forty years.[1]

Eisner and other economists argued that we should use real, rather than nominal, increases in the national debt. For example, Australia's national debt in 2020–21 was estimated to be around $703.2 billion and suppose the CPI level increased by 1.45 per cent in 2020–21. *Ceteris paribus*, the real value of the national debt, therefore, falls by $10.2 billion. According to Eisner, this $10.2 billion inflation adjustment of the national debt should be subtracted from the actual federal budget deficit for 2020–21.

The 2020–21 Australian federal budget deficit was estimated to be $213.7 billion (on an underlying cash balance basis), which is equivalent to 11 per cent of GDP in the 2020–21 Budget. It is expected that the budget position will improve in the short-to medium term and reduce the budget deficit by a significant amount by 2030−31. Using the inflation adjustment approach, the actual 'real' deficit is therefore only $203.5 billion (i.e., $213.7 − $10.2). In effect, the $10.2 billion inflation adjustment may be thought of as just as much a receipt for the federal government as an equal amount of actual tax revenue. Recall from **Chapter 13**, in the discussion of inflation, that increases in the general price

level redistribute buying power from creditors (the general public who hold the government debt securities) to debtors (the federal government).

Eisner also warned that a federal budget balanced using US federal rules of accounting is economically very misleading. Private businesses use two budgets. One is the *operating budget*, which includes salaries, interest payments and other current expenses. The second type of budget is called a *capital budget*. Most governments, including the US government, do not use a capital budget. All expenditures are treated the same way. But capital budgeting allows for spreading the cost of long-lasting assets over future years.

Shutterstock.com/Wael Khalill alfuzai

If a capital budgeting system were utilised, capital spending by the federal government – which would be spending to acquire assets yielding long-term benefits – would be stripped out of the 'operating deficit'. It is worth noting that the Australian Treasury makes an attempt at detailing these distinctions in that the budget 'Operating Balance' excludes net capital investment by the federal government. The budget 'Fiscal Balance' is the result of subtracting net capital investment by the federal government from the operating balance.

[1] R. Eisner, *The Great Deficit Scare*, Century Foundation Press, New York, 1997.

What do you think?

1. Do households make a distinction between spending for current expenses and spending for capital expenses? Compare borrowing $10 000 to take a month's vacation in Hawaii to borrowing $250 000 to buy an apartment and move out of your rented apartment.
2. Using the latest Australian federal budget papers available to you, use Budget Paper 1, Statements 1 and 11 and the annual growth in the CPI from the previous year, and carry out the adjustment outlined in the above box to adjust the current financial year's estimated budget outcome for the change in the real value of the government's stock of outstanding CGS from the previous year.

In summary

- A budget deficit financed by borrowing from the central bank is called **money financing**, and if the funds are borrowing from the public it is called **debt financing**.
- Since the 1980s Australian budget deficits have been financed by debt financing.
- Deficits increase the level of **national debt**.
- Increasing national debt may be a burden on future generations if private sector investment experiences **crowding out** by the deficit spending, and the government spending is for current consumption rather than for capital goods.
- Government borrowing offshore to fund its current consumption is almost certain to lead to future difficulties in servicing and paying back the resulting increase in national debt.

Key concepts

Fiscal policy
Discretionary fiscal policy
Spending multiplier
Balanced budget multiplier
Automatic stabilisers
Budget surplus
Budget deficit
Supply-side fiscal policy
Laffer curve
Fiscal stance
Domestic deficit
Cyclically adjusted budget deficit
National debt
Money financing of the deficit
Debt financing of the deficit
Burden of the debt
Crowding-out effect

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Develop an understanding of discretionary demand-side fiscal policy

- **Discretionary demand-side fiscal policy** is the deliberate use of changes in government spending and/or taxes to alter aggregate demand to try to stabilise the economy's business cycle. It is more commonly used to combat a recession than to reduce inflation.
- In either task the initial action by the government is amplified through the action of the **spending multiplier**, which in turn is influenced by the marginal propensity to consume (MPC).
- A special case of discretionary fiscal policy is the balanced budget situation where any spending change is offset by a tax change of the same size, in which case the **balanced budget multiplier** is one and the impact on aggregate demand is equal to the change in government spending.
- **Automatic stabilisers** are built-in mechanisms that automatically help fight unemployment and inflation without there being any need for government to explicitly alter spending initiatives or tax laws.

2. Understand the concept and rationale of supply-side fiscal policy, and understand how it differs from demand-side fiscal policy

- **Supply-side fiscal policy** emphasises government policies that increase aggregate supply (rather than policies to change aggregate demand) in order to achieve long-run growth in real output and full employment.
- A **shift in aggregate supply** can be accomplished by some combination of: cuts in resource prices (e.g., by enacting various economic reform policies aimed at increasing the efficiency with which such resources are produced), policies enacted to encourage (the pace of) technological advances, increased investment allowances for companies and/or reductions in a range of company taxes, and reductions in government regulation (thereby reducing compliance costs).
- Beginning in the 1980s, Australia implemented quite a comprehensive set of **microeconomic reform policies**, which had the intent of increasing the productive efficiency of Australian industry – depicted in the *AD–AS* framework as a shifting out of aggregate supply – and which therefore have had a strong supply-side flavour to them, as distinct from demand-side fiscal policy.

3. Appreciate the macroeconomic significance of the federal budget

- The **federal budget** is the principal annual fiscal policy statement by the federal government.
- If government expenditure exceeds revenue the budget is in **deficit**; this has been the case in Australia since 2008–09 and is projected to remain the case until 2030–31. This has serious consequences for the level of **government debt** and many believe the budget should have been returned to surplus much earlier.
- In using the budget outcome to gauge the impact on the economy it is important to also look at the **domestic deficit/surplus**, the actual make-up of the proposed spending/taxation changes, as well as adjusting it for

the stage of the business cycle during the period of interest to better determine the **fiscal stance** of the government.

4. Develop an awareness of the implications of the federal budget outcome for government debt levels

- When the government runs a **budget deficit** it finances it by borrowing.
- Borrowing from the central bank is called **money financing** and borrowing from the public is called **debt financing**. Since the 1980s, Australian budget deficits have been financed by debt financing.
- Deficits increase the level of government or **national debt**.
- Debt that is owned by residents results in a redistribution of income over time within the country. It may amount to a burden on future generations if private sector investment is **crowded out** by the deficit spending and the government spending is for current consumption rather than for capital goods.
- Government borrowing offshore to fund its spending on current consumption is almost certain to lead to future difficulties in servicing the debt and paying it back.

Study questions and problems

1 How does each of the following affect the aggregate demand curve?
   - **a** The amount of taxes collected decreases.
   - **b** Government spending decreases.
   - **c** A so-called 'balanced-budget' increase in government spending whereby the amount of taxes collected also increases by the same amount as the government spending.

2 Explain the role that the spending multiplier plays in demand-side discretionary fiscal policy.

3 Which policy is more contractionary, a $100 billion cut in government spending for goods and services or a $100 billion increase in taxes? Explain your answer.

4 What is the difference between discretionary fiscal policy and automatic stabilisers? Give an example of each.

5 Consider an economy that is operating at well below the full-employment level of real GDP. Assuming the MPC is 0.75, predict the effect on aggregate demand and real GDP of a $300 billion increase in government spending balanced by a $300 billion increase in taxes. What would be your answer if the MPC had instead been 0.60?

6 Suppose federal parliament enacts a new tax law in a bid to increase tax revenue and the average income tax rate increases from 35 per cent to 50 per cent. Government tax revenue subsequently actually decreases from $800 billion to $600 billion. In terms of supply-side economics and the Laffer curve, explain this outcome.

7 Indicate the change in either the aggregate demand curve or the aggregate supply curve for each of the following:
   - **a** Expansionary fiscal policy.
   - **b** Contractionary fiscal policy.
   - **c** New government regulations that increase the costs of doing business.
   - **d** The 'corporatisation' of government-owned power companies, which increase their efficiency of operation, thereby reducing electricity prices for businesses.

8 Suppose the percentage of a country's national debt owned by foreigners increases sharply. Should this be of concern?

9 In June 2020 the Australian federal government had around $700 billion in outstanding Commonwealth government securities (its debt). Suppose the average interest rate it paid on that debt in 2020–21 was 3 per cent and that, during 2020–21, in addition to meeting its interest payments obligations, its other expenditure was $200 billion greater than revenue raised.

a What is its interest bill during 2020–21?
b What is the government's budget deficit in 2020–21?
c Other things equal, what then will be the size of the federal government debt by June 2020?
d And therefore what will be its interest bill in 2021–22 (assume the same average interest rate)?

10 'Intergenerational equity' is the idea of fairness between generations. Discuss this concept in the context of the most appropriate way for governments to finance (either through taxation or by borrowing) spending on:
a current consumption.
b public infrastructure.

Answers to 'You're the economist'

Calculating the size of the spending multiplier

For an MPC of 0.6 and $1 of initial spending, the total spending stimulus after three rounds of induced spending will be 1 + 0.6 + 0.36 + 0.216 = $2.176. For $15 billion then, the stimulus after three rounds of induced spending would therefore be $32.64 billion. The next three rounds of induced spending for an initial stimulus of $1 would amount to 0.1296 + 0.0778 + 0.0467 (i.e., 0.2541). Thus, the total stimulus from $15 billion initial stimulus after six rounds of induced spending would be $36.4515 billion. The next two rounds of induced spending would add only a further $0.672 billion spending stimulus, to give a total spending stimulus after eight rounds of induced spending of $37.12 billion. After another two rounds the total stimulus would be $37.3615 billion.

The total spending stimulus thus appears to be approaching $37.5 billion. Thus, for an initial spending stimulus of $15 billion, the total spending stimulus to the economy would be $37.5 billion. If you said the spending multiplier for an MPC of 0.6 is 2.5 (obtained as 1/1 − 0.6), YOU ARE THINKING LIKE AN ECONOMIST.

Calculating the required tax cut

To increase *AD* by $40 billion means that the initial spending stimulus plus all induced spending must equal $40 billion. This means the initial spending stimulus would need to be $10 billion if the MPC was 0.75. This follows because $40 billion = (initial spending stimulus)/(1 − 0.75). Now, given the MPC of 0.75, this means that 25 per cent (i.e., 0.25) of any income tax cut by the government will be saved by the recipients. Therefore, to get the initial spending stimulus of $10 billion, the government would need to cut taxes by $13.34 billion ($13.34 billion × 0.75 = $10 billion). If you said the tax cut needed would be $13.34 billion, YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 Discretionary expansionary fiscal policy is deliberate government action to influence aggregate demand and the level of real GDP through:
a increasing government spending or reducing taxes.
b discretely working with business to change their investment plans.
c encouraging foreign companies to invest in Australia.
d expanding the money supply.

2 If the spending multiplier is 4 the value of the marginal propensity to consume is:
a 0.25.
b 0.5
c 0.75.
d 0.4.

3 Assume the marginal propensity to consume (MPC) is 0.6 and the government increases taxes by $30 billion. The aggregate demand curve will shift:
- **a** inward and to the left by $18 billion.
- **b** inward and to the left by $45 billion.
- **c** outward and to the right by $18 billion.
- **d** outward and to the right by $45 billion.

4 Which of the following is/are an automatic stabiliser?
- **a** Defence spending.
- **b** Unemployment compensation benefits.
- **c** Wages of public servants.
- **d** Personal income tax.

5 Suppose that the economy is currently operating below the full employment level of GDP and the federal government wants to increase the level of aggregate demand in the economy next period by $150 billion (i.e., within the *AD–AS* framework it wants to horizontally shift out the aggregate demand curve by $150 billion), and it wants to do this by using expansionary fiscal policy. Assuming the marginal propensity to consume is 0.6, this increase in aggregate demand could be achieved by:
- **a** increasing government spending by $150 billion.
- **b** decreasing taxes by $150 billion.
- **c** increasing government spending by $60 billion.
- **d** increasing welfare payments by $60 billion.

6 If the federal government runs a budget _____, then the national debt becomes _____.
- **a** surplus; larger
- **b** deficit; larger
- **c** deficit; smaller
- **d** All of the above are incorrect.

7 Suppose during some period outstanding Australian Commonwealth government securities (CGS) on issue amounted to $550 billion and the RBA held $50 billion in CGS on its balance sheet, which of the following statements is/are true?
- **a** Total federal government debt during the period amounted to $600 billion.
- **b** Total federal government debt during the period amounted to $500 billion.
- **c** Total federal government debt during the period amounted to $550 billion, of which $50 billion had been monetised by the RBA carrying $50 billion on its balance sheet.
- **d** Total federal government debt during the period amounted to $600 billion, of which $50 billion had been monetised by the RBA carrying $50 billion on its balance sheet.

8 Which, if any, of the following statements about crowding out is/are false?
- **a** It is argued that it may occur when the government runs a budget deficit.
- **b** It never occurs because the government always finances its deficits by money financing.
- **c** It occurs when the stimulatory effect of deficit government spending is offset by reduced private sector spending.
- **d** It refers to impacts on interest rates and, in turn, consumption and investment spending.

9 The cyclically adjusted budget balance:
- **a** is of little relevance as it is the actual budget balance that is of most relevance as far as fiscal policy is concerned.
- **b** is always balanced out to zero because the effect of the economy's business cycle is taken out of the spending and revenues of the government.
- **c** is the value of the government's actual budget balance adjusted for the stage of the business cycle.
- **d** always makes actual budget surpluses and deficits larger in absolute terms.

10 Which, if any, of the following statements is/are true?

- **a** The domestic deficit is always greater in absolute terms than the actual budget deficit.
- **b** The domestic deficit is the result of subtracting net exports from the government's tax revenues.
- **c** The domestic deficit is the difference between a country's total export revenue and what it has to pay for its imports of goods and services.
- **d** The domestic deficit is the difference between the government's total revenues and its spending on domestically produced goods and services.

11 The portion of a country's national debt held by foreigners:

- **a** is no real problem because the country can simply refuse to pay the debt back.
- **b** may represent a burden because it transfers some future income from the country's citizens to other countries.
- **c** should be eliminated as soon as possible because it is not good for governments to borrow from foreigners.
- **d** is an accounting entry that represents no real burden.

PART 6

FURTHER ASPECTS OF INTERNATIONAL ECONOMICS

18 International trade and finance

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand why nations benefit from trade.
2 Understand what is meant by free trade, fair trade and protectionism.
3 Understand the concepts of a country's balance of payments and its foreign liabilities.
4 Understand the foreign exchange rate and its economic impact.

Just imagine life without world trade: no foreign-made CDs, DVDs, movies, beers, wine, coffee or tea, cars, motorcycles, televisions, computers, phones or iPads; and, of course, overseas travel would not be on the agenda – and the list goes on. So the point is clear; world trade is important because it increases world production and gives consumers more choice by providing a wider array of goods and services that become available from international trade. Today, the speed of transportation and communication means producers must compete on a global basis for the spending dollars of consumers.

Alamy Stock Photo/MG Images Studio

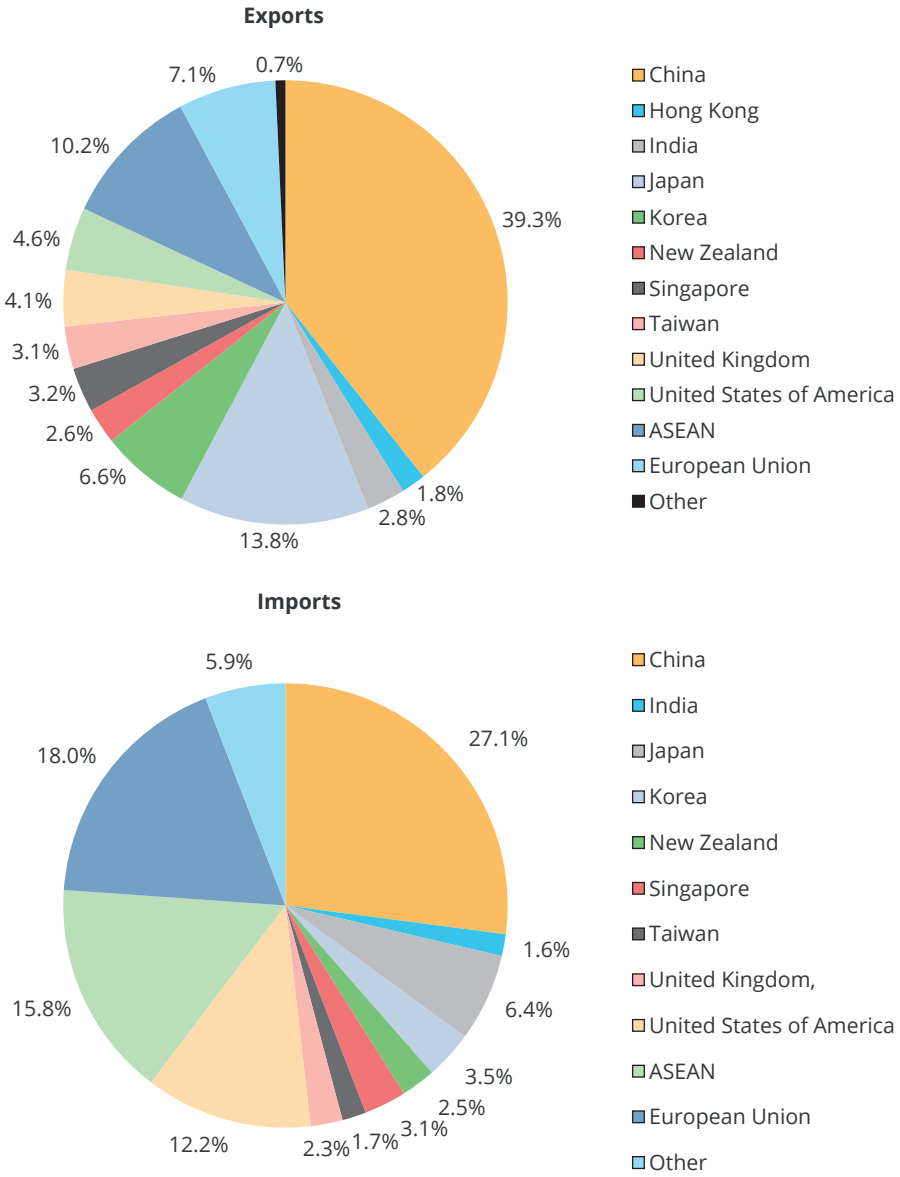
The first part of this chapter explains why countries should specialise in producing certain products and then trade them for imports. Then we will consider the issue of 'fair trade' and whether it should be preferable to 'free trade'. Is there a need for protection for Australian firms from the 'unfair' trade practices of other countries?

The way nations pay each other for world trade will be discussed, and issues relating to a country's current account balance and its levels of foreign debt will be introduced. We will also explain why the value of one Australian dollar can and does vary quite considerably against the Japanese yen, the US$, the euro, £ sterling, the NZ$, the Korean won, and so on.

1 Why nations benefit from trade

Exhibit 18.1 reveals which regions are Australia's major trading partners in merchandise trade (i.e., goods trade). In 2019–20 only about 18 per cent of Australia's merchandise exports went to the European Union (EU), the UK, the US and New Zealand, and about 35 per cent of its merchandise imports came from these areas. On the other hand, about 80 per cent of Australia's merchandise exports went to Asian countries,

Exhibit 18.1 **Australia's merchandise trade trading partners, 2019–20**



Source: Australian Bureau of Statistics, *International Trade in Goods and Services*, October 2020, Cat. No. 5368.0, Tables 14a and 14b.

In 2016–17, Asian countries accounted for about 80 per cent of Australian merchandise exports, while Europe, the UK, the US and New Zealand accounted for just 18 per cent. On the other hand, 59 per cent of Australia's merchandise imports were sourced from Asia, with 35 per cent coming from Europe, the UK, the US and New Zealand.

while about 59 per cent of its merchandise imports were sourced from the same Asian countries. Thus, these days, Australia's merchandise trade is pretty heavily focused on Asia. However, when exports and imports of services are taken into account, the relative importance of Europe, the UK and the US is increased considerably, particularly in respect of imports.

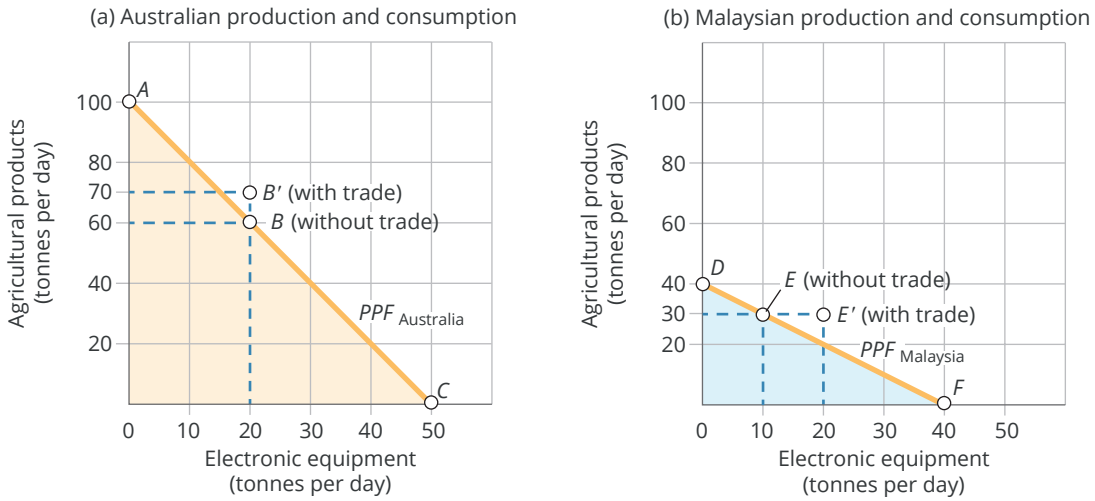
The composition of Australia's exports is also quite different from the composition of its imports. Around a half of the total value of its exports are commodities, while the great bulk of its imports are manufactured goods and services. Leading Australian exports are: commodities, such as coal, iron ore, bauxite, gold, wool, beef, cotton, sugar, cereal grains and other agricultural products; services, such as tourism, medical and health services and education; and manufactures, such as steel products and various other metal products, aluminium and a range of food-related products. Major imports include cars and trucks, heavy machinery, petroleum, electrical and electronic equipment, and clothing.

Why does a nation even bother to trade with the rest of the world? Does it seem strange for Australia to import goods it could probably produce for itself? Indeed, why doesn't Australia become self-sufficient by growing all of its food, making all of its cars, machinery and other manufactured goods, and prohibiting sales of all foreign products? This section explains why specialisation and trade are a nation's keys to unlocking a higher standard of living for its citizens.

The production possibilities frontier revisited

Consider a world with only two countries – Australia and Malaysia. To keep the illustration simple, also assume both countries produce only two generic 'goods' – agricultural products and electronic equipment. Accordingly, we can construct in **Exhibit 18.2** a *production possibilities frontier* for each country.

Exhibit 18.2 **The benefits of trade**



As shown in part (a), assume Australia chooses point *B* on its production possibilities frontier, $PPF_{Australia}$. Without trade, Australia produces and consumes 60 tonnes of agricultural products and 20 tonnes of electronic equipment. In part (b), assume Malaysia also operates along its production possibilities frontier, $PPF_{Malaysia}$, at point *E*. Without trade, Malaysia produces and consumes 30 tonnes of agricultural products and 10 tonnes of electronic equipment. Now assume Australia specialises in producing agricultural products at point *A* and imports 20 tonnes of Malaysian electronic equipment in exchange for 30 tonnes of agricultural products. Through specialisation and trade, Australia moves to consumption possibility point *B´*, outside its production possibilities frontier. Malaysia also specialises in producing electronic equipment (point *F*) and moves to a higher standard of living at consumption possibility point *E´*, outside its production possibilities frontier.

We will also – for the purposes of this discussion – set aside the *law of increasing opportunity costs* (explained in **Chapter 2**). This does not alter the essential validity of the discussion to follow but simplifies things by transforming the bowed-out shape of the production possibilities frontier of **Chapter 2** into a straight line.

By comparing parts (a) and (b) of **Exhibit 18.2**, we see that Australia can produce more agricultural products than Malaysia. If Australia devotes all of its resources of land, labour and capital to this purpose, 100 tonnes of agricultural products may be produced per day, represented by point *A* in part (a) of **Exhibit 18.2**. The maximum agricultural products production of Malaysia, on the other hand, is 40 tonnes per day because Malaysia has less labour, arable land and other factors of production than Australia. This capability is represented by point *D* in part (b) of **Exhibit 18.2**.

Now consider the capacities of the two countries for the production of electronic equipment. If all their respective resources are devoted to this output, Australia produces 50 tonnes per day (point *C*) and Malaysia produces 40 tonnes per day (point *F*). Again, the greater potential maximum electronic equipment output of Australia reflects its greater resources.

However, you should notice in passing that Australia is relatively less efficient at producing electronic equipment than agricultural products – using all of its resources, it can only produce half as much electronic equipment as it can produce agricultural products. Malaysia, on the other hand, is equally efficient at producing both types of goods.

Specialisation without trade

Assuming no world trade, the production possibilities frontier for each country also defines its *consumption possibilities*. Suppose Australia prefers to produce and consume 60 tonnes of agricultural products and 20 tonnes of electronic equipment per day (point *B*). Also assume that Malaysia chooses to produce and consume 30 tonnes of agricultural products and 10 tonnes of electronic equipment (point *E*).

Exhibit 18.3 lists data corresponding to points *B* and *E* and shows that the total two-country world output is 90 tonnes of agricultural products and 30 tonnes of electronic equipment.

Exhibit 18.3 **Effect of specialisation on world output**

	Agricultural products production (tonnes per day)	Electronic equipment production (tonnes per day)
Before specialisation		
Australia (at point *B*)	60	20
Malaysia (at point *E*)	30	10
Total world output	90	30
After specialisation		
Australia (at point *A*)	100	0
Malaysia (at point *F*)	0	40
Total world output	100	40

Now suppose Australia specialises by producing and consuming at point *A*, rather than point *B*. Suppose also that Malaysia specialises by producing and consuming at point *F*, rather than point *E*. As shown in **Exhibit 18.3**, specialisation by each country increases total world output per day by 10 tonnes of agricultural products and 10 tonnes of electronic equipment.

However, without trade, A and F are not desirable production points for each country since neither country wants to consume these combinations of output. Australia prefers to consume fewer agricultural products and some electronic equipment at point *B* compared to point *A*. Malaysia, on the other hand, prefers to consume some agricultural products and less electronic equipment at point *E*, rather than point *F*.

Specialisation with trade

Now, returning to **Exhibit 18.2**, suppose Australia agrees to specialise in agricultural products production at point *A* and to import 20 tonnes of Malaysian electronic equipment in exchange for 30 tonnes of its agricultural products output. Does Australia gain from trade?

The answer is yes. At point *A*, Australia produces 100 tonnes of agricultural products per day. Subtracting the 30 tonnes traded to Malaysia leaves Australia with 70 tonnes for its own consumption. In return for the agricultural products, Malaysia trades 20 tonnes of electronic equipment to Australia. Hence, specialisation and trade allow Australia to move from point *A* to point *B'*, which is a consumption possibility outside its production possibilities frontier in part (a) of **Exhibit 18.2**. At point *B'*, Australia consumes the same amount of electronic equipment and 10 more tonnes of agricultural products compared to point *B* (without trade).

Malaysia also has an incentive to specialise by moving its production mix from point *E* to point *F*. With trade, Malaysia's consumption would be at point *E'*. At point *E'*, Malaysia has as much agricultural production to consume as it had at point *E*, plus 10 more tonnes of electronic equipment. After trading 20 tonnes of the 40 tonnes of electronic equipment produced at point *F* for agricultural products, Malaysia can still consume 20 tonnes of electronic equipment from its production, rather than only 10 tonnes of electronic equipment at point *E*. Thus, point *E'* is a consumption possibility that lies *outside* Malaysia's production possibilities frontier also.

Of course, this highly simplified and stylised example is purely for expositional purposes and clearly is not meant to capture the much more elaborate production, trade and consumption patterns between Australia and Malaysia in the real world. Nonetheless, it should be noted in passing that the example does contain a kernel of the reality of the actual trading situation between the two countries. Australia is one of Malaysia's largest export destinations, with products heavily concentrated in information and communications technology manufactures, as well as crude oil and gas. On the other hand, Australia's exports to Malaysia are predominantly agricultural products, simply transformed mineral-based manufactures and education-based services.[1]

Comparative and absolute advantage

Here we introduce the economic principle that drives international specialisation and trade. It is one of the most fundamental and important principles of economics: the principle of comparative advantage.

Comparative advantage

Comparative advantage The ability of a country to produce a product at a lower opportunity cost than another country.

Engaging in world trade permits countries to escape the consumption prison of their own production possibilities frontiers by producing products in which they have a **comparative advantage** while using that production to trade with other countries to acquire other goods and services they also wish to consume. Comparative advantage is the ability of a country to produce a product at a lower opportunity cost than another country.

Returning to our earlier example, if Australia produces 50 tonnes of electronic equipment, 100 tonnes of agricultural products are not produced. This results in 2 tonnes of agricultural products being forgone for every 1 tonne of electronic equipment.

For Malaysia, the opportunity cost of producing 40 tonnes of electronic equipment is 40 tonnes of forgone agricultural products; that is, only 1 tonne of agricultural products is forgone for every 1 tonne of electronic

[1] 'Economic synergies between Australia and Malaysia', by Nicholas Coppel, DFAT, a paper presented at the Australia Malaysia Free Trade Agreement Conference, Melbourne, March 2005.

equipment. Malaysia's electronic equipment is therefore relatively cheaper than Australia's in terms of agricultural products forgone. This means Malaysia has a comparative advantage over Australia in electronic equipment production because less agricultural production has to be given up to produce electronic equipment in Malaysia than is required in Australia.

Conversely, for Australia, 1 tonne of agricultural products costs half a tonne of forgone electronic equipment. Since in Malaysia, 1 tonne of agricultural products costs 1 tonne of forgone electronic equipment, Australia has a comparative advantage over Malaysia in agricultural products because its opportunity cost in terms of forgone electronic equipment is lower compared to Malaysia.

Thus, it is in both their interests for Australia to specialise in agricultural products because it is relatively more efficient in their production, and for Malaysia to specialise in producing electronic equipment and then for the two countries to engage in trade.

Absolute advantage

So far, it has been shown that a country's production and international trade decisions depend on comparing what a country gives up to produce more of any given product. It is important to note that comparative advantage is based on opportunity costs, regardless of the absolute costs of resources used in production.

In the above analysis, we did not consider how much labour, land or capital either Australia or Malaysia uses to produce a tonne of agricultural products or electronic equipment. **Absolute advantage** is the ability of a country to produce a product using fewer resources than another country.

Absolute advantage The ability of a country to produce a product using fewer resources than another country.

What if Australia or Malaysia had an absolute advantage in producing both agricultural products and electronic equipment, meaning that country needed to use fewer resources per tonne to produce both agricultural products and electronic equipment than the other? Can it nonetheless still gain from specialisation and trade? The answer is yes and we explain why next.

Comparative and absolute advantage compared

Recall the example from **Chapter 2** where it was pointed out that a successful fashion designer may well be able to undertake more designing and more fabric-cutting in an hour than her cutter. However, while she may well have been quite an efficient fabric cutter, her design skill was truly exceptional and she was much better-off specialising in the design work and employing a fabric-cutter for that role in the business.

Consider another example of a lawyer and her typing needs. Suppose the lawyer has truly exceptional legal skills and can charge say $400 per hour for providing those skills. She also needs a lot of word-processing support and it turns out she is also quite adept at this as well and has a typing speed in the top 5 per cent.

The market rate for a competent typist is $40 per hour. The lawyer knows that, if she did them herself, she could get a set of briefs typed up in two hours, which would take such a typist four hours to do. Should she therefore do all her own typing since she has an absolute advantage in both the provision of quality legal advice as well as typing? No, of course not. The opportunity cost of the two hours she would spend on the typing is $800 whereas she could get the same result from paying someone else $160 (4 x $40) for this support. While she has an absolute advantage in both skills, her comparative advantage is clearly in the provision of legal advice.

Do nations with an absolute advantage gain from trade?

You're the economist

Suppose for the sake of argument that Australia has an absolute advantage over Indonesia in the production of both calculators and towels. In Australia, for a given AUD cost of production inputs, four calculators or 400 towels can be produced. In Indonesia, for the same AUD-equivalent cost of production inputs, one calculator or 50 towels can be produced. Under these conditions, would Australia benefit economically by trading with Indonesia?

The same argument explains why most of us specialise in some particular thing as our livelihood rather than becoming 'jacks of all trades'. By specialising, and then trading our output, we are able to expand our consumption possibilities. In theory, a person could possibly grow all their food, make their clothes, educate their children, build their own house, do all their own mechanical repairs, and so on. However, individuals typically choose to specialise in a small number of things in which they have a comparative advantage compared with others in their communities. They then trade their production (through spending their earned income) for that of others who have similarly specialised in other activities. In this way, the whole community achieves higher levels of production, consumption and economic welfare than would otherwise have been the case had individual members tried to do everything for themselves.

What makes obvious sense for the individual and domestic communities is equally sensible for entire countries. The idea that nations concentrate on production in their areas of comparative advantage was first articulated by one of the most famous economists of the 19th century, David Ricardo (see **Exhibit 18.4**). Over the last 50 years or so, quite a number of empirical tests of the idea have been carried out. One of the most recent was a 2012 study by National Bureau of Economic Research (NBER) economists Costinot and Donaldson who claim to have found strong empirical support. As they said:

> Despite all of the real-world considerations from which this theory abstracts, we find that Ricardo's theory of comparative advantage has significant explanatory power in the data, at least within the scope of our analysis.[2]

Exhibit 18.4 **19th century economist David Ricardo**

Alamy Stock Photo/Classic Image

In summary

- *International specialisation* allows the potential for greater total world consumption.
- World output and consumption are maximised when each country specialises in producing goods for which it has a comparative advantage (i.e., a relatively lower opportunity cost of production) and trading them for those in which it does not have a comparative advantage.

[2] A. Costinot and D. Donaldson, 'Ricardo's theory of comparative advantage: old idea, new evidence', *American Economic Review*, Vol. 102, No. 3, 2012, pp. 453–8. First published as NBER Working Paper No. 17969, 2012. Quote in the text is from the Abstract to the Working Paper.

2 Free trade, fair trade and protectionism

Free trade is the flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow. In practice, every nation protects its own domestic producers from foreign competition to some degree. **Protectionism** is the government's use of embargoes, tariffs, quotas and other restrictions to protect domestic producers from foreign competition.

Free trade
The flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow.

Protectionism
The government's use of embargoes, tariffs, quotas and other restrictions to protect domestic producers from foreign competition.

Embargoes, tariffs and quotas

Embargoes are the strongest limit on trade. An embargo is a law that bars trade in a range of products with another country. In a recent example, the United Nations authorised trade and financial sanctions (like embargoes) against North Korea in response to its continued breaches of international law in developing its nuclear weapons and intercontinental delivery capabilities.

A **tariff** – also called a customs duty – is a tax levied by the government on an import, which increases its domestic sale price compared with similar domestic goods and thereby potentially reduces demand for the imported product. The average Australian tariff is now less than 5 per cent (as a percentage of import prices). Motor vehicle and textile production in Australia, which had been historically afforded very high tariff protection, was progressively wound back over recent decades.

In response to the onset of the worldwide Great Depression in 1929–30, some nations decided to try to protect their industries and workers from foreign competition by erecting high tariffs on imported goods. It was thought that imports meant the loss of domestic jobs. Unfortunately, as one nation raised its tariffs to protect its industries and jobs, other nations retaliated by raising their tariffs.

In the end, all that these so-called 'beggar thy neighbour' policies achieved was a worldwide reduction in demand and production, thereby actually bringing about the higher unemployment such ill-conceived policies were aimed at avoiding. Very fortunately, after the onset of the 2008–09 GFEC, the world's leading nations demonstrated they had learnt the lessons of the 1930s and resisted pressure to impose such trade barriers. However, the COVID-19 pandemic affected international trade and travel negatively, leading to disruptions in the supply chain, border closures, mobility restrictions, lockdowns, closures of travel, tourism and entertainment industries and so on. However, countries are expected not to escalate the ongoing trade tensions further, not to impose new tariffs or trade restrictions, and increase the commitment to trade without violating rules.

Embargo
A law that bars trade with another country in a range of products.

Tariff
A tax on an import that increases the domestic price of the imported product.

Most countries today are members of the World Trade Organization (WTO) and they meet periodically to negotiate multilateral reductions in tariff rates around the world with the aim of stimulating world trade and output. In November 2001 at Doha, Qatar, the trade ministers of the 144 WTO member nations agreed to initiate a new trade negotiation round (the ninth since the Second World War). The Doha Round had been scheduled to take three years, but, unfortunately, by the end of 2015, progress had effectively stalled.

Notwithstanding this, it is nonetheless hoped that reductions in international trade barriers will ultimately be achieved, either through some sort of WTO-sponsored multilateral approach or through a range of emerging bilateral and regional agreements between various pairs of countries and within defined regions around the world. Such reductions in protection would further enrich the world through greater trade, economic activity and employment opportunities, particularly for those suffering severe economic hardships in less-developed countries.

Another way to limit foreign competition is to impose a **quota**. A quota is a limit on the quantity of a good that may be imported into a country in a given time period. For example, a country may allow 10 million tonnes of sugar to be imported over a one-year period. Once this quantity is reached, no more sugar can be imported for the year.

Quota
A limit on the quantity of a good that may be imported in a given time period.

Quotas, like all barriers to trade, invite other nations to retaliate with more measures to restrict trade, the end result being that all nations use their scarce resources in sub-optimal ways and thereby become worse off than they otherwise could have been.

Analyse the issue — Applicable concept: the cost of tariff protection

The cost of Australia's tariff protection

A Productivity Commission report entitled 'Trade and Assistance Review, 2017–18' and 'Trade and Assistance Review, 2018–19', published in July 2019 and 2020 respectively, contains the Commission's latest quantitative estimates of Australian government assistance to industry.

Exhibit 18.5 has been extracted from the Commission's report and provides what the Commission refers to as 'Net Tariff Assistance' to the major sectors of the Australian economy. Aggregate Net Tariff Assistance provided to Australian producers in 2017–18 and 2018–19 was estimated to be just $0.26 billion. In aggregate this is small, but what is most important is how this net tariff assistance is distributed across the major sectors of the economy.

Exhibit 18.5 **Net tariff assistance by industry sector, 2012–13 to 2018–19, $million (nominal)**

	2012–13	2013–14	2014–15	2015–16	2016–17	2017–18	2018–19
Primary production	272.5	157.1	195.5	239.3	338.4	355.8	186.5
Mining	-151.3	-165.2	-164.8	-148.5	-114.9	-102.1	-55.4
Manufacturing	2603.9	2280.7	2180.4	1839.9	1270.6	1169.0	1162.3
Services	-2012.6	-2082.3	-2006.6	-1663.6	-1217.9	-1158.4	-1028.6
Total	**712.5**	**190.3**	**204.4**	**267.1**	**276.2**	**264.3**	**264.8**

Source: Productivity Commission, 'Trade & Assistance Review 2017–18', July 2019, Chapter 2, Table 2.2. and 'Trade & Assistance Review 2018–19', July 2020, Chapter 1, Table 1.1.

What do you think?

1 Using the data for 2018–19, what was the estimated net tariff assistance provided to the manufacturing sector that year?
2 What was the estimated size of tariff-imposed input-cost penalties – i.e., negative net assistance – borne by the services sector? (This would have been ultimately borne by consumers of services who had to pay that much more for those services than would have been the case in the absence of the tariff protection provided to manufacturing.)

In addition to embargoes, tariffs and quotas, some nations use more subtle measures to discourage trade. For example, some countries set up an overwhelming number of bureaucratic steps that must be taken in order to import a product. Some countries would, for example, claim Australia has used quarantine and health regulations to restrict – and, in their view, unnecessarily – the importation of various agricultural products from other nations.[3]

[3] For example, Australia levied a 90-year embargo (from 1921 until 2011) on importing apples from New Zealand, citing fear of spreading 'fire blight' to Australia as the reason. Australia has also placed a continuing embargo on importing fresh bananas – from countries like the Philippines – again citing the ongoing need to protect Australia from diseases of bananas, such as Moko, Black Sigatoka and Freckle.

Fair trade, the 'Fair Trade Movement' and strategic trade

Free trade was defined above as the free flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow. In recent years, much has been made of the notion of fair trade. The term has a nice ring to it, so it is important to understand what exactly is meant by it, in what sense the word 'fair' is used, and whether it is or isn't desirable.

Fair trade is the term used to represent the view that a country should only reduce its barriers to imports from another country if the other country does not have some sort of 'unfair' competitive advantage over it and the other country is also willing to reduce its import barriers reciprocally. The notion of fair trade has popular appeal, so it is important to spend a little time thinking critically about it.

Fair trade
Term used to represent the view that a country should only reduce its barriers to imports from another country if the other country does not have some sort of 'unfair' competitive advantage over it and the other country is also willing to reduce its import barriers reciprocally.

First and foremost, how should one define 'unfair' competitive advantage? For example, is it unfair if another country has a lot more rainfall and easily accessible fertile and arable land to use in its production of agricultural products than does the home country? If this means the foreign country can produce a range of such products at 50 per cent of the cost of the home country, is that a reason for the home country to claim 'unfair competition', impose high tariffs on agricultural imports from it and thereby deny itself this source of high-quality, relatively cheap food products?

Another common example of supposed 'unfair' advantages is differential wage rates between developing and developed nations. (For more on this, see the discussion later in the chapter titled 'Cheap foreign labour argument'.

Second, you also need to ask who is being hurt when a country imposes high tariffs on the importation of another country's products. The foreign country is most likely being hurt because it probably can't sell as much as it could into the home country's markets. But, very importantly, the home country is also being hurt by its actions. The tariffs will result in people having to pay higher prices for such products than they need to and too much of the home country's scarce resources will be used up in inefficient production rather than being freed up and redirected into more productive areas. As a well-known Australian economist once remarked, imposing tariffs in such situations is like shooting yourself in the foot with a shotgun in the hope that a few of the pellets will also hit your competitors!

The moral that should be drawn and remembered here is that the prime motivation for a country to lower its tariffs is not to give a helping hand to firms in other countries. The most important reason – and the prime motivation – for reducing import barriers at home is that the home country will itself derive very significant benefits from such action. Furthermore, as has already been explained, this will happen irrespective of whether another country lowers its tariffs in return for such action.

Despite the aforementioned discussion, there is nonetheless a growing level of concern in the developed world about issues such as international differences in workplace health and safety regulation, the use of child labour in some countries, and the lack in many developing countries of appropriate environmental regulation.[4] Such concerns have driven increasing support around the world for the so-called 'Fair Trade Movement'.

As part of the fair-trade movement, in the late 1980s a labelling system was created to help introduce 'certified' products into mainstream international product markets. As one example, 'fair trade certification' of small farms now occurs in several developing countries to ensure certified agricultural products in those

[4] For example, in April 2013, the Rana Plaza garment factory in Dhaka, Bangladesh collapsed, killing many textile workers and injuring many others. Subsequently over 150 international clothing companies have signed an 'accord' that allows staff to stop work if their safety is under threat. Similarly, some environmentalist commentators would probably question why Australia recently signed a Free Trade Agreement with China (see later for further discussion on this) and instead ask why Australia does not impose a tariff on imports of manufactured goods from China, the major global contributor to carbon emissions and many other forms of pollution.

countries are produced by small farms that are operated following basic environmental, health, safety, labour and human rights regulations.

Usually, the prices of such 'fair trade certified' products are significantly higher than similar non-certified products. However, a growing number of consumers in developed countries around the world are happy to pay the higher prices as they have a strong desire to support the movement, provided, of course, they can be assured of the proper policing of the certification process.

Strategic trade theory
The idea that governments should seek to be strategic in their use of their spending and taxing powers in order to actively facilitate some sectors of the economy that they may feel have the potential to be strong export earners for their nations.

Strategic trade theory is the term used for the idea that governments should seek to be strategic in the use of their spending and taxing powers in order to actively facilitate economic growth in some sectors of the economy that they may feel have the potential to be strong export earners for their nations.

Proponents argue that a country's comparative advantages are not static. On the contrary, they would argue that nations – such as Singapore, for example – can develop comparative advantages over other nations by careful and selective support for some domestic industries over others. This idea is not without its critics, who argue that the private sector is better placed, and better skilled, to develop export opportunities. For example, another very prosperous south-east Asian nation, Hong Kong, followed a market-based approach to developing its export markets with little strategic direction from the government. Critics also argue that there are many other counter-examples to the case of Singapore that demonstrate how wasteful and counterproductive such interventionist government policies can be when they go wrong.

Be that as it may, on one thing there is agreement: it is undoubtedly true that comparative advantages are dynamic. Just as a person's comparative advantages can – and usually do – evolve over time, so too do those of an entire country. Thus, rather than remaining static, a nation's comparative advantages evolve over time as technology and international tastes change. What is not agreed upon is the appropriate role for government to play in such circumstances.

Global perspective — Applicable concept: comparative advantage

The 'Dutch Disease'

The **Dutch Disease** refers to the possible relationship between a significant and sudden increase in natural resource exports and a decline in the manufacturing sector. The proposed mechanism is that an increase in revenues from natural resource exports will appreciate a nation's currency, resulting in its other exports – particularly exports of its manufacturing sector – becoming less profitable as well as potentially hurting all those import-competing businesses in the economy. The term was coined in the late 1970s to describe the decline of the manufacturing sector in the Netherlands after the discovery of a large natural gas field in 1959. However, there is disagreement among economists about the extent to which the 'Disease' is something about which policymakers need to worry. Being a resources rich country, Australia has experienced resource booms several times in history. The resources and mining boom in 2000s brought economic prosperity to the whole economy but some sectors in Agriculture, such as grains and wool were damaged.[1]

As we learnt earlier, a country ought to specialise in industries in which it has a comparative advantage, so a country rich in natural resources could be expected to be better off specialising in the extraction of those natural resources. However, this may certainly result in short-term transitional impacts as some of the nation's production capacity is redirected into the natural resources sector, and these short-term reallocation adjustment effects, along with the impact of a possibly higher-valued exchange rate on the nation's other production sectors, need to be managed properly by policymakers.

One way to reduce the threat of Dutch Disease is to seek to boost the competitiveness of the nation's other production

Getty Images/Bloomberg

sectors, including the manufacturing sector. Setting appropriate government royalties accruing from the extraction of the natural resources, and then investing them in education, research, innovation and public infrastructure, would be one way to increase the competitiveness of the rest of the economy, including the manufacturing sector.

What a government should not do is to resort to protection of its affected manufacturing industries by, for example, using tariffs. Imposing tariffs on imported goods will act to reduce demand for imported products and most likely simply lead to further appreciation of the exchange rate. Similarly, a government should not try to 'pick winners' and use the resource royalties to give specific subsidies to the production of those selected industries – better to improve the general international competitiveness of the economy, as just described, and then allow the 'winners' to self-select.

What do you think?

Analyse this issue from the point of view that nations should specialise in those activities in which they have a comparative advantage. What role should government play in mitigating any negative impacts that may flow from a country being in possession of such internationally sought-after natural resources?

1 C. Ebrahim-Zadeh, 'Back to basics – Dutch Disease: too much wealth managed unwisely', *Finance and Development*, IMF, Vol. 40, No. 1, March 2003; 'The Dutch Disease', *The Economist*, 26 November 1977, pp. 82–3; 'Strong forex inflows now hurting economy', GMANews.TV, http:// www.gmanews.tv/story/69158/Strong-forex-inflows-now-hurting-economy–BSP, accessed 10 May 2015; Battellino, R. (2010), Mining Booms and the Australian Economy. Address to The Sydney Institute, Sydney, https://www.rba.gov.au/speeches/2010/sp-dg-230210.htm.

The political economy of reducing barriers to trade

Dutch Disease
A term referring to the possible relationship between a significant increase in natural resource exports and a consequential decline in the manufacturing sector. The proposed mechanism is the increase in natural resource export revenues appreciating the currency, resulting in other exports becoming less profitable, and import-competing businesses less able to compete with imports made cheaper by the appreciated currency.

If the trade is so obviously beneficial to the economic prosperity of countries, why is free trade vehemently opposed by some and why do some countries severely restrict trade? The reason essentially stems from the political influence of narrow special-interest groups that are likely to be negatively affected by any reductions in a country's trade restrictions.

Free trade provides consumers with lower prices and more products from which to choose. Thus, reducing tariffs might take a few hundred dollars a year off the cost of living for an individual family. When added up across the entire community, these savings are considerable, but for each family, it might not be a huge amount and they are unlikely to link their reduced cost of living directly back to the tariff reductions anyway.

The resultant extra buying power of consumers is likely also to stimulate other areas of domestic production, with concomitant increases in employment. However, again, the firms who experience the extra demand would not necessarily link back to this being a benefit of the reduced tariffs and neither would the people who receive the additional jobs that those firms are then able to offer.

Finally, reduced import prices arising from the lower tariffs will generally lower the costs of production of firms since many firms use imported components and machinery in their production. This will be of great benefit to exporting firms, in particular, allowing them to sell more output into international markets, earn higher income for their owners and provide more significant numbers of jobs. However, again, such firms and employees are unlikely to appreciate the link between the reduced tariffs and their increased good fortunes and extra opportunities.

On the other hand, the cheaper imports resulting from the reduced tariffs may significantly reduce the incomes of the owners of the firms competing directly with the cheaper imports. Furthermore, some employees in such firms may lose their jobs, resulting in considerable short-term hardship for them.

Thus, the benefits of the tariff reduction are dispersed across a wide cross-section of the community and are enjoyed by many who do not necessarily link back to the tariff reductions. However, the costs of the tariff reductions are felt by a small, easily identified group who are usually very much aware of the reason for their reduced fortunes.

Thus, despite the great total benefits to the wider community, trade barriers continue to exist in most countries. The employees and owners of import-competing firms have a great deal at stake, and their plight often makes for good sensationalist news and current affairs shows. They are therefore also likely to find a sympathetic ear in politicians – particularly those in regions most negatively affected by the tariff reductions – who have at least one eye fixed firmly on the next election.

Some common (but specious) arguments for protection

The following are some of the most popular rationalisations used by those seeking protection. These arguments have strong political or emotional appeal, but weak support from most economists.

Infant industry argument

The infant industry argument suggests that a new domestic industry needs protection because it is not yet ready to compete with established foreign competitors. With time to grow and with protection, an infant industry can 'catch up' with established foreign firms.

Economists ask where one draws the arbitrary line between an 'infant' and a 'grown-up' industry. It is also challenging to make a convincing case for protecting an infant industry in a developed country where industries that are consistent with the country's natural comparative advantage should be well established.

The infant industry argument may have some validity for less-developed countries. However, even for these countries, there is a danger. Once protection is granted, the new industry will not necessarily feel the competitive pressures necessary to encourage the adoption of the most efficient production technologies and best management and work practices. Also, once an industry is given protection, it can prove extremely difficult to take away that protection for all the reasons already mentioned.

National security argument

By protecting critical defence industries, a nation will not be dependent on foreign countries for the essential defence-related goods it needs to defend itself in wartime. For instance, most recently in Australia, there was a considerable debate – with some of it resting on the national security argument – about whether a new federal defence contract for submarine construction should go to local Australian companies or whether better value could be had by going offshore.

Of course, a country must make a sensible investment in the necessary military assets and personnel to provide for its own security. However, again, rather than trying to be self-sufficient in the actual production of all such assets, the much more sensible approach is to purchase through trade those military assets the production of which the country does not have a comparative advantage.

Employment argument

The employment argument suggests that restricting imports increases domestic jobs in the protected industries. Protectionism may indeed increase output and save jobs in some domestic industries. What is ignored, however, are all the other employment-reducing effects that were covered earlier. The net outcome could well be that the tariffs act to reduce the total number of domestic jobs available compared with what would have been the case in the absence of the tariffs.

For example, suppose higher tariff protection is given to domestically produced motor vehicles, and this results in them being, say, $5000 per unit dearer in Australia than they would be otherwise. This may well protect a few hundred jobs in the domestic motor vehicle production industry.[5]

However, consumers will have that much less to spend on other products, possibly costing jobs in those firms whose demand is reduced on account of the reduced consumer buying power. The higher-cost vehicles also increase the production costs of all firms that use the vehicles in some way in the production of their products, again possibly costing jobs in those firms, particularly those export-oriented firms that need to compete in international markets.

In short, while protection for one industry may well protect some jobs in that industry, it may cause a net reduction in the nation's total employment, and most thorough, economy-wide analyses of the output and employment impacts of protection find just this outcome.

Cheap foreign labour argument

If labour costs a minimum of around $17 an hour in Australia and firms in many developing countries pay only $1 to $2 an hour, then without protection from imports from those countries, how can Australian industry possibly compete? This argument neglects the essential reason for the difference in the wage rates between countries.

Australian workers have more education, training, capital and access to more advanced technology than do workers in less developed countries. They, therefore, have a higher rate of labour productivity. Thus, if Australian workers produce more output per hour than workers in another country, then, even with the higher wages per worker in Australia, the labour costs per unit of production should remain internationally competitive.

Free trade (or preferential trade) agreements

Given the lack of success – despite more than 14 years of negotiations – in the WTO Doha round of multilateral trade talks, a growing trend has been for nations to negotiate reductions in trade barriers on bilateral and regional bases. A **free trade agreement (FTA)** is an agreement whereby participating members eliminate all (or most) trade barriers among themselves.

Free trade agreement (FTA) Parties to the agreement agree to eliminate or significantly reduce tariffs and other trade barriers among themselves. However, each individual country maintains its own tariff policy against non-parties to the agreement.

For instance, Australia has several FTAs in place. Existing signed FTA agreements and other agreements currently under negotiation account for a total of around 75 per cent of Australia's total international trade.

Other FTAs around the world include:

- In Europe, 15 nations joined together in 1958 to form a customs union called the European Economic Community (EEC), which, over time, evolved into the EU, now consisting of 27 countries.[6] The EU has removed all trade barriers among its members virtually, thereby creating, by the 1990s, a single European member-state economy somewhat comparable in size to the US economy.
- In addition, a new currency, the euro, was launched on 1 January 1999, has now replaced the individual currencies of a 19-country subset of the EU countries (originally there were 11 countries). The 19 euro-currency countries, accounting for around 320 million people, are all subject to the same monetary policy controlled by the European Central Bank (the ECB).
- In 1993, the US entered into an FTA with Canada and Mexico called the North American Free Trade Agreement (NAFTA). Under NAFTA, tariffs and other impediments to trade among the three nations

[5] Until the 1970s, Australia had one of the most protected manufacturing sectors in the Organisation for Economic Co-operation and Development (OECD). Unilateral tariff reductions by Australia resulted in protection falling to under 5 per cent on average by 2001. The tariff on imported motor vehicles was reduced to 5 per cent in January 2010.

[6] In a 'customs union', the parties to the union agree to eliminate all tariffs among each other and erect a common tariff structure for those countries outside of the union. The UK voted in 2016 to leave the EU and left the EU on the 31 January 2020 with a transition of 11 months.

were to be progressively phased out over 15 years.[7] In late 2014, the US had 14 FTAs covering 20 countries, including one with Australia, and several others under discussion, although Singapore probably holds the world record with 25 registered FTAs (both bilateral and regional).

Critics of bilateral and/or regional FTAs fear they may largely result in trade diversion rather than net trade creation, and that they will make global, multilateral agreements increasingly challenging to achieve in the future. It is undoubtedly true that as FTAs proliferate, these may increasingly act as a barrier to trade for those countries not included and, *ceteris paribus*, this may have quite a deleterious impact on the international trade and economic growth of such non-participants. Some believe the name itself is a misnomer, and that a more accurate name would be preferential trade agreement (PTA), since such agreements only free up trade between the parties to the agreement while other excluded countries can be significantly negatively affected.

Economics and ethics — Importance of free trade

Despite the incredible income and wealth available in developed countries at the start of the 21st century, very significant poverty continues to exist in quite a large number of less-developed countries, particularly in Africa. This poverty continues to plague generation after generation in such countries, with the result that the quality of life is extremely poor, child mortality is shamefully high and life expectancy is low.

Each year, the developed countries of the world dutifully contribute aid to such countries – usually around 1 per cent or so of their annual GDP. However, the aid never seems to permanently improve an economic lot of the people in the recipient countries, and the poverty cycle just goes on. What else can be done?

Several commentators are increasingly arguing that freer international trade, particularly in agricultural products – in tandem with better internal governance and anti-corruption measures in some of the countries concerned – may be the best way forward. It is, unfortunately, confirmed that many richer developed countries, despite their wealth, nonetheless continue to maintain significant trade barriers, such as import tariffs, quotas and embargoes, and inappropriate quarantine laws, as well as various forms of government subsidies to their exporters and so on.

The existence of these trade barriers means that developing countries may have little opportunity to use their comparative advantage – including relatively low labour costs – to trade their way to a better life. Thus, as is often the case, allowing markets to operate without undue government interference can well provide a superior outcome from many perspectives, including ethical concerns.

It is, therefore, most unfortunate for such poor countries that the 2001 WTO Doha Development Agenda Round was effectively abandoned in 2015. The Doha Round had great potential to help such countries help themselves through using their comparative advantage to trade their way out of poverty.

In summary

- **Free trade** is the flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow.
- **Embargoes**, **tariffs**, **quotas** and other restrictions are mechanisms to protect domestic producers from foreign competition.
- Arguments put forward against reducing protection include: the so-called infant industry argument, the national security argument, the employment argument and the cheap foreign labour argument.
- **Free trade agreements (FTAs)** involve countries agreeing to eliminate or significantly reduce tariffs and other trade barriers among themselves.

[7] US President Donald Trump announced in 2017 that the US would be evaluating NAFTA with a possible outcome being it could exit the arrangement unless 'a better deal' could be found for the US.

3 The balance of payments

The **balance of payments** is a bookkeeping record of the international transactions between a country and other countries during a given period of time. This summary records the value of a nation's spending inflows and outflows made by individuals, firms and governments.

Balance of payments A summary record of the international transactions between a country and other countries during a given period of time.

Exhibit 18.6 is a simplified balance of payments for Australia for 2019–20. Note the minuses in the table. A transaction that results in a dollar flow to Australia is entered as a positive amount, while payment by Australia, resulting in a dollar outflow to another country is entered with a minus sign.

Exhibit 18.6 **Australia's balance of payments, 2019–20**

Current account transactions	Amount (billions of dollars)
Goods	
1 Exports (credits)	382.632
2 Imports (debits)	–312.097
3 Balance on merchandise trade (1 + 2)	70.535
Services	
4 Exports (credits)	94.016
5 Imports (debits)	–87.138
6 Net services (4 + 5)	–6.878
7 Balance on goods and services (3 + 6)	77.413
Foreign income	
8 Credits	69.018
9 Debits	–109.181
10 Net income paid overseas (8 + 9)	–40.163
Transfers	
11 Credits	9.878
12 Debits	–11.320
13 Net transfers abroad (11 + 12)	–1.442
14 Balance on current account (7 + 10 + 13)	35.808
Capital and financial account transactions	
15 Capital account	–1.116
Financial account	
16 Net Direct investment – Incoming to Australia	27.599
17 Net Portfolio investment – Incoming to Australia	–29.691
18 Net Financial derivatives Incoming	4.755
19 Net Other investment Incoming	–55.791
20 Change in official reserve assets	18.522
21 Balance on capital and financial account (14 + ... + 19)	–35.722
Balancing item (Net errors and omissions)	+0.086

Source: Australian Bureau of Statistics, Cat. No. 5302, Table 30, https://www.abs.gov.au

Current account

Current account balance on merchandise trade
The value of a nation's merchandise imports subtracted from its merchandise exports.

The first section of the balance of payments is the current account, which includes trade in currently produced goods and services. The first component of the current account is the **current account balance on merchandise trade** (line 3 in **Exhibit 18.6**).

The current account balance on merchandise trade is the value of a nation's merchandise imports subtracted from its merchandise exports. As shown in **Exhibit 18.6**, Australia had a merchandise trade surplus of $70.535 billion in 2019–20.

Of course, a nation's trade consists of more than just goods. For example, each year Australia enjoys many visits from tourists, and many international students come here for education. Their spending in Australia is counted as services exports. On the other hand, Australians enjoy, for example, going to movies made overseas and making trips overseas. Such spending is recorded as services imports.

Current account balance on goods and services
This represents the nation's net position in respect of its exports and imports of both goods and services.

When services exports and imports are added to the balance of merchandise trade, we have the so-called **balance on goods and services** (item 7 in **Exhibit 18.6**). In 2019–20, Australia's balance on goods and services was a surplus of $77.413 billion.

You should recall at this point the fundamental national income identity of GDP = $C + I + G + (X - M)$, which was first introduced in **Chapter 11**. The current account balance on goods and services is the $(X - M)$ component in this identity. In 2019–20 this amounted to a surplus of $12.2 billion; that is, $(X - M)$ = $77.413 billion.

This means, therefore, that Australia's total production of goods and services (GDP) exceeded total spending that year by consumers, business and governments $(C + I + G)$ by $77.413 billion. This excess of domestic production over domestic demand was sold abroad.

Items 10 and 13 of the current account in **Exhibit 18.6** list ways other than direct trade in goods and services that result inflows of dollars into and out of Australia in a particular reporting period on account of the current period activity.

Income flowing into Australia from the investments abroad of Australian residents – such as shares in foreign companies, real estate owned abroad and loans made to offshore residents – are payments for the services of this Australian investment abroad. These income flows would include interest payments on the loans, rent on the property and dividends on the shares. Analogously, residents of foreign countries also receive income flowing from the application of their investments in Australia. The difference between these flows in aggregate is known as **net foreign income** (also referred to in recent times by the ABS as 'net foreign primary income').

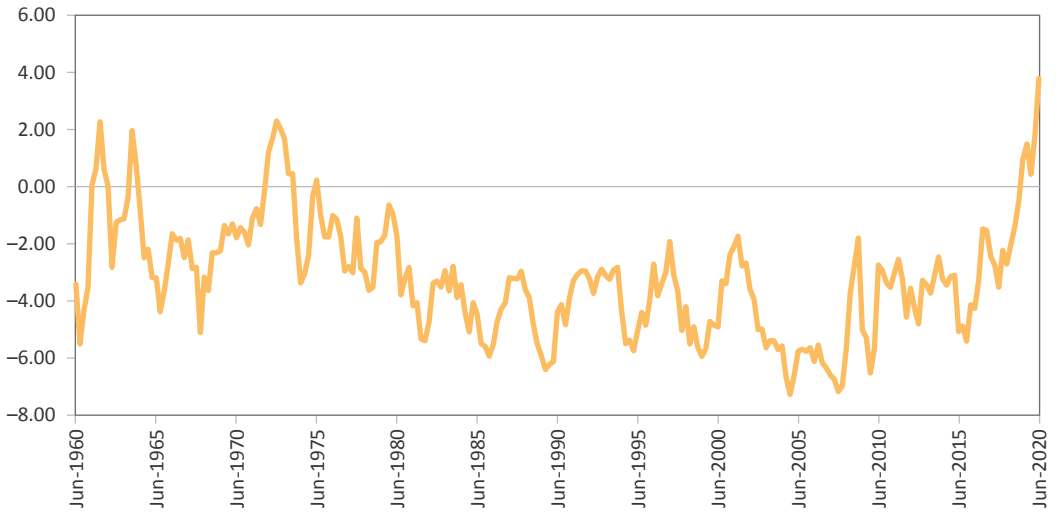
Net foreign income
Also known as 'net foreign primary income'; the difference between aggregate investment income flows into and out of a country. Such income includes interest payments on loans, rent on property and dividends on shares.

Item 10 of **Exhibit 18.6** reports a net flow of such investment income out of Australia of $40.163 billion in 2019–20, which was smaller than the goods and services trade surplus of $77.413 billion! This is unusual for Australia and usually Australia's investment income out of the country is much larger than the goods and services trade surplus. The above arises on account of the very significant foreign investment in the country over the past century. More will be said on this below.

Finally, we consider item 13: net transfers abroad (also referred to in recent times by the ABS as 'net foreign secondary income'). This category includes gifts made by the Australian government, charitable organisations or private individuals to other governments or private parties elsewhere in the world. Also included here would be pensions paid to Australian residents living abroad, for example. Net transfers for Australia in 2019–20 was estimated to be an outflow of 1.442 billion.

Adding items 7, 10 and 13 gives the current account balance (item 14) for 2019–20, namely a surplus of $35.808 billion. This is a current account surplus and means that Australians receive more payments from foreigners than the payments Australians made to foreigners. Usually, Australia experienced current account deficits in the past. A current account deficit means that, in aggregate net terms, Australians undertook a higher level of current payments to foreigners than foreigners made to Australians; or, equivalently, the current account deficit arose because the nation in aggregate spent more than its income. Australia's current account balance – as a percentage of GDP since 1959–60 – is provided in **Exhibit 18.7**.

Exhibit 18.7 **Australia's current account balance, percentage of GDP, 1959–60 to 2019–20**



Source: ABS data catalogue 5206.0, Table 3, https://www.abs.gov.au/statistics/economy/national-accounts/australian-national-accounts-national-income-expenditure-and-product/latest-release; and 5302.0 Table 4, https://www.abs.gov.au/statistics/economy/international-trade/balance-payments-and-international-investment-position-australia/latest-release

In the almost 60 years since 1959–60, Australia has experienced current account deficits in every year except 1961–62, 1963–64, 1972–73 and 2019-20.

As can be seen from the exhibit, Australia has consistently registered a current account deficit in almost every year except 1961–62, 1963–64, 1972–73 and 2019-20. More will be said on this shortly.

Capital and financial account

The second section of the balance of payments is the capital and financial account, which records the dollar payment flows on account of the purchases of foreign assets by Australians or purchases of Australian assets by foreigners. Examples include purchasing shares or bonds, extending loans, buying government securities or directly purchasing property.

You should be sure to recognise that what are recorded on the capital and financial accounts are the net new purchases (or sales) of such investment assets during the period in question, whereas what are recorded on the current accounts (in the net foreign income section) are the net dividends, rent and interest payments that flow into or out of Australia during the period in question on account of the total existing stock of such investment assets.

As **Exhibit 18.6** shows, in 2019–20 it was estimated that Australia experienced a capital account deficit of $35.722 billion (item 21). This means that, in 2019–20, in aggregate net terms, Australians made investments in foreign countries $35.722 billion in excess of foreign investments in Australia. In the past, Australia experienced capital and financial account surpluses, which means more foreigners made investments in Australia than Australians made investments abroad.

Capital and current accounts exactly offset

You will also notice that included in **Exhibit 18.6** is a line titled 'balancing item' (also referred to by the ABS as 'Net Errors and Omissions'). The inclusion of this item illustrates that, in principle, the two accounts should

exactly balance, and that any difference is simply due to measurement errors involved in trying to account for all of a country's many millions (or even billions) of transactions.

In 2019–20 the estimated value of the current account surplus was $35.808 billion but the ABS estimate of the capital account deficit amounted to only $35.722 billion. The balancing item takes the value equal to the difference between the two estimates and, by convention, its value is such that, when added to the capital account estimate the result exactly balances the estimated current account outcome. In 2019–20 it was $0.086 billion. But why must Australia's current account surplus of $35.808 billion in 2019–20, in principle, be matched with an exactly offsetting capital account deficit for the same year? The reason is very simple. As noted above, the current account surplus represents the extent to which a nation's income in a particular period exceeds its spending. In Australia's case, in 2019–20 the nation's residents earned an estimated $35.808 billion more than their aggregate spending. If Australia had current account deficits, then this means the nation's spending in a particular period exceeds its income.

The excess of $35.808 billion was provided to foreigners by Australia, and something needed to be provided to Australians in return to induce them to do this. What was provided were claims to foreign-based assets – either loans made to foreigners, property purchases by Australians or purchases by Australians of shares in foreign companies. In effect, Australians were willing to accommodate the additional expenditure by becoming investors in other countries.

An analogy may clarify this important point. Imagine that in a particular year a photocopying business owner earned $60 000 after running costs from her small photocopying business. Of this, she spent $40 000 on her own personal consumption but also invested in some new photocopiers to the value of $10 000, and the other $10 000 being was saved in her bank. The business owner's spending of $50 000 was less than her income by $10 000 and so her 'current account surplus' was $10 000. This current account surplus was exactly offset by a 'capital account deficit' of $10 000 in the form of the savings in the bank. She now has asset in the bank.

We can use a similar example to illustrate how current account deficit will be offset by a capital and financial account surplus. If the photocopying business owner earned $60 000 and of this, she spent $50 000 on her consumption but also invested in some new photocopiers to the value of $20 000. For the purchase of these machines, $10 000 was provided from her saving and the other $10 000 being provided by her bank. The business owner's spending of $70 000 was greater than her income by $10 000 and so her 'current account deficit' was $10 000. This current account deficit was exactly offset by a 'capital account surplus' of $10 000 in the form of the loan from the bank. She now has a liability to the bank.

Are current account deficits a bad thing?

In **Exhibit 18.7** we saw that Australia almost always experiences current account deficits, and we now understand that a current account deficit means that as a nation we are spending more than the income we generate. A result of this is that, in net terms, foreigners are continually accumulating Australian assets. Surely this must be a bad thing and cannot possibly be sustained indefinitely!

Let's look a little more closely at this question, using the photocopying business example again. The business owner's spending consisted of consumption, C, of $50 000 plus investment, I, of $20 000 for a total spending ($C + I$) of $70 000, which exceeded her income (her GDP as it were) of $60 000. However, notice that out of her income of $60 000 she spent only $50 000 on consumption; the rest she saved. In this case, her saving of $10 000 took the form of investment back into her business.

The reason for her 'current account deficit' was, therefore that her desired investment of $20 000 exceeded her saving of $10 000. To fund the other $10 000 of desired investment, she sought a bank loan. The bank would have made a business decision that she was very likely to be able to pay the required interest as well as repay the loan principal. For her part, she would have judged that her business prospects were such that she would

be able to pay the interest on the loan, pay back the principal, and also be able to generate some additional net economic benefits for herself.

The same is true for a nation's current account deficit. Provided it is being incurred to fund additional investment beyond the nation's available saving, it is unlikely to prove a problem. There will undoubtedly be additional liabilities to foreigners being accumulated. However, the resulting dividends, interest and rent payments payable overseas on account of the extra foreign liabilities should be able to be made out of the additional future production made possible from the extra capital stock. Furthermore, there is also likely to be a net residual benefit remaining that will accrue to Australian residents so that their standard of living will be higher than it otherwise would have been without the use of the foreign investment.

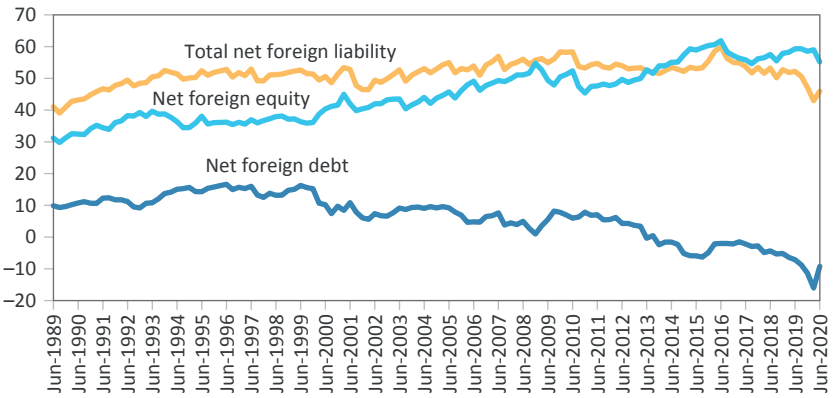
This has certainly been the case for Australia. Over the past two centuries, it has had an abundance of investment opportunities but has been short of both labour and capital. To solve this, it has imported both. Its workforce has been increased much more rapidly by having a significant immigration program. It has also imported foreign capital to supplement domestic saving. Given that the provision of this capital stock has been shared between residents and foreigners, the benefits deriving from it have also quite properly been shared by both Australian and foreign residents.

Australia's foreign liabilities – should we be concerned?

As noted above, every time Australia runs a current account deficit, foreigners either purchase Australian assets or grant loans to Australians equivalent to the current account deficit. A loan adds to our net foreign debt liabilities, while the purchase of a share in an Australian asset adds to Australia's net foreign equity liabilities. The sum of these two types of liabilities is referred to as Australia's total net foreign liabilities.

Exhibit 18.8 provides a graphical representation of Australia's net foreign liabilities expressed as a percentage of Australia's annual GDP since 1990. As can be seen, there has been a rise in both total net foreign liabilities and the debt component over the past 27 years. The net foreign equity component has actually steadily reduced since the late 1990s and has in fact become negative since 2014, meaning that, since then, Australians have had greater foreign equity holdings offshore than foreigners have had in Australia.

Exhibit 18.8 **Australia's net foreign debt, net foreign equity and total net foreign liabilities, 1989–2020, per cent of GDP**



Source: Reserve Bank of Australia, Table I5: Australia's Net Foreign Liabilities, https://www.rba.gov.au/statistics/tables/xls/i05hist.xls

In 1990, net foreign debt stood at around 32 per cent of GDP, and over the following 27 years it increased to 55 per cent in 2020. Total net foreign liabilities stood at around 43 per cent of GDP in 1990, and by 2020, stood at 45 per cent. It is reasonable to ask whether the current levels are too large and will prove to be a problem. In fact, back in the 1970s net foreign debt was around 3 per cent of GDP and total net foreign liabilities were around 10 per cent of GDP! Why are they so much larger now than they were 40 years ago?

Of course, an alternative question might be: were the ratios in evidence in the 1970s too small? Remember that in the 1970s the world had a very restricted international financial system that severely constrained international investment. A country's capital accumulation was therefore much more constrained to be commensurate with the saving of its own residents. This may have resulted in lower foreign liabilities in some countries, but it also meant slower capital stock accumulation and economic growth than might otherwise have been the case.

Today, financial capital and saving can flow much more easily across international borders, allowing savings to flow to wherever its likely maximum rate of return is located. The increase in Australia's foreign liabilities is most likely simply a reflection of the more deregulated international financial system of the past few decades.[8]

While the freeing up of international financial markets may explain the increase in Australia's foreign liabilities, one can still ask whether Australia's foreign liabilities may become a problem. This will only occur if servicing them becomes a problem, so if they have been incurred to fund income-generating capital assets, the servicing of these liabilities is not likely to become a problem.

One way to gain insight into this issue is to see what has happened over time to a country's net foreign income payable overseas as a share of its annual GNP. This ratio represents the proportion of a nation's income needed to pay the net interest, rent and dividends owing to foreigners on account of their ownership share of the country's assets. A large and sustained increase in the ratio over time means that a country is needing an increasing proportion of its income to service its existing net foreign liabilities, and this may be some cause for concern.

In Australia's case, the ratio averaged 3.4 per cent through the 1990s, then marginally edged up to average 3.5 per cent in the 2000s, and then averaged just 2.4 per cent in the five-year period from 2012–13 to 2016–17 inclusive, and was approximately what the ratio was in the 1980s (2.1%). In other words, a reasonable conclusion from this relatively long 37-year period would probably be that the ratio has essentially remained pretty stable.

However, it should be noted that the recent very low servicing ratios are no doubt a reflection of the very low levels of international interest rates prevailing in the post-GFEC years, and the fact that all of Australia's net foreign liabilities were, by 2014, comprised of net foreign debt. With international interest rates on the increase from 2017, the ratio could be expected to increase back above 3 per cent. Notwithstanding this probability, it is nonetheless highly unlikely that the current level of foreign liabilities in Australia is problematic.[9]

In summary

- The **balance of payments** is a summary record of the international transactions of a country. It consists of a current account and a capital account, which in principle exactly balance.
- A current account deficit occurs when a nation's residents spend more than their income.
- A current account deficit implies an increase in a nation's foreign liabilities – either increased foreign debt or increased foreign equity in the nation's assets.
- Servicing foreign liabilities is not likely to become a problem if the liabilities were incurred to finance an increase in the nation's capital stock.

[8] For further discussion on this, see A. P. Layton and T. Makin, 'Estimates of the macroeconomic impact of foreign investment in Australia', *International Economic Journal*, Vol. 7, pp. 35–42, 1993, cited by the Chairperson of the US Federal Reserve System (US Fed), Janet Yellen, in a conference address, 'Reaping the full benefits of financial openness', which she gave at the Bank of Finland's 200th Anniversary Conference, Helsinki, May 2011.

[9] Figures derived from various entries in ABS, Cat 5302, Table 30 and Cat. 5206, Table 38. Another interesting ratio relevant to this discussion is the ratio of net foreign income payable overseas to a country's total exports of goods and services.

4 The exchange rate and its economic impact

In the previous discussion, no mention was made about currencies and exchange rates. And yet, different countries have different currencies, and when investors want to channel some of their funds into another country they need to exchange their currency for that of the foreign country. How is the price of another country's currency in terms of the home country's currency determined? It is to this issue that we now turn.

Each transaction recorded in the balance of payments requires an exchange of one country's currency for that of another. Suppose an Australian buys a Japanese-made Mazda car. Mazda wants to be paid in yen, not dollars, so dollars must be traded for yen. On the other hand, a Japanese meat-processing company needs to pay its Australian beef suppliers in dollars, and so it has to convert some yen into dollars.

What is the exchange rate at which these currency conversions take place? The **exchange rate** is the number of units of one nation's currency that equals one unit of another nation's currency. For example, assume NZ$1.25 is exchangeable for one Australian dollar. This means the exchange rate is A$1 = NZ$1.25. Alternatively, the exchange rate can be expressed as a reciprocal. Dividing A$1 by NZ$1.25 gives A$0.80 per NZ$1.

Exchange rate
The number of units of one nation's currency that equals one unit of another nation's currency.

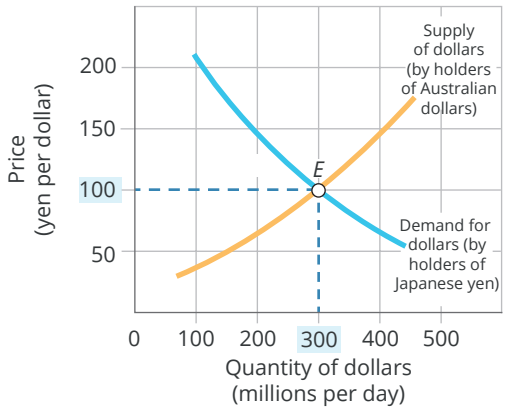
We now turn to how an exchange rate is determined.

Supply and demand for foreign exchange

The exchange rate for dollars – or any nation's currency – is determined by international forces of supply and demand. For example, consider the exchange rate for yen and Australian dollars (AUD) shown in **Exhibit 18.9**. The quantity of dollars exchanged is measured on the horizontal axis and the value of the AUD expressed as a number of yen per dollar is measured on the vertical axis.

The demand for AUD (from holders of yen) comes from Japanese individuals, corporations and governments who want to buy Australian exports and assets, as well as from other international investors around the world who may be holding Japanese yen but believe that, in the future, the Australian dollar may gain in value against it. The Japanese buyers seek to exchange their yen for dollars and the investors will seek to buy a quantity of Australian dollars in exchange for a quantity of Japanese yen they may be holding.

Exhibit 18.9 **The supply of and demand for Australian dollars (in exchange for Japanese yen)**



The number of Japanese yen per Australian dollar in the foreign exchange market is determined by the demand for dollars by holders of Japanese yen and the supply of dollars by holders of Australian dollars. The equilibrium exchange rate is 100 yen per dollar and the equilibrium quantity is A$300 million exchanged per day.

The demand curve for dollars – or any foreign currency – is downward-sloping. As the number of yen per dollar declines this means Australian goods and investment opportunities are less expensive to Japanese buyers because they must pay fewer yen for each dollar. Thus, as the yen price of an AUD decreases, the quantity of dollars demanded to purchase beef, shares, land and other Australian products and investments increases.

For example, suppose a BHP Billiton share has an A$40 price tag. If the exchange rate is 100 yen to the dollar, a Japanese investor would pay 4000 yen. If the price of dollars falls to 80 yen each, the same $40 share will cost Japanese investors only 3200 yen. This lower price will encourage Japanese investors to increase their purchases of Australian shares and other goods and services, which in turn increases the quantity of dollars demanded.

The supply curve of dollars is upward-sloping. The supply of dollars in this market flows, for example, from individuals, corporations and governments in Australia that want to buy Japanese goods and services, shares and land, and make other investments in Japan.[10]

Because the Japanese goods and services must be paid for in yen, as the number of yen per dollar increases, the prices of Japanese products fall to those holding Australian dollars. When the number of yen per dollar rises, Australians respond by seeking to purchase more Japanese goods, services and assets, which in turn increases the quantity of dollars supplied in exchange.

The foreign exchange market in **Exhibit 18.9** is in equilibrium at an exchange rate of 100 yen for $1. As was explained in Chapter 3, it is the interaction of demand and supply for a product that determines its price and quantity traded. In this case it is the interaction of demand and supply for dollars in exchange for yen that will lead to the prevailing exchange rate and the quantity of dollars exchanged for yen.

Shifts in supply and demand for foreign exchange

For the years between the Second World War and 1971, the world's exchange rates were fixed and the values of currencies and the rates at which they exchanged were based primarily on their accepted values in terms of gold. In 1971, most western nations – but not Australia – agreed to abandon the 'gold standard', stop fixing their exchange rates, and to allow their currencies to adjust continuously according to the forces of supply and demand. This type of system is known as a **flexible exchange rate system**. In December 1983, Australia also adopted a flexible exchange rate system.

Flexible exchange rate system
A system in which countries allow their currencies to adjust continuously according to the forces of supply and demand. Most developed countries in the world today employ this type of exchange rate system.

Exhibit 18.10 illustrates that the exchange rate can and does fluctuate widely. For example, in 1984 one AUD was worth about 210 Japanese yen and US$0.85. After fluctuating during the 1980s and 1990s – but essentially trending downwards – the AUD hit post-war lows of about 59 yen per dollar and US$0.49 in 2001. At the time of writing in September 2020, the AUD was worth about 75 yen and around US$0.71, respectively.

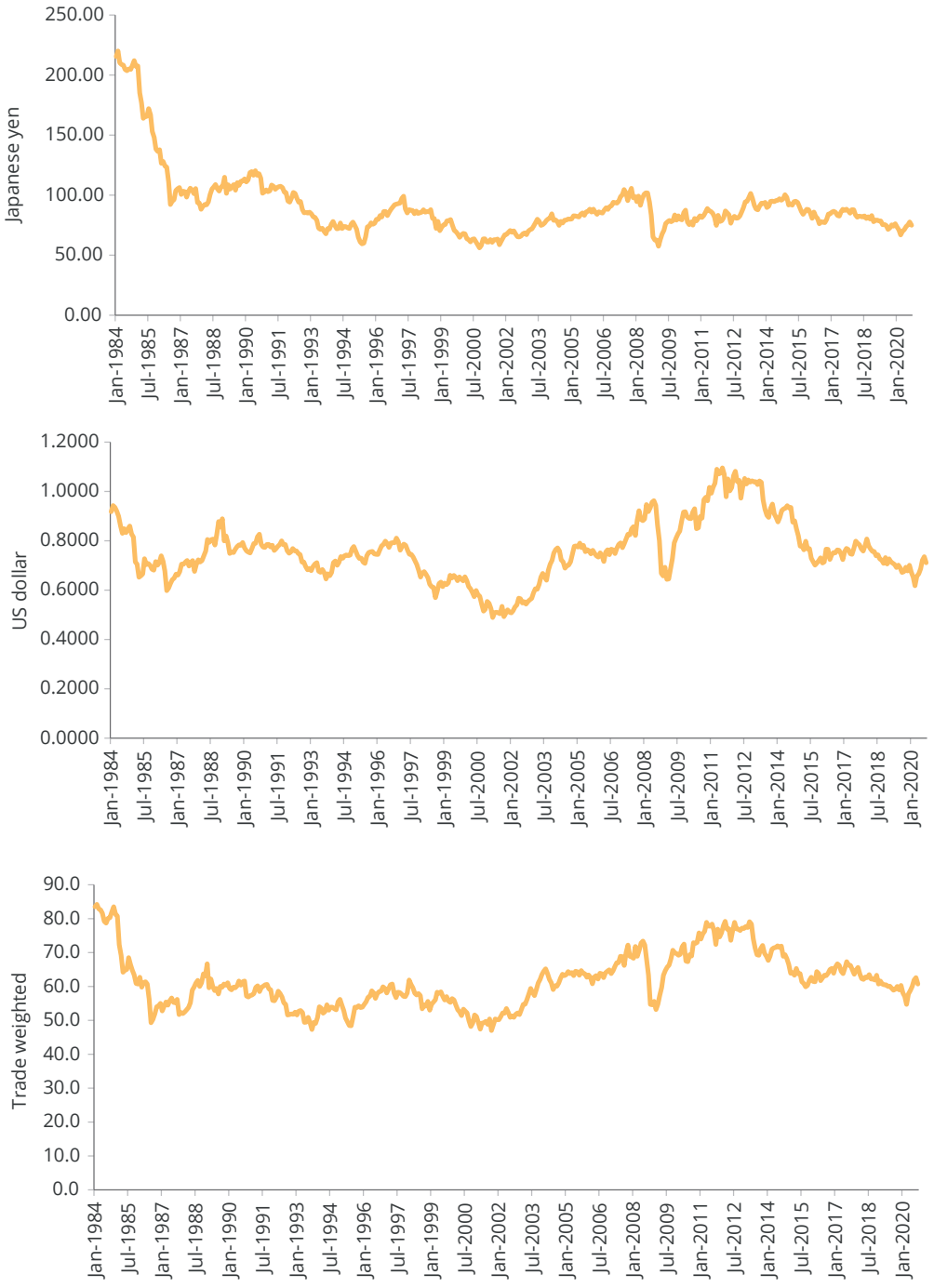
A broader measure of the value of the AUD is provided in the RBA's Trade Weighted Index (TWI, with base of 100 in May 1970) value of the AUD. The TWI is a weighted average (weighted by the relative amount of trade Australia has with the countries whose currencies are included in the index) of the AUD against the currencies of Australia's major trading partners. In late 2020 its value was around 60.7.

Recall from **Chapter 4** that the equilibrium price for products changes in response to shifts in the supply and demand curves. The same supply and demand analysis applies to equilibrium exchange rates for foreign currencies. Four important sources of shifts in the supply and demand curves for foreign exchange are shifts in tastes and preferences across countries, changes in relative incomes across countries, changes in relative price levels across countries, and changes in relative real interest rates across countries.

As demand and supply for currencies adjust over time, their rate of exchange changes. For example, if changes to demand and supply of yen for the AUD result in a reduction in the equilibrium number of yen per dollar, then

[10] Again, there will also be international investors holding Australian dollars in their portfolios who may, depending on the current value of the dollar in terms of yen compared with their perceptions of what might be the relative value in the future, also want to offer dollars for sale in exchange for yen.

Exhibit 18.10 **Changes in the AUD/yen and the AUD/USD exchange rates, and the AUD in terms of the TWI, 1984–2020**



Source: Reserve Bank of Australia, *Statistics, Exchange Rates*, Tables F11, http://www.rba.gov.au/statistics/tables/index.html#exchange-rates

Today, most economies are on a system of flexible exchange rates. As demand and supply curves for currencies shift, exchange rates also change. In 1984, one AUD was worth about 210 Japanese yen and US$0.85. By 2001, the AUD had dropped to lows of about 59 yen and US$0.49 but subsequently rose to be around US$0.71 and 75 yen in late 2020. The TWI is a weighted average value of the AUD against the currencies of Australia's major trading partners. In late 2020 it was around 60.7 (with base = 100 in 1970).

Depreciation of a currency
A fall in the price of one currency relative to another.

Appreciation of a currency
A rise in the price of one currency relative to another.

Terms of trade (TOT)
Defined as the ratio of an index of a country's export prices to an index of its import prices.

the AUD is said to depreciate against the yen. **Depreciation of a currency** is a fall in the value of one currency relative to another.

Alternatively, if changes to demand and supply of say the Malaysian ringgit for the AUD result in a rise in the equilibrium number of ringgit per dollar, then the AUD is said to appreciate against the ringgit. **Appreciation of a currency** is a rise in the value of one currency relative to another.

Over the longer term a nation's exchange rate is, *inter alia*, very significantly influenced by its **terms of trade (TOT)**, defined as the ratio of an index of its export prices to an index of its import prices. Higher values of the TOT mean a given quantity of a nation's exports can buy relatively more quantities of imports and so higher values of the TOT are usually regarded as a good thing. Furthermore, since Australia is a major international commodity exporter, a very important determinant of its TOT is the international price of commodities, especially mineral commodities.

The economic impact of exchange rate fluctuations

Now it is time to stop a minute, take a breath and draw some important conclusions. As has just been explained, exchange rates between most major currencies are flexible. Instead of being pegged to gold or another fixed standard, their value is determined by the laws of supply and demand. Consequently, shifts in supply and demand result in a lower or higher exchange rate for the AUD.

However, it should be understood that exchange rates do not fluctuate without any influence from the monetary authorities. Central banks like the RBA will often buy and sell currencies to prevent wide swings in exchange rates. Notwithstanding this, it needs to be emphasised that its objective is not to try to reverse a change in a currency's value resulting from changes to its fundamental determinants. Instead, its actions are an attempt to smooth out excessive fluctuations in the exchange rate and to bring about an orderly market adjustment of the currency's value to whatever is its apparent new equilibrium level.

It is also essential to understand that the international price of any nation's currency has a profound impact on its economy. A lower value for the AUD, for example, enhances the profitability of Australian products exported and sold on international markets – and, in so doing, will enhance employment prospects in those businesses. For example, suppose an Australian exporter's production cost is A$20 per unit, the current international price of the product is US$25,[11] and the exchange rate is A$1 = US$1. This means the exporter's AUD profit per unit sold is A$5 (= US$25 − A$20 = A$25 − A$20 = A$5).

Now, suppose the AUD depreciates to say, US$1 = A$1.25 (or, equivalently, A$1 = US$0.8), the exporter's AUD profit per unit sold is now A$11.25 (= US$25 − A$20 = A$31.25 − A$20 = A$11.25). This means more AUD profit per sales unit for the exporter, which may well encourage them to produce more for export, and quite possibly increase employment in the business. Such a depreciation also means import-competing firms in Australia are better able to compete domestically with imported goods and services because the AUD prices of imports (whose prices are mostly determined offshore in foreign currencies) go up on account of the depreciated value of the AUD. Thus, Australian economic activity is likely to be stimulated by a depreciated dollar with a commensurate increase in jobs.

[11] For most Australian export markets, Australia is a price taker, meaning the product price is usually set in foreign currencies by the forces of international market demand and supply. In a few markets Australia has some pricing power and therefore AUD currency movements may result in changes in the international price itself.

The downside is that it also means that those Australians who like to consume imports, travel overseas or need to use imported goods in their production processes, face higher AUD costs. A lower value for the AUD also carries the risk that domestic inflation may increase as a result of the higher AUD prices of imported goods and services.

Australia's flexible exchange rate has acted to buffer Australia's economy from quite a number of very significant international shocks since its floating in 1983. It most certainly played an important role in the Asian Financial Crisis of 1997–98. It played an even more important role in the GFEC and was no doubt a significant factor in Australia escaping recession. Some details are provided in the next chapter.

A higher value for the AUD, on the other hand, has the opposite effects to those described above. It reduces the AUD profits of Australian exporters, thus impacting negatively on their business, possibly leading them to scale back their workforce. It also means import-competing firms are less able to compete with imported goods and services because the AUD prices of imports go down. Thus, due to these two impacts, Australian economic activity is likely to be retarded by an appreciated dollar, with a commensurate loss of jobs.

In summary

- An **exchange rate** is the price of one currency in terms of another.
- Today most countries maintain a **flexible exchange rate** system.
- Australian export and import-competing firms benefit when the Australian dollar **depreciates**.
- Australian export and import-competing firms are detrimentally affected when the Australian dollar **appreciates**.

Key concepts

Comparative advantage
Absolute advantage
Free trade
Protectionism
Embargo
Tariff
Quota
Fair trade
Strategic trade theory
Dutch Disease
Free trade agreement (FTA)
Balance of payments
Current account balance on merchandise trade
Current account balance on goods and services
Net foreign income
Exchange rate
Flexible exchange rate system
Depreciation of a currency
Appreciation of a currency
Terms of trade (TOT)

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand why nations benefit from trade

- When countries specialise, total world output increases and, therefore, the potential for greater total world consumption also increases.

- **Comparative advantage** refers to the relative opportunity costs of producing the same goods between different countries.
- World output and consumption are maximised when each country specialises in producing goods for which it has a comparative advantage and trading them for those in which it does not have a comparative advantage.

2. Understand what is meant by free trade, fair trade and protectionism

- **Free trade** is the flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow.
- Governments around the world use **embargoes**, **tariffs**, **quotas** and other restrictions to protect domestic producers from foreign competition.
- Some support a so-called '**fair trade**' approach, and some governments opt to pursue strategic trade approaches to their international trade development.
- The political economy considerations that militate against reducing **protectionism** include the so-called *infant industry argument*, the *national security argument*, the *employment argument* and the *cheap foreign labour argument*.
- **Free trade agreements** (sometimes known as preferential trade agreements [PTAs]), involve countries party to a free trade area agreeing to eliminate or significantly reduce tariffs and other trade barriers among themselves. They have become more prevalent in recent decades as a multilateral approach – via the WTO – to reducing trade barriers has been increasingly fraught.

3. Understand the concepts of a country's balance of payments and its foreign liabilities

- The **balance of payments** is a summary record of the international transactions between a country and other countries during a given period of time, and it consists of a *current account* and a *capital and financial account*, which in principle exactly balance each other.
- A *current account surplus* occurs when a nation's residents spend less than their income during a period, and this shortage of spending is reflected by a capital and financial account deficit.
- In contrast, the *current account deficit* implies that a nation has insufficient saving and excessive spending, and this will be financed by capital and financial account surplus. It also implies an increase in a nation's foreign liabilities – either increased foreign debt owed by the nation's residents or an increase in foreign equity in the nation's assets. Servicing – that is, by the payment of interest, dividends and rent – of the nation's foreign liabilities is not likely to become a problem if the liabilities were incurred to finance an increase in the nation's capital stock.

4. Understand the foreign exchange rate and its economic impact

- An **exchange rate** is the price of one currency in terms of another.
- Today most countries maintain a **flexible exchange rate system** where the level of an exchange rate continuously changes with the relative demand and supply for the two currencies involved.
- A **depreciation** of the Australian dollar assists Australian export and import-competing firms, thereby potentially generating more employment.
- An **appreciation** of the Australian dollar negatively affects export and import-competing firms, thereby potentially leading to losses in employment.

Study questions and problems

1 Caleb can paint either two walls or two window frames in one hour. In the same time, Declan can paint either three walls or one window frame. There are four window frames and four walls to be painted in a job. How many person-hours would the total job take if Caleb did it all? How many if Declan did it all? How many person-hours would the total job take if Caleb did only the windows and Declan did only the walls?

2 The countries of Alpha and Beta produce steel and copper. The production possibilities schedule shown here describes their potential output in tonnes per year.

Points on PPF	Alpha		Beta	
	Steel tonnes	Copper tonnes	Steel tonnes	Copper tonnes
A	135	0	120	0
B	90	15	80	20
C	45	30	40	40
D	0	45	0	60

Using the data in the table, answer the following questions:

a In which good does Alpha have a comparative advantage?

b In which good does Beta have a comparative advantage?

c Suppose Alpha is producing and consuming at point *B* on its production possibilities frontier and Beta is producing and consuming at point *C* on its production possibilities frontier. Use a table, such as the one in **Exhibit 18.3**, to explain why both nations would benefit if they specialised.

d Draw a graph and use it to explain how Alpha and Beta benefit if each completely specialises in the production of the product for which they have a comparative advantage and agree to trade steel and copper to the other.

3 Consider this statement: 'The principles of specialisation and trade according to comparative advantage among nations also apply to states in Australia.' Do you agree or disagree? Explain.

4 Suppose Australia passed a law stating that it would not purchase imports from any country that imposed any trade restrictions on its exports. Abstracting from Australia's exporters for the time being, analyse who would benefit directly and who would lose directly from such retaliation? What of the indirect impacts? Could this policy adversely affect Australian exporters in any way?

5 Consider this statement: 'Reducing tariffs on imports, by increasing competition for Australian-made products from imported products, destroys Australian jobs.' Do you agree or disagree? Explain.

6 Discuss this statement: 'Because each nation's balance of payments equals zero, it follows that there is actually no significance to a current account deficit or surplus.'

7 Discuss this statement: 'Since current account deficits, by definition, increase a country's net foreign liabilities, they should be avoided and all countries should seek to achieve current account surpluses.'

8 The following table summarises supply and demand for the yen.

	Yen per AUD			
	Y75	Y85	Y95	Y105
Quantity of AUD demanded (per day)	500	400	300	200
Quantity of AUD supplied (per day)	200	300	400	500

Using the above table, do the following:

a Graph the supply and demand curves for AUD.

b Determine the equilibrium exchange rate.

c Determine what the effect of a fixed exchange rate at Y85 per AUD would be.

Answers to 'You're the Economist'

Do nations with an absolute advantage gain from trade?

In Australia, the opportunity cost of producing one calculator is 100 towels. In Indonesia, the opportunity cost of producing one calculator is 50 towels and is therefore lower in Indonesia than in Australia. Due to this difference in opportunity costs between the two countries, Indonesia has a comparative advantage in calculator production and Australia has a comparative advantage in towel production.

If Australia diverted the required production inputs to produce four calculators to towel production, 400 more towels would be produced at the cost of the four calculators. If Indonesia diverted the production inputs required to make 200 towels into calculator production, four more calculators could be produced. Australia could export 200 of its additional towels to Indonesia in return for Indonesia's four additional calculators. This leaves Indonesia no worse off for the same level of production inputs and Australia better off by an extra 200 towels for the same quantity of its production inputs.

If you said that even though it has an absolute advantage in both goods, Australia can benefit by trading with Indonesia because it has a comparative advantage YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 Free trade theory suggests that when trade takes place between two nations:

a goods and services are traded free of charge similar to a barter system of exchange.

b both nations will be better off.

c both nations will be worse off.

d one nation must gain at the other nation's expense.

2 In the table shown here depicting production possibilities for two countries, which country has the comparative advantage in the production of tomatoes?

Tomatoes and corn output (tonnes per hour)		
Country	**Tomatoes**	**Corn**
Australia	2	3
South Africa	3	2

a Australia because it has higher total productivity per hour than South Africa.

b South Africa because it can make more tonnes of tomatoes in an hour than Australia can.

c South Africa because, 1 tonne of tomatoes has an opportunity cost of 0.5 of a tonne of corn production.

d Australia because it can produce 1.5 tonnes more corn for every tonne of tomatoes it forgoes.

3 In the previous table, the opportunity cost of 1 tonne of corn is:

a 1.5 tonnes of tomatoes in Australia and 0.66 tonnes of tomatoes in South Africa.

b 0.66 tonnes of tomatoes in Australia and 1.5 tonnes of tomatoes in South Africa.

c 1 tonne of tomatoes in Australia and 1 tonne of tomatoes in South Africa.

d 2 tonnes of tomatoes in Australia and 3 tonnes of tomatoes in South Africa.

4 If the countries in the table in question 2 follow the principle of comparative advantage, South Africa should, *ceteris paribus*:
- **a** buy all of its tomatoes from Australia.
- **b** specialise in tomato production and buy the corn it needs from Australia.
- **c** produce both tomatoes and corn and not trade with Australia.
- **d** buy all of its tomatoes and corn from Australia.

5 A tariff increase directly increases:
- **a** the quantity of imports.
- **b** the ability of foreign goods to compete with domestic goods.
- **c** the prices of imports to domestic buyers.
- **d** The quantity of exports.

6 Which of the following international accounts records payments for exports and imports of goods and services as well as foreign investment income and foreign transfers?
- **a** The capital account.
- **b** The current account.
- **c** The federal budget account.
- **d** The income and expense account.

7 If a Japanese mobile phone priced at 17 000 yen can be purchased for A$200, the exchange rate is:
- **a** 34 dollars per yen.
- **b** 85 yen per dollar.
- **c** 85 dollars per yen.
- **d** 34 yen per dollar.

8 Australia:
- **a** has been on a fixed exchange rate system since the Bretton Woods agreement in 1945.
- **b** was on a flexible exchange rate system prior to late 1971, but now is on a fixed exchange rate system.
- **c** now has a floating exchange rate against any bilateral currency but maintains a fixed peg against a trade-weighted basket of currencies of our major trading partners (the TWI peg).
- **d** has been on a floating exchange rate system since 1983.

9 Which of the following international accounts records the purchase and sale of assets between Australia and other nations?
- **a** The net asset trade balance account.
- **b** The net external asset acquisition account.
- **c** The current asset account.
- **d** The capital and financial account.

10 If the AUD appreciates, this causes:
- **a** the AUD value of export sales to increase.
- **b** the AUD value of export sales to fall.
- **c** the AUD price of imports to fall.
- **d** both b and c.

19 Macroeconomic policy issues: reflecting on economic and debt crises

Learning objectives

In this chapter, you will be pursuing the following key learning objectives:

1 Understand the essential cause of the global financial and economic crisis (GFEC) and its timing.
2 Gain some appreciation of the economic consequences of the GFEC.
3 Gain insight into the economic policy responses around the world to the GFEC and develop an understanding of policy lessons learnt from the crisis.
4 Gain economic insights into its aftermath, especially the so-called European sovereign debt crisis, itself at least partially brought about by the GFEC.
5 Understand the economic impact and policies of the COVID-19 pandemic.

Global financial and economic crisis (GFEC)
Also referred to as the Global Financial Crisis (GFC), the GFEC was a period of tremendous international financial system instability, which began in 2007, intensified in 2008, and ushered in a synchronised international recession dating from April 2008 to June 2009. In the US, the recession in that country associated with the GFEC is often referred to as the 'Great Recession'.

In this chapter we will outline in some detail a range of matters pertinent to the **global financial and economic crisis (GFEC)** of 2008–09. We will also discuss matters relating to the European sovereign debt crisis that was, at least in part, the result of the GFEC. The GFEC – also known as simply the global financial crisis (GFC) – has undoubtedly been the most significant macroeconomic event in recent decades. In the US, the recession in that country associated with the GFEC is sometimes referred to as the 'Great Recession'.

The descriptor of 'Great Recession' serves to simultaneously label it as an unusually severe recession, whilst at the same time consciously ranking it below the Great Depression of the 1930s in terms of its macroeconomic and social consequences. Indeed, many argue that it is because of the implementation of a range of macroeconomic policy lessons drawn from the Great Depression that the severity of the Great Recession was much less than it otherwise might have been. It is argued that without such policy initiatives, the Great Recession may well have turned into the Great Depression II.

Science Photo Library/NASA

But others argue that the actual policy responses put in place to combat the GFEC merely served as a Band-Aid, short-term solution that largely left unchanged the underlying root causes of the crisis. In fact, some would say that by mitigating the cleansing 'creative destruction'[1] of the economic downturn, the enacted policies may simply result in a postponed, but much more pronounced, pervasive and protracted economic adjustment sometime in the future.

The global COVID-19 pandemic in 2020 was a once-in-a-century shock to the world. With steps to stop the spread of the virus, we began to see disruption to global economies, including Australia's, since the movement of people became restricted and supply chains became disrupted. The Australian government and the Reserve Bank of Australia acted swiftly and decisively to address the economic impact of COVID-19. The government provided significant support to the health system and the economy and jobs by assisting the households and businesses.[2]

1 The unfolding of a worldwide financial maelstrom

In this section, the focus will be on understanding the essential cause of the GFEC and its timing.

The GFEC of 2008–09 had its origins in the so-called US sub-prime mortgage crisis of 2007–08. The term '**sub-prime mortgage**' refers to a housing loan extended to a borrower assessed by the originator of the loan as having a relatively higher level of difficulty in paying it off. A slang pejorative term in the US for such a loan is a so-called NINJA loan (no income, no job and no assets – on the part of the borrower). In the US, such loans historically accounted for only 7 or 8 per cent of outstanding housing loans. However, in the early 2000s – with US house prices rising quickly during that period – the percentage began to increase to a peak of around 20 per cent of all outstanding housing loans by 2007. This represented a considerable volume of some $1.3 trillion of financial assets of doubtful value.

Sub-prime mortgage
A housing loan extended to a borrower assessed by the originator of the loan as having a relatively higher level of difficulty in paying it off.

An important driver of the increase in such loans by originators is that the loans, once made, could then be *securitised* and sold by the originating lending institution to other financial institutions. **Securitisation** is the practice of bundling up a large number of similar financial assets – say, housing loans – into a new derivative financial asset. These new derivative assets (or 'securities') – offering a promised fixed rate of interest funded by the income stream from payments by the original borrowers – are then on-sold, usually to substantial international institutional investors like insurance and superannuation companies. **Exhibit 19.1** provides a simple visual schematic of the basic process.

Securitisation
The practice of bundling up a large number of similar financial assets – say, housing loans – into a new derivative financial asset (or 'security').

When the underlying assets are mortgages, which are being bundled together into the new financial security, the resulting derivative security is commonly called a **mortgage-backed security (MBS)**. Since traditionally there had been a very low rate of default on the repayment of interest and principal on the original underlying loans, these MBSs proved very popular with international investors. Furthermore, since the original lenders were able to 'get the loans off their books', they could make more loans with the proceeds of the sale. The original lenders do well, the investors are happy and, because the quantity of housing finance available is increased substantially (because the funds available for loan effectively come from around the world), the home buyers – and the housing construction industry – are also happy.

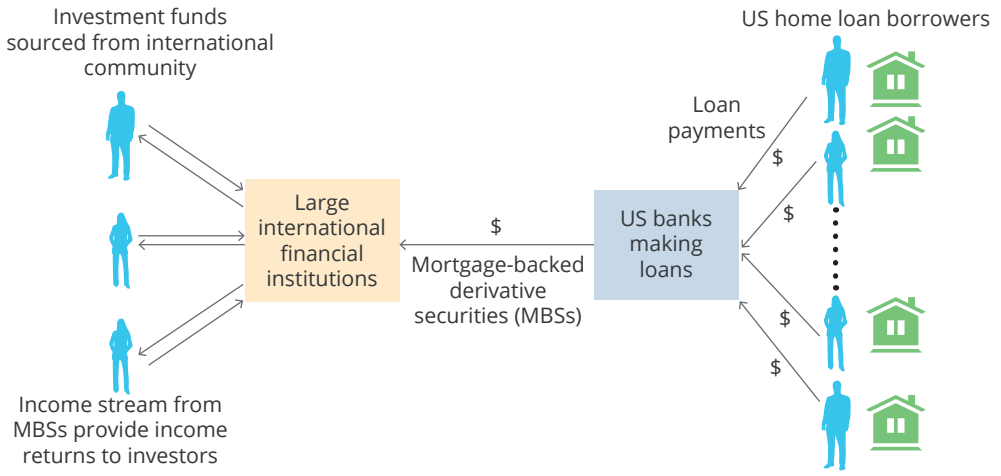
Mortgage-backed security (MBS)
In securitisation, when the underlying assets are mortgages that are being bundled together, the resulting derivative security is called a mortgage-backed security.

---

1 A term first coined by Joseph Schumpeter in his famous book *Capitalism, Socialism and Democracy* (1942), used to capture the beneficial aspects of economic downturns as scarce resources are freed from less efficient uses in marginal enterprises to be used more productively in others. M. Shahiduzzaman and A. Layton also recently refer to the idea in their *Applied Energy* 2015 paper, 'Changes in CO_2 emissions over business cycle recessions and expansions in the United States: A decomposition analysis', as one possible explanation of how a recession – in particular, the US recession of 2007–09 – may have longer-term impacts on carbon dioxide emissions beyond the end of the recession.

2 Economic and Fiscal update of the Australian government, available at https://budget.gov.au/2020-efu/economic-fiscal-update.htm.

Exhibit 19.1 A schematic of the securitisation process using home loan mortgages as the underlying original assets



In this process, everyone seems to win – but only so long as the default rates (i.e., borrowers failing to meet their interest payment obligations and/or failing to repay the loan principal) on the underlying loans stay small. And therein lay the basic problem with the whole process. The traditional lender/borrower scenario sees the lender continuing to bear the risk of default, ensuring that loans are only made to individuals who have a reasonable probability of repayment. However, if the risk can be shifted to another party, as in the securitisation process described above, a '**moral hazard**' situation arises. A moral hazard occurs when one person takes on more risk because someone else bears the costs. Here, as a result of the inherent moral hazard, lending criteria may, over time, be weakened and housing loans of considerable size end up being made available to borrowers with no deposit and possibly no real prospect of reliable repayment of interest and principal. Of course, the ultimate buyers and holders of the MBSs may – at least for a time – not be aware of this, and the increased riskiness of the securities goes unpriced in the market.

Moral hazard
A moral hazard occurs when one person takes on more risk because someone else bears the costs.

This is exactly what happened throughout the 2000s. By 2007, US households had increased their debt-to-income ratio up to just under 130 per cent after ranging between 80 per cent and 90 per cent through the 1990s up to 2000 (implying a 45 per cent increase on the 2000 ratio of 90 per cent!); the increasing indebtedness being mostly to buy increasingly more expensive housing stock. At the same time, US housing prices had peaked in the second half of 2006 and declined quite steeply through 2007.

The leverage levels of the borrowers were becoming excessive at a time when their properties financed from their debt were decreasing in value. The MBSs associated with the US housing mortgage markets were becoming increasingly risky, but this increased risk was not yet being adequately priced into the value of the MBSs. Eventually, default rates on the underlying loans started to increase significantly compared with what had been the previous experience in the US housing market. Investors around the world eventually began to realise the value of the MBSs they were holding was much less than they had thought (due to the higher real underlying risk).

This issue was further compounded by the inherent difficulty in trying to put an excellent risk-adjusted value on a given set of MBSs in one's portfolio. This arose from the derivative nature of the MBSs and the lack of any real knowledge of the creditworthiness of the original housing loan borrowers underpinning a particular set of MBSs. Their value began to be seriously questioned from mid-2007 and, as a result, their market value was quickly and significantly repriced downwards.

Moral hazard and the health sector

You're the economist

Australia has a universal healthcare system wherein those who choose not to have private health insurance nonetheless are guaranteed access to a doctor at either zero cost to themselves – in the case of a doctor who 'bulk bills' – or at a significant discount to the true cost of the visit. The full cost of the doctor's visit must nonetheless be paid by somebody, and the cost not paid by the individuals themselves is paid by the government from taxation revenue.

Discuss whether you think a moral hazard element is involved in this situation and, if so, what might also be done – by way of government policy – to mitigate the consequences.

All of a sudden such US-based MBSs became very 'sub-prime', and they had to be very cheap for anyone to want them! This rapid and dramatic reduction in market value had very dramatic consequences for companies that had purchased large volumes of MBSs and were holding them in their portfolios. Any such company was in sudden, severe financial trouble, with loss of investor confidence and a rapidly falling share market price the result.

The financial cataclysm that was the year 2008

When this house of cards finally came down, it happened with devastating effect and rapidity. Here are some of the more significant events:[3]

- January 2008 – the US share market was formally declared a bear market (a loss of at least 20% in value).
- March 2008 – JP Morgan bank took over investment bank Bear Stearns (then the fifth-largest investment bank in the US!)
- June 2008 – in Australia, prominent investment bank, Babcock & Brown, began its eventual slide down to 1 per cent of its peak value and ultimate takeover. And while well-known investment bank Macquarie Bank would eventually survive the cataclysm, it nonetheless suffered a devastating impact on its share price and market reputation.
- July 2008 – the oil price peaked at US$147 a barrel (by December 2008, just five months later, it was back down below US$40!)
- September 2008 – a month that may well go down in history as the most financially catastrophic of all time. In the space of a couple of weeks:
  - the UK's fifth-largest home lender, Northern Rock, suffered a bank run and was eventually taken over by the UK government
  - in the US, the federal government took control of the country's two home mortgage giants, Fannie Mae and Freddie Mac, and also took control of one of the world's largest insurance giants, AIG
  - investment bank Lehman Brothers failed (after the US Treasury Department refused to bail it out)
  - Bank of America took over investment bank Merrill Lynch
  - the two remaining giant US investment banks, Goldman Sachs and Morgan Stanley, announced they would seek to become commercial banks and be subject to Fed regulatory oversight
  - the US Treasury announced the US$700 billion **Troubled Asset Relief Program (TARP)**, with the aim of trying to bail out other troubled financial companies.
- October 2008 – the Australian government announced, somewhat controversially, that it was 'guaranteeing bank deposits' of the major banks. It also announced that a fiscal stimulus package of $10.4 billion was being rolled out. A much larger stimulus package was to follow in early 2009.

Troubled Asset Relief Program (TARP) Announced in October 2008 by the US Treasury announced, costing US$700 billion; aimed at trying to bail out a range of troubled financial companies whose problems it deemed to be relatively short-term in nature.

[3] For a detailed timeline of the GFEC, the Federal Reserve Bank of St. Louis has a webpage dedicated to the events starting in February of 2007. See https://www.stlouisfed.org/financial-crisis/full-timeline, accessed 29 May 2018.

- By November 2008 – all major central banks around the world were rapidly cutting interest rates in a bid to avert or ameliorate recessionary forces. The International Monetary Fund (IMF) announced the world was in the grip of a 'simultaneous recession across the developed world (the first since WWII)'. And most of the world's share markets were trading at less than half their peaks reached in late 2007. Australia's share market reached its nadir in early 2009 (ASX200, Feb. 2009 = 3344) and was less than half its pre-GFEC peak (Oct. 2007 = 6754).
- By December 2008 – the US central bank, the Fed, had cut its key interest rate to 0.5 per cent; The UK central bank, the Bank of England, had cut its key interest rate to 0.5 per cent, the lowest since it commenced in 1694; the Bank of Canada similarly cut its key rate to 0.5 per cent, the lowest in its history. In Australia, the RBA eventually cut the cash rate to a similarly historic low of 3 per cent (from 7.25 per cent) by April 2009, and, for the first time since December 2000, Australian quarterly real GDP contracted in the December quarter by 0.7 per cent.

All of the above made for a truly remarkable set of financial/economic events in the year 2008! It really was quite a year.

In the aftermath, one consequence has been a worldwide push towards more regulated financial markets to curtail the use of exotic derivative financial products, such as MBSs. In particular, the reader is referred to the Third Basel Accord, commonly referred to as Basel III, agreed by the Basel Committee on Banking Supervision.[4]

Economics and ethics **Conflict of interest among economists and advisers**

As noted in the text, the widespread adoption of a range of derivative financial products was a very significant cause of the GFEC. On this, some commentators believe a number of influential academic economists had significant conflicts of interest in the advice and analysis they provided in relation to the potential risks to the financial system of such derivative products. One such prominent commentator was Charles Ferguson, the producer and director of the documentary, *Inside Job*.[1]

More generally, Ferguson argues that senior economic advisers to the US government have increasingly been involved in serious conflicts of interest and that this is a growing trend in the US. He further argues the academic economics profession in the US has been penetrated by special interests, particularly the financial services industry. He is also critical of the so-called revolving door in the US whereby academic economists regularly move among industry, government and academia.

Ferguson is concerned that many prominent academic economists, whom the public would hope would provide balanced and objective analyses of a wide range of economic matters pertaining to individual companies or government policies, are often also paid consultancy fees by companies to be involved in regulatory proceedings. They may also serve on related company boards of directors or otherwise act as advisers to such companies. Ferguson is very concerned that those payments and the conflicts of interest they generate could easily colour the advice/testimony provided.

In his film, Ferguson focuses on several very well-known academic economists in the years leading up to the unfolding of the GFEC and points to their involvement in major financial services companies. In particular, he questions whether their public positions concerning financial regulatory matters and the pros and cons of emerging derivative products could have been adequately independent of their involvements with such companies.

Ferguson's view is not unchallenged, but it is nonetheless vital for the academic economics profession to consider it seriously. At the very least, in the same way, that ethics plays quite an essential role in the teaching of, for example, law, accounting, marketing and so on, perhaps in the future, it also needs to be more prominent in the teaching of economics.

[1] **This box draws significantly from the article 'Why economists are part of the problem', by Charles Ferguson, published by Reuters as part of the *Great Debate Series*, 12 October 2010.**

[4] A committee of banking supervisory authorities created under the auspices of the world's central banks.

Another consequence was that some of the expansionary fiscal policies taken around the world – along with other contributing factors – produced a set of negative second-round effects resulting from financial market perceptions of significantly increased risk of some governments within the Eurozone area – particularly those of Portugal, Ireland, Greece, Spain and Italy – defaulting on their loans. This issue is also discussed in greater detail in the final section of this chapter.

Several commentators believe there was another significant ingredient in the mix of factors giving rise to the GFEC, and that was the insufficient vigilance and professionalism displayed by prominent financial and economic consultancy firms, credit rating agencies (like Moody's, S&P and Fitch), financial regulatory agencies (like the US Securities Exchange Commission), and other major public sector institutions (like the US Fed) in the years leading up to the crisis. Some commentators even lay a portion of the blame at the door of the academic economist fraternity; this issue is discussed in the 'Economics and ethics' box that follows.

In summary

- The **global financial and economic crisis (GFEC)** originated in the US home mortgage market. Low-quality **mortgage-backed securities (MBSs)** were created and sold to international institutional investors at much higher prices than their inherent risks warranted. Any financial institution that had bought large volumes of these MBSs found themselves in very serious financial trouble. As a result of the resulting financial crisis, the world went into simultaneous recession in the developed countries.

2 The economic consequences of the GFEC

After explaining the 'how' and the 'why' of the GFEC, it is important to gain some understanding of the economic consequences. First, we will investigate the actual precise dating of the international recession of 2008–09 and whether it was predictable in some sense. Then we will look at the impact of the recession on economic output in selected countries around the world. Finally, we will outline the effects the recession had on unemployment and inflation.

Dating and forecasting the international recession of 2008–09[5]

In the US, the dates of the peaks and troughs in the US business cycle are officially determined by the Business Cycle Dating Panel under the auspices of the National Bureau of Economic Research (NBER). For other countries, using the NBER methods, the Economic Cycle Research Institute (ECRI) in New York City determines and maintains dates.

Exhibit 19.2 provides the dates for the US, Japan, Germany, the UK (the world's 20th-century locomotor countries), China and India (emerging dynamo countries of the 21st century) for the last 40 years. From the coincident indexes (a coincident index is a single aggregate measure of the economic activity in each country) of these six countries, a 'world coincident index' of economic activity has also been constructed, and its peaks and troughs are also provided in the exhibit. At least three things are of interest from this exhibit.

[5] This section is based on a recent journal article by Banerji, Layton and Achutan (2012), viz., A. Banerji, A. P. Layton and L. Achutan, 'Dating the world business cycle', *Applied Economics*, Vol. 44, 2012, pp. 2051–63.

Exhibit 19.2 Six countries' business cycle dates: US, Japan, Germany, UK, China and India

Peak/ trough	US	Japan	Germany	UK	China	India	World Coincident Index	World Leading Index
P						6/72		
T						5/73		
P	11/73	11/73	8/73	9/74		11/73	11/1973	
T	3/75	2/75	7/75	8/75		2/75	05/1975	12/1974
P	1/80		1/80	6/79		4/79	02/1980	12/1978
T	7/80					3/80		04/1980
P	7/81							
T	11/82		10/82	5/81			10/1982	
P					8/88			12/1989
T					12/89			
P	7/90		1/91	5/90		3/91		
T	3/91			3/92		9/91		12/1990
P		4/92						
T		2/94	4/94					
P						5/96		
T						11/96		
P		3/97						
T		7/99						
P	3/01	8/00	1/01				02/2001	04/2000
T	11/01	4/03	8/03				12/2001	01/2001
P	12/07	2/08	4/08	5/08			04/2008	08/2007
T	6/09	3/09	1/09	1/10			06/2009	11/2008

Source: A. Banerji, A. P. Layton and L. Achutan, 'Dating the world business cycle', *Applied Economics*, Vol. 44, 2012, pp. 2051–63.

First, looking at the country columns and the last two rows, the US was the first major country to enter recession and did so in January 2008 (its peak date was December 2007). The other three major locomotor countries joined it within five months so that, by mid-2008, all four major developed countries were simultaneously in recession in an event triggered by the GFEC. This was a relatively rare phenomenon and had last happened in the first oil price crisis in the early 1970s.

World Coincident Index An index of the current state of 'world' economic activity constructed from the coincident indexes for the US, UK, Japan, Germany, India and China maintained by the Economic Cycle Research Institute.

Second, looking at the column headed '**World Coincident Index**', the research suggests there have been four synchronised international recessions over the 40 years, the most recent being dated from April 2008 to June 2009, this last coming about as a result of the impacts of the GFEC. The period from mid-1990 through to late 1991/early 1992 was also a period of international economic weakness but is probably better characterised as a period of rolling national recessions among the major economies rather than a period of the synchronised international recession.

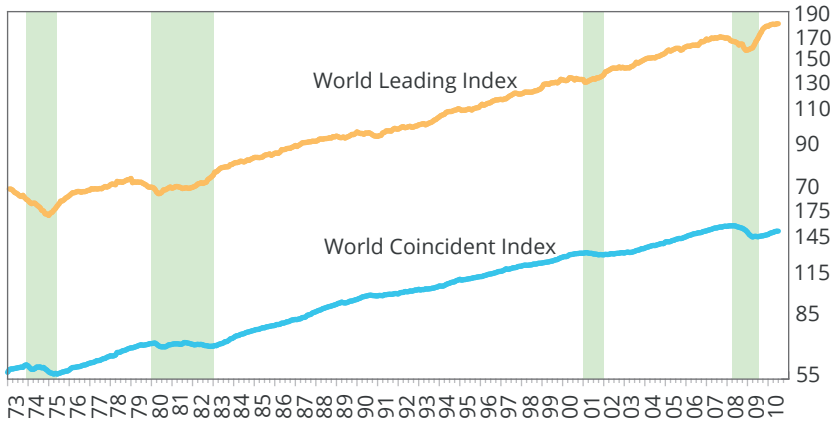
Third, looking at the last column, the '**World Leading Index**' constructed from the leading indexes of 19 countries covered by ECRI (a leading economic index is constructed to forecast a phase change in the business cycle within a six-month time horizon) appears to have anticipated the most recent international recession very well, turning into a downswing in August 2007, some eight months ahead of the onset of the international

recession of 2008–09. It also troughed in November 2008, again seven months ahead of the international recession's trough in June 2009. In addition, it should be noted that this leading index also anticipated the period of international economic weakness in the early 1990s discussed in the paragraph above.

Exhibit 19.3 provides a graphical representation of the 'world coincident index' and the 'world leading index' and this provides a clear pictorial presentation of the reliable predictive ability of the leading index.

World Leading Index Constructed from the leading indexes of 19 countries covered by the Economic Cycle Research Institute (ECRI).

Exhibit 19.3 World coincident index and world (ECRI 19-country) leading index



Note: Shaded areas represent world recessions.

Source: A. Banerji, A. P. Layton and L. Achutan, 'Dating the world business cycle', *Applied Economics*, Vol. 44, 2012, pp. 2051–63.

The GFEC impact on the world's economic output

In **Chapter 16** in the 'Global perspectives' box we outlined the impact that the Great Depression had on US economic output and showed that real GDP in 1933 was an extraordinary 30 per cent lower than just four years before in 1929. In the Global perspectives box below we focus on the impact the GFEC had on the economic output of some of the world's major economies. As you will see by analysing the reduced output in the US, by comparison, the Great Depression was genuinely catastrophic in its economic consequences compared to the economic impacts of the GFEC.

Applicable concept: the impact of the GFEC on economic output

Global perspective

The effect of the GFEC on the economic output of the US, the UK, Germany and Japan

The US economy officially reached its business cycle peak in December 2007. The US share market also reached its peak in October of that year. As outlined above, during 2008, global financial problems became increasingly evident and the world's other major share markets began to recede from previous peaks. By September 2008 all of the world's major share markets were tumbling.

With the tumbling share markets, plummeting business and consumer confidence, and severe world liquidity problems resulting from ailing financial institutions and evaporating confidence in them, the crisis, which began as a financial one, quickly transformed into an economic one. Many countries around the world went into recession during 2008 and 2009.

In fact, in only the second occasion since the Second World War, all four of the world's major 20th-century powerhouse economies – the US, the UK, Japan and Germany – entered into a period of simultaneous business cycle recessions, and did so within a six-month period during the calendar year 2008 (the only previous post-war occasion was the synchronised international recession of the early 1970s that resulted from the first world oil price shock). This in turn precipitated many other smaller dependent economies around the world to also go into recession. Fortunately, for reasons that will be explained later in this chapter, the world managed to avert a repeat of a 1930s-style full-scale Great Depression.

For now, consider **Exhibit 19.4** where movements in GDP for Germany, Japan, the UK and the US for the calendar years 2006–12 are provided in index number form (with base = 100 in 2006 for each country). Using these data, the impact of the global financial and economic crisis (GFEC) of 2008–09 is stark! The GDP growth rates for these countries between 2006 and 2007 were 3.4, 2.1, 3.6 and 1.8 per cent, respectively. Extrapolating these growth rates for the following two years would have seen the economies of Germany, Japan, the UK and the US grow by 6.8, 4.2, 7.2 and 3.6 per cent respectively from 2007 to 2009. However, due to the GFEC, the economies of these countries actually *shrank*.

Exhibit 19.4 **GDP index numbers for Germany, Japan, the UK and the US, 2006–09**

Country	2006	2007	2008	2009	2010	2011	2012
Germany	100.0	103.4	104.2	98.9	102.7	106.2	107.2
Japan	100.0	102.1	101.0	95.4	99.9	99.5	101.2
UK	100.0	103.6	102.6	98.6	100.3	101.3	101.6
US	100.0	101.8	101.5	98.7	101.2	102.8	105.1

Base = 100 in 2006 for each country

Source: Economagic.com, 'Country Tables' and author calculations, http://economagic.com, base data accessed 5 April 2017.

What do you think?

1 What was the percentage reduction in each country's annual GDP in 2009 compared with its GDP in 2007? *Hint:* refer back to **Chapter 11** on how to use index numbers to calculate percentage changes.

The contractions are quite significant, but compared with the Great Depression were all much less pronounced than the 30 per cent experienced between 1929 and 1933 in the US. Also all economies were producing more GDP in each of 2010, 2011 and 2012 than in 2009 so the economic downturns were much less protracted in the GFEC than in the Great Depression in the US, where output continued to shrink each year for four consecutive years.

2 Using the index numbers, which countries had regained their previous peak by 2012 and which had not? Of those that had, in what year did each regain its previous peak?

The impact of the crisis on inflation and unemployment

As was explained in the previous sub-section, the GFEC produced a relatively rare event, a synchronised international recession that saw most developed countries around the world in almost simultaneous recession. As outlined above, research conducted by one of the authors (Layton) and his colleagues at the Economic Cycle Research Institute (ECRI) dates the onset of this international recession as April 2008 and its conclusion as June 2009.

A recession is defined as being over as soon as the level of aggregate economic activity begins to expand again after the recession's trough is plumbed. However, this does not mean that a country's unemployment

rate immediately starts to reduce at that time. If the economy's growth is slow, unemployment can continue to increase because an insufficient number of jobs are being created concerning the growth in the workforce. Similarly, inflation pressure may remain quite muted in the early years of recovery if the pace of that recovery is relatively moderate.

The impact of the GFEC on inflation and unemployment in selected countries can be seen in **Exhibit 19.5**, where data for 2008–12 (2012 being three years since the dated end of the 2008–09 synchronised international recession) are presented. Concentrating on the two years 2008 and 2009 for the moment, as can be seen from the exhibit, the significantly reduced aggregate demand for goods and services coming about as a result of the GFEC exerted downward pressure on rates of inflation around the world (this is the obverse of the 'demand-pull' inflation pressure described earlier in the text). Only in India did the rate of inflation not reduce from 2008 to 2009.

Exhibit 19.5 **Inflation and unemployment during the GFEC: selected countries**

Country	Inflation					Unemployment rates				
	2012	2011	2010	2009	2008	2012	2011	2010	2009	2008
Australia	1.8	3.3	2.8	1.8	4.4	5.2	5.1	5.2	5.6	4.3
Brazil	5.4	6.6	5.0	4.9	5.7	5.5	6.0	6.8	8.1	7.9
China	2.6	5.4	3.3	–0.7	5.9	4.1	4.1	4.2	4.3	4.1
France	2.0	2.1	1.5	0.1	2.8	9.8	9.2	9.3	9.1	7.4
Germany	2.0	2.1	1.1	0.3	2.6	6.8	7.1	7.7	8.1	7.6
Greece	1.5	3.3	4.7	1.2	4.2	24.7	18.1	12.9	9.8	7.9
India	9.3	8.9	12.0	10.9	8.4	5.7	5.6	5.6	5.6	5.3
Indonesia	4.3	5.4	5.1	4.8	9.8	6.2	6.7	7.3	8.0	8.4
Japan	–0.0	–0.3	–0.7	–1.3	1.4	4.6	4.8	5.3	5.3	4.2
Malaysia	1.7	3.2	1.7	0.6	5.4	3.0	3.0	3.25	3.7	3.3
New Zealand	0.9	4.4	2.3	2.1	4.0	6.9	6.5	6.5	6.1	4.2
Russia	5.1	8.4	6.8	11.7	14.1	5.5	6.5	7.5	8.4	6.4
South Korea	2.2	4.0	3.0	2.8	4.7	3.3	3.5	3.8	3.8	3.2
United Kingdom	2.8	4.5	3.3	2.2	3.6	7.8	8.0	7.8	7.6	5.6
United States	2.1	3.2	1.6	–0.4	3.8	8.2	9.1	9.8	9.4	5.8

Source: Federal Reserve Bank of St Louis (FRED) Economic Database, Various Country Tables, and world bank data on unemployment rates, accessed 1 November 2020.

In other words, most countries listed experienced disinflation in 2009 compared with 2008 – the rate of inflation reduced between the two years. And, in a few countries such was the reduction in demand for goods and services that the inflation rate was negative between the two years, meaning the average price levels fell from 2008 to 2009. In other words, some countries actually experienced deflation as defined in **Chapter 13**. This occurred, for example, in Japan, China and the US (and in France and Germany the annual inflation rate in 2009 was barely above zero).

The reduced demand, in turn, caused production to fall significantly in many countries and employment growth slowed below the growth in countries' workforces, with the result of increasing unemployment. In some cases, for example, in the US, the increase in unemployment was very quick and very substantial. In that country alone a rough estimate of the increase in the number of unemployed from one year to the next as a result of the GFEC would be around 5 million people!

Similarly, a rough estimate of the total increase in the number of unemployed across the 15 countries listed would undoubtedly be well above 10 million people. And, of course, there are many other countries in the world that were significantly affected by the GFEC. So the total number of people worldwide thrown out of work on account of the crisis – and perhaps for a very long time –would have been very large indeed. The economic and social cost of this increased unemployment is likely to have been genuinely enormous.

Before moving on, a few other things are of interest from **Exhibit 19.5** by looking at the data for the later years. Very notable is Japan that experienced continually falling prices from 2009 right through till 2012. From 2013 (not shown) it began to experience modestly rising prices again. Another important aspect is the escalating unemployment rate in Greece. Unemployment in that country climbed to just under 30 per cent and among young adults climbed to above 50 per cent. More will be said about Greece's situation later in this chapter. Finally, unemployment in the US continued to remain persistently high for quite a few years after the GFEC. It was for this reason that the Fed resorted to the unconventional monetary policy tool of quantitative easing (QE) in 2013. More will be said on this in the next section.

In summary

- The GFEC of 2008–09 produced a synchronised international recession – one of only four over the last half -century – which has been dated as lasting from April 2008 to June 2009.
- The world's 20th-century locomotor countries of the US, the UK, Germany and Japan all went into recession within six months of each other and, as a result, output in these countries dropped significantly.
- Unemployment increased significantly and disinflation occurred in most countries.

3 Macroeconomic policy responses to the crisis

In this section, you will gain insight into the economic policy responses around the world to the crisis, and obtain an understanding of policy lessons learnt from the crisis. First, we shall discuss monetary responses by the world's major central banks and then we will move to fiscal policy actions.

Monetary policy during the GFEC crisis

In the 'Global perspectives' box in **Chapter 16**, it was suggested that the Fed in the US – and central banks elsewhere around the world – did not employ the appropriate policies in the face of the unfolding 1929–30 financial crisis. Banks were allowed to fail, the money supply shrank rapidly, liquidity and loan funds dried up overnight, very short-sighted trade protectionism sprang up quickly around the world ('beggar thy neighbour' trade policies), and the US and the world spiralled rapidly into what became the Great Depression.

Fortunately, much was learnt from that crisis, and it was quite a different story in 2008–09 when the GFEC began to unfold. When the full horrific extent of the sub-prime financial crisis was realised (refer back to the previous section for a discussion of the roots of the GFEC), the Fed and the US Treasury moved quickly to shore up the US financial system. While some fundamentally unsound financial institutions were allowed to fail (most significantly, and famously, Lehman Brothers investment bank), those judged to have only short-run liquidity problems were supported by the Fed. The financially huge Troubled Asset Relief Program (TARP) was quickly passed by the US Congress and was designed to re-liquify banks and other financial institutions.

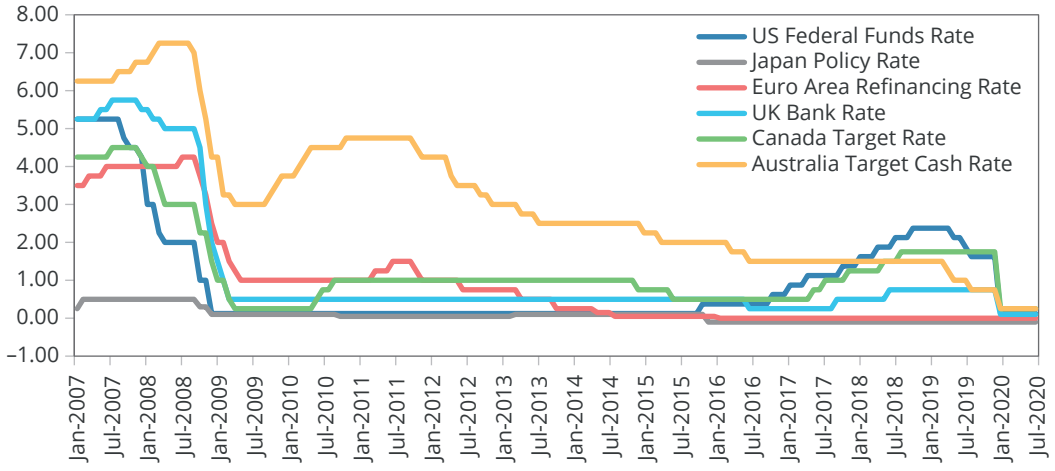
The role played by interest rate reductions by the world's major central banks

Whilst the above actions were important in stemming contagion among the world's major financial institutions, of utmost importance was the September 2007 pre-emptory action by the US Fed to begin a dramatic monetary policy easing, which saw its key monetary policy interest rate instrument, the **Federal Funds Rate (FFR)**, drop quickly from 5.25 per cent (in August 2007) to just 0.13 per cent in December 2008, and it remained at that level – actually maintained at zero to 0.25 per cent – right through to December 2015, when it was increased by 0.25 per cent. This was the first increase in the FFR by the Fed in almost 10 years!

Federal Funds Rate (FFR) The key monetary policy interest rate instrument of the Federal Reserve System, or Fed, the US central bank. It is the equivalent to the 'overnight cash rate' in Australia.

Refer to **Exhibit 19.6** for a graph of the course of a few central bank monetary policy interest rate instruments in the immediate lead-up to the GFEC through to 2020. The monetary policy responses during 2019 and 2020 will be discussed in the section below on COVID-19 pandemic.

Exhibit 19.6 Central bank monetary policy instruments: monthly 2007–20



Data source: Reserve Bank of Australia

We now know the US business cycle peaked in December 2007, so the relatively early, decisive and substantial action by the Fed in September was exactly what was needed – although with perfect hindsight one might suggest that a move in June would have been the most optimal timing. However, given the ever-present policy information and determination lags (refer back to **Chapter 16** for discussion of these), the Fed's move in September was in practice pretty good.

As an aside – but an important one – it is fortunate indeed that, after an earlier easing cycle, the US Fed elected to 'reload the cannon' and get the US FFR setting back to more normal levels in 2004 and 2005 while the US economy was growing briskly, even though there had been no sign of any breakout in inflation at that time. Many criticised the Fed for doing so, but fortunately, in 2007–08 it was able to react quickly and substantially to give the US economy a crucially needed monetary policy boost.

Japan, on the other hand, was not so fortunate. The Bank of Japan's 'Target Call Rate' had been left very low at just 0.5 per cent in the year leading up to the GFEC and so Japan's central bank had little interest rate stimulus ammunition to use. It reduced its target interest rate to 0.1 per cent in late 2008 (and then to 0.05 per cent in October 2010), but it is unlikely this minuscule interest rate reduction had any significant countercyclical impact on the Japanese economy.

You're the economist **Can nominal interest rates ever go below zero?**

In **Chapter 13** you learnt that real interest rates – being nominal interest rates minus the rate of inflation – could go negative. But can nominal interest rates ever be negative? This would mean that the lender, instead of receiving interest from the borrower, actually pays the borrower for the privilege of lending to them!

In the UK, the Bank of England began easing its official bank rate from 5.75 per cent in November 2007 – but less quickly than the US Fed – and still had it set at 5 per cent in September 2008. But then, in the space of just six months, it dropped it to just 0.5 per cent – a 300-year low!

In Europe, in the Eurozone countries, the ECB only began easing its key interest rate setting (its Repo rate) relatively late in the crisis, in October 2008. In fact, it somewhat curiously tightened in August 2008 by 0.25 per cent to 4.25 per cent only to begin aggressively cutting it again from October, and had reduced it to 1 per cent by May 2009. Ultimately, given the continuing fragility of the Eurozone economic recovery, the ECB eventually had cut it to zero in March 2016!

The actions throughout the GFEC by the RBA are extremely interesting. Australia had been at a very different point in its business cycle when the crisis hit in 2008. Local inflation was the major concern of the RBA throughout 2007 and well into 2008, and so it had raised its cash rate from 6.25 per cent to 7.25 per cent from July 2007 to March 2008, and kept it at that level until August 2008. In September 2008 it then began a very dramatic easing of monetary conditions, which saw the cash rate down at just 3 per cent by April 2009, its lowest level during the crisis and its lowest level in decades.

Once it became clear the Australian economy was not going to slide into recession, the RBA began raising the cash rate again in October 2009, and had raised it back up to 4.75 per cent by November 2010. However, the perceived balance of risks to the economy – domestically and internationally – were such that the RBA had again eased the cash rate back down to 1.5 per cent by August 2016.

Furthermore, on top of these dramatic reductions in their usual monetary policy interest rate instruments, as was mentioned in **Chapter 16**, the central banks of the US, Japan, the UK and the Eurozone also all vigorously pursued large-scale QE to shore up their fragile economies. This is discussed next.

The use of quantitative easing by central banks

As was explained in **Chapter 16**, quantitative easing (QE) is considered an unconventional monetary policy tool and consists of a central bank expanding its balance sheet by buying long-dated government and corporate bonds from the financial sector. In paying for these purchases the central bank expands the Monetary Base. The action may have an impact on the yield of the long-dated securities purchased by the central bank but the primary goal of the strategy is to significantly expand the available liquidity in the financial system. The aim of QE is to stimulate economic growth and inflation and, in recent times, has been pursued by central banks when they had reduced their interest rate instruments to levels where further reductions may have been expected to have negligible impact on economic activity (on this refer back to **Exhibit 19.6** for the path of the interest rate instruments of the world's major central banks).

For the purposes of this QE discussion we will focus on the US, the UK and Europe. The US Fed began its QE program in November 2008 when it began buying MBSs from the private sector. There were three stages of QE in the US, known as QE1, QE2 and QE3, with QE3 being wound up in October 2014. Shortly, in the 'Analyse the issue' box, you will analyse the extent to which QE in the US impacted on the US Monetary Base.

In the UK, the Bank of England began its QE program in March 2009 and by July 2012 had increased the UK Monetary Base by £375 billion. Then, in August 2016, it announced it would buy an additional £70 billion on account of uncertainty brought about by Brexit. And in Europe, the ECB commenced a QE program in May 2009, but in January 2015 it was ramped up very substantially when it was announced that from March that year the

ECB would purchase €60 billion in government bonds per month until at least September 2016, meaning the intention at that time was to expand its balance sheet by at least €1.1 trillion. Later the ECB advised the program would also be expanded to include corporate bonds. Then, in June 2017, it was announced the purchases would continue at the same rate at least until the end of 2017, or beyond if necessary. The buying would continue until the ECB saw a sustained rise in inflation to a target of below, but close to, 2 per cent.

As explained in **Chapter 16**, the purpose of all this QE is to use growth in the money supply – namely the Monetary Base – to stimulate growth in nominal GDP. In the Analyse the issue box below we have a look at what has actually occurred to the Monetary Base of the US as a result of that country's QE program.

Applicable concepts: quantitative easing and Monetary Base growth

Analyse the issue

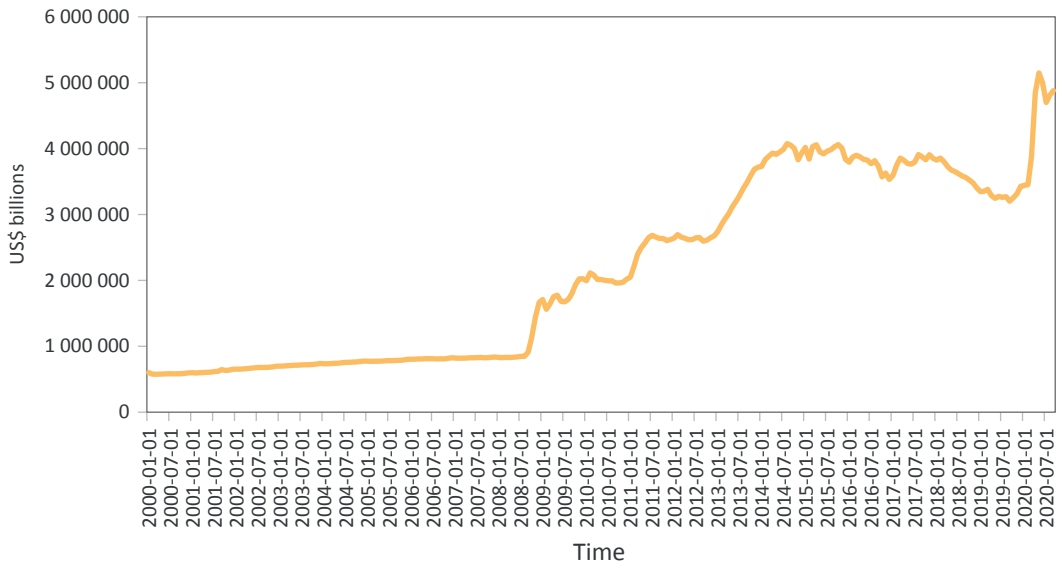
The impact of QE on the US Monetary Base growth

As stated above, the idea behind quantitative easing (QE) is that the central bank deliberately expands its balance sheet by buying government, and sometimes corporate, bonds with the purpose being to increase the country's Monetary Base. The policy draws on the quantity theory of money – explained in **Chapter 16** – and the idea is that nominal GDP growth will be stimulated by the action.

Operationally, as also explained in **Chapter 16**, the mechanism is that the central bank pays for its purchases by crediting the sellers of the bonds – namely banks – with newly created exchange settlement deposits (or 'cash'). The expectation on the part of the central bank is that those banks would then use the newly created deposits to fund new lending to the private sector to stimulate economic activity. So, just how much has QE in the US acted to expand that country's Monetary Base?

Refer to **Exhibit 19.7** below and answer the following questions.

Exhibit 19.7 **Growth in US Monetary Base, 2000–20 (monthly, in billion $US)**



Source: Board of Governors of the Federal Reserve System (US), Monetary Base; Total [BOGMBASE], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/BOGMBASE

What do you think?

1 QE began in the US in November 2008 and this can be seen quite starkly in the figure. What approximately was the level of US Monetary Base at that time?
2 The three stages of QE in the US are quite evident from the figure when QE was finally wound up in October 2014. Approximately what was the value of US Monetary Base at that time? What then was the per cent increase in US Monetary Base over this six-year period?
3 What might be the explanation for the reduction in US Monetary Base since December 2015 (*Hint*: recall the discussion in **Chapter 16** on the method by which the central bank influences its interest rate instrument)?

Has QE been successful? Whilst the opinion is certainly not unanimous many analysts believe it has been, especially in the US whose economic performance has continued to improve during 2015 and 2016, and into 2017. Some commentators remain concerned about the possibility of inflation becoming a problem on account of QE but there was certainly no evidence of this by the end of 2017. There is also the issue of what the impact might be on the US economy when the Fed begins the process of reducing its bond holdings, which will have the effect of reducing the US Monetary Base. We will soon know the answer to this as the Fed had begun this process during 2017 and was intending to continue reducing its balance sheet throughout 2018 and 2019.

And so, what impact did all this international central bank monetary policy easing have? In short, a great deal! While the GFEC was a painful recessionary episode for most of the world's economies – Australia, having averted a recession, being one notable exception – there is no doubt this action from the world's major central banks ensured a much reduced severity and duration in the international recession that occurred in 2008–09.

The international recession is now believed to have ended in June 2009 (having commenced most likely in April 2008), with Germany, Japan, the US and the UK commencing an expansionary phase of their business cycles in January, March and June 2009, and January 2010, respectively.[6] Countercyclical monetary policy was certainly a very significant factor in ameliorating the severity of the economic impact of the GFEC.

The role of Australia's flexible exchange rate pre- and post-GFEC

While in Australia we have a flexible and floating exchange rate (since 1983), and therefore it is not an instrument of monetary policy, the exchange rate was certainly an important factor in assisting Australia to avoid a recession as a result of the GFEC.

Since it was floated in 1983, as is evident from **Exhibit 18.9**, the AUD exchange rate has fluctuated across quite a wide range. Indeed, to a very great extent the flexibility of the AUD – and its ability to adjust quickly, especially downwards – has acted as a very useful buffer for the Australian economy against a number of significant and challenging external economic shocks that have occurred over the last 25 years. The most recent example of this was the GFEC of 2008–09. In the months leading up to this crisis the AUD was trading at over US$0.90 and peaked at US$0.98 in July 2008. At this same time, in trade weighted index (TWI) terms it was valued in the low 70s range, and against the Japanese yen it was trading at a little over 100 yen. Within a few months, by late 2008, it had depreciated to around US$0.65, around 60 yen, and had declined to around 54 on a TWI basis.

Using the TWI as the broader indicator of the international value of the AUD, this represented more than a 25 per cent depreciation in a very short period of time. It continued to trade at these much lower levels (US$0.60–0.70; 60–70 yen; and 50–60 on a TWI basis) through until mid-2009. From then it began a fairly steady appreciation back to the levels mentioned above.

[6] However, the UK and Japanese expansions were both relatively short-lived, with both falling back into recession in August 2010. The UK again entered a lasting expansion in February 2012, whilst Japan continued to struggle with a continuing very tepid economy.

As has been emphasised on a number of occasions earlier in this book, while the depreciated AUD might irritate people wanting to buy imported products or travel overseas, it is a great boon to Australian exporters and import-competing companies as it makes Australian products sold overseas more profitable to Australian exporters, converting overseas export revenue into more AUDs. This is good for Australian production and employment, not only in those directly affected companies, but also throughout the rest of the economy due to the linkages that these sectors have with other parts of the economy.

However, more recently, by early 2013 the AUD had appreciated quite significantly from the levels of early 2009 (on a TWI basis, the AUD was trading just below 80 in early 2013). This was almost certainly due to relatively higher real interest rates available in Australia (along with Australia's AAA international credit rating) and was, at least for a couple of the more recent years in the period, in contradiction to moderating international commodity prices – historically, one of the main drivers of the value of the AUD. This presented considerable challenges to many exporters – like education and tourism service exporters for instance, as well as a range of manufacturing exporters – and a number of import-competing industries in the Australian economy.

Fortunately, by late 2015, the AUD had depreciated quite significantly from the early 2013 highs, and was probably regarded then by the RBA to be trading around 'fair value' in terms of Australia's economic fundamentals (particularly international commodity prices). While there had been a modest appreciation in the AUD by 2017 it would seem that the lower AUD will continue long enough to provide significant long-lasting benefit to Australian businesses.

The other major arm of macroeconomic policy, fiscal policy, also played a significant role in dealing with the GFEC, and this is discussed next.

The role of fiscal policy during the GFEC

As explained above, the world's major central banks quickly went into gear to provide support to the financial systems of their own countries as well as the international financial system as a whole, and also speedily put in place unprecedentedly low interest rates via extremely expansionary monetary policy, all in order to prevent the GFEC becoming another Great Depression.

Many of the world's governments also actively engaged in discretionary expansionary fiscal policy in order to stimulate aggregate demand to help ameliorate the severity of the economic downturn. Some, like Australia, were able to do this from a position of having a strong government balance sheet (i.e., low existing government debt) from which to operate. Others – notably in Europe – were less fortunate and had much weaker balance sheets (i.e., high existing government debt levels) but nonetheless provided the same sort of fiscal stimulus to their economies. As a result, some of these countries had very significant subsequent sovereign debt problems (see below). For the purposes of the discussion here we will focus on the Australian case, which captures the essence of how many governments reacted to the crisis.

As mentioned above, the Australian federal government had, from 1996, reduced its gross debt from $111 billion (about 21% of annual nominal GDP at that time) to just $53 billion by June 2007 (about 5% of GDP at that time). Other countries had much higher government debt levels going into the GFEC, with average net government debt of Organisation for Economic Co-operation and Development (OECD) countries around 40 per cent of GDP. Australia was therefore able to weather the impact that the GFEC had on its federal government's budget – reduced tax revenue from lower economic activity and higher expenditure from higher unemployment – quite well compared with other countries. Furthermore, it was able to deploy a massive short-term fiscal stimulus to aggregate demand in the Australian economy with the aim of averting a recession.

In late 2008, in response to the rapidly deteriorating world economic situation, the Australian government took action to support the economy through announcing additional government expenditure of $42 billion, a fiscal stimulus of about 3 per cent of 2009 GDP. The expenditure consisted of about $29 billion in spending on schools, housing, energy efficiency initiatives, community infrastructure and roads, such spending to occur over the course of 2009 and 2010.

In addition, the government announced that from early March 2009 it would provide about $13 billion in various one-off bonus payments to different categories of Australians. The idea was to deliver an immediate short-term stimulus to demand in the economy to support growth before the more medium-term spending, and RBA-engineered lower interest rates, were able to have a significant effect on the economy.

Using the *AD–AS* framework, these actions can be understood by referring back to **Exhibit 17.2**. As a result of the onset of the GFEC the Australian economy could be represented as being in the process of shifting to a point like E_1 where unemployment was growing due to insufficient aggregate demand. The government's stimulus package was designed to increase aggregate demand back to something like what was required to maintain full employment and would be represented in the exhibit by the *AD* curve shifting outward from AD_1 towards AD_2. The monetary policy easing by the RBA, which was also occurring from late 2008, was also seeking to stimulate aggregate demand and would also be represented in the *AD–AS* framework by an outward shifting *AD* curve.

What was the result? The combined actions of Australia's macroeconomic policymakers – the fiscal policy action by the government described here, along with the pronounced and rapid reductions in interest rates brought about by the rapid and sharp monetary policy easing by the RBA discussed above – quite probably did much to help Australia, virtually alone among the developed countries in the world, to avoid experiencing a GFEC-induced recession.

However, the fiscal action – which, in the opinion of a number of academic economists and other commentators may in hindsight have actually been overdone – certainly significantly impacted on the country's federal government budget balance and increased the level of government debt dramatically. After a string of budget deficits from 2008–09, government debt had risen to an estimated $320 billion in 2013–14, or up from 5 per cent of GDP to 20 per cent of GDP, representing a quite extraordinary quadrupling of the relative size of the debt in just five years! By 2016–17 it had risen further to around 29 per cent.

Referring back to **Exhibit 17.7**, the impact on the budget of the fiscal stimulus is very evident. In 2009–10 the deficit for that year was a quite massive 4.2 per cent of GDP. This followed a 2.1 per cent deficit in 2008–09 and was the largest budget deficit ever recorded, being even larger than the 4 per cent deficit that occurred as a result of Australia's very long and severe 1990–92 recession. Furthermore, while in the case of the 1990–92 recession the budget was returned to surplus within five years of its peak deficit, Australia's federal budget was projected in the 2017–18 Budget to remain in deficit for a total of at least 12 years from 2008–09 to 2019–20 inclusive! However, things changed remarkably with the onset of the COVID-19 pandemic, and it is forecasted the budget deficit would last until 2030–31. We have discussed this in detail in **Chapter 17** and will discuss further in the sections below.

In conclusion to this section, in the immediate aftermath of the crisis, what was clear was that the GFEC had left relatively long-lasting heightened perceptions of risk among economic players – both in financial markets and more broadly throughout the world's economies – so that many countries were still struggling to return to more normal rates of economic growth several years after 2009. There were also significant concerns held by many economists that the super-accommodative monetary policies by the world's major central banks – still in force through to 2017 – would eventually lead to another financial and economic crisis sometime in the not-too-distant future.

Finally, whilst Australia's government debt situation going into the crisis was quite low, in other countries this action came on the back of already high debt levels. This produced serious consequential problems for those countries in terms of the perceptions of the levels of risk associated with their country's government debt issuance. We turn to this in the next section.

In summary

- The severity of the GFEC was most certainly ameliorated by quick monetary policy easing actions of the world's major central banks.
- This monetary policy consisted first in interest rate reductions to historical lows, and subsequently using quantitative easing to dramatically increase the monetary bases of their respective countries.
- Fiscal policy was also deployed by the world's governments whereby government expenditure was increased and funded by issuing government debt to the private sector.

4 The European sovereign debt – and banking – crisis

European Sovereign Debt Crisis
Within the 19-country Eurozone area there arose in 2009 widespread fears that a particular five-country sub-group of countries – namely, Portugal, Ireland, Italy, Greece and Spain – may default on their government debt.

The **European Sovereign Debt Crisis** refers to a situation that began in 2009 – and most probably had largely passed by early 2014. Within the 19-country Eurozone area,[7] there arose in 2009 widespread fears that a particular five-country sub-group may default (i.e., not repay) on their government debt. This sent financial system tremors around Europe and beyond as many banks – both inside Europe and also around the rest of the world – had lent funds to the countries in question by purchasing billions of euros of their governments' bonds. This called into question the financial soundness of those banks, which in turn affected their ability to raise further funds from financial markets, as well as their willingness to lend.

As mentioned earlier in this chapter, the countries at the centre of the crisis were Portugal, Ireland, Italy, Greece and Spain (sometimes referred to by the unfortunate acronym 'PIIGS'). Of these, Portugal, Ireland and Greece are relatively small, accounting collectively for only around 5 per cent of Eurozone GDP. Nonetheless, a default by any of those countries was still of great concern due to fears this would lead to a contagion of falling confidence throughout the Eurozone area, possibly precipitating further economic stagnation for all, and perhaps resulting in the international community losing confidence in the euro currency itself, thus putting its future at risk.

The reasons put forward by analysts as to why these five countries found themselves in the debt predicament varied across countries, but included, for example, real estate bubbles, which led to questionable government bailouts of banks, financed by government debt; sustained and very significant debt-financed government spending of a non-capital goods kind; the worldwide cheap borrowing rates in the years leading up to the GFEC; the 'Great Recession' of 2008–09 and the economic pressure that it put on the countries in question; and the lack of a 'fiscal union' among the Eurozone countries to complement their currency union.

To gain an insight into the crisis, refer to **Exhibit 19.8**. What is depicted in the exhibit is the difference between the market-determined yield on 10-year government bonds for each of the five countries and the yield on German 10-year government bonds. Germany is the largest – by far – of the Eurozone economies and its bonds would usually be considered to be the least risky of any country's debt issuance in the Eurozone. The plotted differential therefore represents the financial community's perceptions of idiosyncratic changes in the riskiness of each country's government debt.

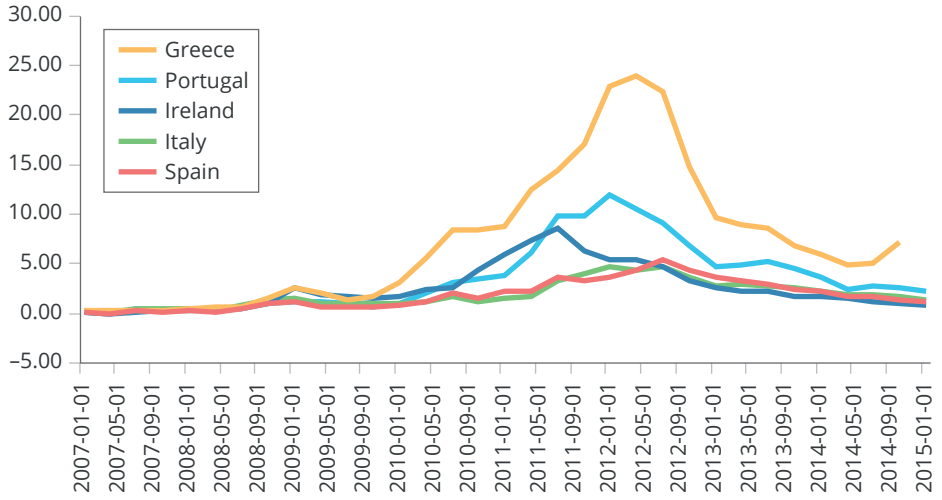
As can be seen, until late 2009, the yield differentials were really mostly quite modest for all five countries. After this, the increased financial market perceptions of the riskiness of each country's sovereign debt are quite evident. A few points are worthy of mention. First, from the figure it is quite clear that the market never became overly concerned about the riskiness of the sovereign debt of Italy and Spain.

In Italy's case, the differential reached its maximum of 4.7 per cent in January 2012. At this time, its 10-year bond yield was 6.5 per cent, high but quite manageable from a debt servicing point of view. In Spain's case, the differential reached its maximum of 5.4 per cent in July 2012. At this time, its 10-year bond yield was 6.8 per cent – again high but still quite manageable from a debt servicing point of view.

Greece, Ireland and Portugal were in quite a different situation. Of these, the least affected was Ireland. Its yield differential peaked at 8.6 per cent in July 2011. At that time, its 10-year bond yield was 10.9 per cent, an unsustainable debt servicing level over the longer term. Portugal's yield differential peaked at 12 per cent in January 2012 when its 10-year bond yield reached 13.8 per cent, again an unsustainable debt servicing level over the longer term. Finally, by far the most affected of the five countries was – and continues to be – Greece. Its yield differential peaked at a whopping 24 per cent in April 2012. At that time, its 10-year bond yield was 25.4 per cent, a completely unsustainable debt servicing level. Indeed, Greece's 10-year government bond yield was in double digits for three years!

[7] This group is part of the EU 'customs union' but in addition, they use the same common currency, the euro, established in 1999. They are therefore also part of a 'currency union'.

Exhibit 19.8 **10-year government bond yield differentials with Germany, monthly, 2007–15**



Source: Country tables in FRED® (Federal Reserve Bank of St Louis), https://fred.stlouisfed.org/categories/32264

The sovereign debt situation – particularly that of Greece, Portugal and Ireland – required significant action to be taken by the Eurozone authorities, the detail of which follows. The crisis was resolved by the coordinated actions of the so-called 'Troika' of the International Monetary Fund (IMF), the European Commission (EC) and the European Central Bank (ECB), along with significant actions by the European Banking Authority (EBA). The EC and the IMF together raised and distributed support funds to ailing countries in exchange for binding agreements to significantly restructure their public finances (*inter alia*, cutting unaffordable expenditure on public sector wages and transfer payments of various kinds, and raising taxes).

Meanwhile, the ECB – as well as lowering official interest rates to near zero to support all Eurozone countries – announced in 2012 that it would provide unlimited support to those countries involved in an EC/IMF bailout program and that it would do whatever was necessary to ensure the future of the euro.[8] Finally, the EBA worked with troubled Eurozone banks using the so-called 'European Financial Stability Facility' in an operation known as the 'European Capital Exercise' to recapitalise troubled European banks and raise their required holdings of equity capital so as to increase investors' confidence in them.[9, 10]

All of these measures worked in concert to eliminate the crisis, which, as previously mentioned, most likely had passed by the end of 2013 or early 2014. This is not to say that some countries – for example, Greece, and the small island country of Cyprus – were not still receiving external support, but the danger of a widespread collapse in confidence in the Eurozone seemed to have largely dissipated by early 2014. By then, widely accepted

8 This assurance was delivered in what has now become a quite famous address by the ECB's President, Mario Draghi, in July 2012.

9 The European Financial Stability Facility was a very large fund used to assist in recapitalising troubled Eurozone banks. It was similar in nature to the TARP (Troubled Asset Relief Program) used earlier in the GFEC in the US.

10 The ECB also announced that in September 2014 it was commencing a Targeted Long-Term Refinancing Operation (TLTRO). Under this program, which represented a significant expansion of its quantitative easing operation, the ECB would provide additional liquidity to European banks up to a maximum of one trillion euros. *Inter alia*, this action by the ECB was expected to exert significant downward pressure on the value of the euro – a result apparently desired by the ECB.

quantitative measures of financial risk suggested that international perceptions of the risk of sovereign debt of Eurozone countries had returned to more normal levels, and that international investor confidence in the Eurozone banking system had been restored. For example, the positive impact of the policy measures on each of the country's 10-year government bond yield differentials can be clearly seen in **Exhibit 19.8**.

As a final note, it should be said that the solution to the crisis for a number of countries included a range of quite painful economic austerity measures being imposed on their citizens. These measures, along with unemployment rates in excess of 25 per cent in some countries, led to very significant – and sometimes quite violent – social unrest and political upheaval. On this, with the election of Alexis Tsipras and his 'anti-austerity' Syriza Party in Greece in January 2015, there had re-emerged by July the very real possibility that Greece and its creditors would ultimately fail to agree to the conditions for continued financial support. Fortunately, this eventuality did not occur and a recurrence of financial and economic turmoil for the Eurozone did not materialise.

The European sovereign debt crisis therefore represents a very recent and sombre reminder of what can happen when governments fail to keep their fiscal finances in order. Running larger budget deficits than necessary, for longer periods than prudent, and incurred for purposes other than for economically productive public investment, can lead to very high levels of government debt (as a ratio to GDP). If the debt is significantly externally owed, this can ultimately result in very serious and painful adjustments being wrought on a nation's citizens by the international community to whom the debt is owed. The reader is referred back to **Chapter 17** for the discussion there on issues pertaining to a country's national debt.

In summary

- The European sovereign debt – and banking – crisis demonstrates what can occur when governments fail to keep their public debt under control.
- Significant policy actions by the so-called Troika of the EC, the IMF and the ECB brought the situation under control.
- A number of the countries, especially Greece, experienced a range of quite painful economic austerity measures, which in turn ushered in significant social and political unrest and turmoil.

5 The COVID-19 pandemic economic impacts and policy responses in Australia

The COVID-19 pandemic was a once-in-a-century shock to almost all the countries in the world. This crisis has put enormous pressure on countries' health systems and also the economy and jobs. It is probably too early to analyse the economic impact of the virus and the policies during the time of the writing the book (late 2020). The actions taken to stop the spread of the virus started to disrupt many economies in the world by restricting the movement of the people and disturbing the supply chains. The global economy faced its worst economic recession since the Great Depression, despite the significant monetary and fiscal policy actions. The International Monetary Fund (IMF) forecast that the economic contraction in the US, Euro area, UK, and Canada will be more severe than that of Australia. Further, the economic contraction on the Asian economies will be less than the whole world's average and IMF forecasts that the China's output will increase.[11]

We can gain insight into the economic crisis due to the COVID-19 pandemic by referring to **Exhibit 19.9**. This exhibit shows the economic growth rates, inflation rates and unemployment rates for selected countries from 2017 until mid-2020. Here it evident that all the countries in the world experience negative economic growth rates except China. India experienced double-digit negative growth rates in 2020 while France, Greece and

[11] International Monetary Fund, World Economic Outlook Update, June 2020: A crisis like no other, an uncertain recovery, June 2020, https://www.imf.org/en/Publications/WEO/Issues/2020/06/24/WEOUp dateJune2020.

Exhibit 19.9 **Economic growth, Inflation and unemployment 2017–20: selected countries**

Country	Economic growth rate				Inflation rate				Unemployment rate			
	2020	2019	2018	2017	2020	2019	2018	2017	2020	2019	2018	2017
Australia	–4.2	1.8	2.8	2.4	0.7	1.6	1.9	2	6.9	5.2	5.3	5.6
Brazil	–5.8	1.1	1.3	1.3	2.7	3.7	3.7	3.4	13.4	11.9	12.3	12.8
China	1.9	6.1	6.7	6.9	2.9	2.9	2.1	1.6	3.8	3.6	3.8	3.9
France	–9.8	1.5	1.8	2.3	0.5	1.3	2.1	1.2	8.9	8.5	9	9.4
Germany	–6.0	0.6	1.3	2.6	0.5	1.3	2	1.7	4.3	3.1	3.4	3.8
Greece	–9.5	1.9	1.9	1.5	–0.6	0.5	0.8	1.1	19.9	17.3	19.3	21.5
India	–10.3	4.2	6.1	7	4.9	4.8	3.4	3.6	N/A	N/A	N/A	N/A
Indonesia	–1.5	5	5.2	5.1	2.1	2.8	3.3	3.8	8	5.3	5.3	5.5
Japan	–5.3	0.7	0.3	2.2	–0.1	0.5	1	0.5	3.3	2.4	2.4	2.8
Malaysia	–6.0	4.3	4.8	5.8	–1.1	0.7	1	3.8	4.9	3.3	3.3	3.4
New Zealand	–6.1	2.2	3.2	3.8	1.7	1.6	1.6	1.9	6	4.1	4.3	4.7
Russia	–4.1	1.3	2.5	1.8	3.2	4.5	2.9	3.7	5.6	4.6	4.8	5.2
South Korea	–1.9	2	2.9	3.2	0.5	0.4	1.5	1.9	4.1	3.8	3.8	3.7
United Kingdom	–9.8	1.5	1.3	1.9	0.8	1.8	2.5	2.7	5.4	3.8	4.1	4.4
United States	–4.3	2.2	3	2.3	1.5	1.8	2.4	2.1	8.9	3.7	3.9	4.3

N/A means the data were not available in that database at that time

Source: International Monetary Fund, Various country tables, https://www.imf.org/en/Publications/WEO/weo-database/2020/October, accessed 9 November 2020.

the UK experienced very high negative growth rates closer to 10 per cent. Almost all the countries presented in the exhibit had significantly low inflation rates except India, whose inflation rate has risen compared to pre-COVID-19. Unemployment rates were very high in 2020 for the countries selected, with Brazil and Greece having the highest. Unemployment rate data was not available for India and Indonesia, while the US experienced significant increases in unemployment rates.

In Australia, these impacts were initially felt by domestic businesses due to lockdowns and then by the businesses related to tourism and international students due to border closures. The majority believes that the handling of this crisis has been more successful in Australia compared to many countries in the world. Many professionals consider that the economic and policy response was also swift, well-targeted and substantial. Australia is on its way to recovery after the second wave of infection in the state of Victoria (at least by the time of writing this book); however, the path to economic recovery is more challenging than one could imagine. Economic policies in Australia are constrained by the economic policies of major trading partners as well as more enormous powers. Australia has benefitted from international engagement with the two biggest economies – the US and China – and many other countries. It is scary to think that the post-COVID world will be less likely to be open and liberal, and Australia will have to make difficult decisions.

Similar to most countries, Australia's GDP per person has fallen significantly, and it will take many years to get back to the pre-COVID level of living standards. Unemployment is rising at a rate that has not happened in the recent past, and the Australian government's budget deficit is heading to a new record high. The pandemic has increased not only government debt but also business and household debts.

The Australian government has committed a large sum of money as their fiscal policy response for the economic crisis due to COVID-19. This included an allocation of a huge sum of money to the health sector to protect the vulnerable and the elderly and to stop the spread of the virus. Also, the government's economic response to COVID-19 continued to provide temporary assistance to the affected workers and businesses. At the time of writing, the government is providing further assistance to get businesses back open and to get people back into jobs. The government's JobKeeper Payment Scheme was created to support job retention, maintain business cash flows and provide income support to employees. In addition, the government is provided households with income support and JobSeeker Payments for the unemployed. The government also supported small- and medium-sized businesses by providing temporary cash flow support for them to stay in business. Moreover, the government allocated funding for training for Australians to access critical skills they need to return to work when the economy recovers.[12]

Exhibit 19.6 showed the responses by the worldwide major central banks to the crisis in 2019–20 by decreasing the official cash rates to zero or near zero per cent. The Reserve Bank of Australia implemented monetary policy measures to assist the Australian economy during these extraordinary and challenging times. The RBA brought in further reductions to the cash rate, and the official cash rate sat at 0.1 per cent in November 2020, and this is a cumulative decline of 1.25 per cent over the last year. This easing of monetary policy is expected to boost the cash flow of businesses and households and also help the trade-exposed industries through exchange rates. Also, the RBA has lowered the funding costs for the entire banking system to provide low-cost credit to businesses and households and to provide incentives for lenders to support credit to businesses. In the past, the RBA usually targeted the cash rate; however, they are now complementing this by targeting the risk-free interest rate that is the yield on the three-year Australian Government Securities. Moreover, the RBA announced that they implement QE for the first time in November 2020 since it ran out of its traditional monetary policy tool (cash rate) and decided to increase the purchase of Australian Government bonds. The US continued QE measures as shown in the **Exhibit 19.7** to recover the economic crisis in 2019-20 due to the COVID-19 pandemic.

The other prominent economies during the COVID-19 pandemic implemented expansionary fiscal stance mainly by increasing spending on the virus spread, production of medical equipment, unemployment benefits and social security payments and public investments. Also, the governments implemented multiple steps to provide financial relief to the households and businesses by providing them with low-interest loans and loan-holidays to the existing loans. Most of the central banks around the world adopted monetary easing by reducing the official cash rates and injecting liquidity into the economy.

It is believed that Australia's economic integration into Asia, particularly to East Asia, will help its economic recovery since the Asian economies perform much better than Europe and the US. However, this pandemic has somewhat deteriorated the relationship between Australia and China. Also the trade and technology disagreement between America and China threatens the global openness of trade and technology. Even though Australia's government debt is not as high as in many other advanced economies, Australia may not be able to escape the consequences. The RBA may need to maintain very low-interest rates for a very long time and should not keep high levels of government debts in its balance sheets. Since the RBA will not be able to lower its interest rates further, managing unemployment will fall entirely on the fiscal policy. According to the Australian government, the Australian economy will recover faster. However, it will be a long and painful process, and the unemployment rate will be high for some time, particularly for women and young people. The economic outlook for Australia and the rest of the world is uncertain and controlling the spread of the virus remains a significant challenge.[13]

[12] Economic and Fiscal update of the Australian Government, https://budget.gov.au/2020-efu/economic-fiscal-update.htm.

[13] RBA responses for COVID-19, https://www.rba.gov.au/publications/smp/2020/aug/box-c-central-bank-policy-responses-to-covid-19.html.

In summary

- The COVID-19 pandemic has put enormous pressure on countries' health systems, the economy and jobs.
- As a response to the crisis, worldwide major central banks decreased the official cash rates to zero or near zero per cent.
- The health and economic outlook for Australia and the rest of the world is uncertain and depends on how efficiently and effectively countries can control the spread of the virus.

Key concepts

Global financial and economic crisis (GFEC)
Sub-prime mortgage
Securitisation
Mortgage-backed security (MBS)
Moral hazard
Troubled Asset Relief Program (TARP)
World Coincident Index
World Leading Index
Federal Funds Rate (FFR)
Quantitative easing (QE)
European Sovereign Debt Crisis

Summary

Here we summarise the key ideas we have discussed under each of this chapter's learning objectives.

1. Understand the essential cause of the global financial and economic crisis (GFEC) and its timing

- The **Global financial and economic crisis (GFEC)** of 2008–09 originated in the US home mortgage market. Extremely large volumes of low-quality **mortgage-backed securities (MBSs)** were created and sold to international institutional investors at much higher prices than their inherent risks warranted. Once these products were properly priced in international financial markets, many very well-known financial institutions found themselves in very serious financial trouble. As a result of the resulting financial crisis, the world went into simultaneous recession across the developed world and macroeconomic policymakers acted quickly to avert what could have become the second Great Depression.

2. Gain some appreciation of the economic consequences of the GFEC

- The GFEC, dated as lasting from April 2008 to June 2009, was a relatively rare and quite catastrophic event in that it produced a synchronised international recession, one of only four over the last half century.
- The world's 20th-century locomotor countries of the US, the UK, Germany and Japan all went into recession within six months of each other and as a result, output in these countries dropped significantly, although much less than in the Great Depression of the 1930s.
- With the substantially reduced demand, unemployment increased significantly and disinflation occurred in most countries, and in some, deflation actually occurred in 2009.
- As serious as the crisis was, its severity was significantly ameliorated by quick action by the world's economic policymakers.

3. Gain insight into the economic policy responses around the world to the crisis, and obtain an understanding of policy lessons learnt from the crisis

- The impact of the GFEC on the world's economies was most certainly ameliorated by the quick and quite decisive monetary policy easing actions of the world's major central banks. This monetary policy consisted in first quickly reducing the interest rate instruments of the central banks to historical lows, and secondly deploying the unconventional policy of **quantitative easing** whereby the central banks bought up both long-dated government and corporate bonds from the private sector to consciously and dramatically increase the Monetary Bases of their respective countries.
- Fiscal policy was also deployed by the world's governments whereby government expenditure was increased and funded by issuing government debt to the private sector. This caused some serious consequential sovereign debt problems in some countries.

4. Gain economic insights into its aftermath, especially the so-called European sovereign debt crisis, itself at least partially brought about by the GFEC

- The European sovereign debt and banking crisis is a vivid and very recent demonstration of what can occur when governments fail to keep their public debt under control. Whilst the GFEC certainly exacerbated the situation in which a number of Eurozone countries found themselves in 2009–10, their fundamental fiscal finance problems pre-dated the GFEC.
- Significant policy actions by the so-called 'Troika' of the EC, the IMF and the ECB brought the situation under control but not without a number of the countries in question, especially Greece, experiencing a range of quite painful economic austerity measures, which in turn ushered in significant social and political unrest and turmoil.

5. Understand the economic impact and policies of the COVID-19 pandemic

- The COVID-19 pandemic economic crisis that began in late 2019 is considered a once-in-a-century shock to the entire world, leading to the disruption of many economies in the world with reduced standards of living and rising unemployment.
- The government of Australia and the Reserve Bank of Australia acted swiftly and decisively to minimise the severe economic impacts. The government supported the health system, economy and jobs but the budget deficit increased to an extraordinarily high level.

Study questions and problems

1 Using **Exhibit 19.2** answer the following questions. Of the four 20th-century locomotor economies of the US, Japan, Germany and the UK, which country was the first into recession and what was the date? Which country was last out of recession and what was the date? Which of the four countries experienced the longest recession?

2 By searching the Fed and the RBA websites, find out the levels of the US target FFR and the Australian target overnight cash rate at the time you are answering this question. Have those rates been on the rise, decline or unchanged over the two years before? What does that suggest to you has been the stance of monetary policy (tightening or loosening) in the two countries in the recent period up to your reading this chapter?

3 Describe the process of securitisation and define mortgage-backed securities (MBSs). What role did these securities, the originating of 'sub-prime mortgages', and moral hazard play in the onset of the GFEC?

4 Apart from those cited in the chapter, provide two examples of moral hazard and in each case explain how its presence can lead to undesirable behaviours.

5 Explain what you think may have been learnt from the 1930s Great Depression by policymakers around the world that helped prevent the GFEC becoming the second Great Depression.

6 By thinking about what is commonly regarded as the traditional transmission mechanism of monetary policy, explain why policymakers eventually felt the need to resort to quantitative easing (QE) to assist in combating the recessionary forces of the GFEC and the economic crisis of the COVID-19 pandemic.

7 Following from question (6), discuss the operation of QE, how it is thought to work in stimulating the economy, and how this differs from the more traditional conception of the monetary policy transmission mechanism. (*Hint:* Refer back to **Chapter 16** and the discussion there on the quantity theory of money and the monetarist view of the monetary policy transmission mechanism.)

8 At the time you are doing this question, find out what the 10-year government bond yield is for Greece and for Germany and compare the differential with the data for Greece in **Exhibit 19.8**. What does this comparison suggest to you as far as Greece's sovereign debt situation is concerned?

Answers to 'You're the economist'

Moral hazard and the health sector

If the cost to the individual of the doctor's visit is zero then some commentators have argued that it is quite possible a person may 'overuse' the service, thus incurring a heavier burden on the taxpayer than necessary. Similarly, a range of commentators have argued that a wide range of investigative procedures that may really not be entirely crucial to the satisfactory treatment of a patient may be undertaken anyway, 'just to be sure', since the patient does not need to meet the cost themselves; but again, this comes at great cost to the rest of the community. If such actions are in fact prevalent, then over time the total social cost of the universal health cover system can become so burdensome as to put its continuation at risk. If policymakers are persuaded by these considerations, a range of policy options could be deployed to assist in keeping the total social cost of the universal health system manageable. For example, *inter alia*, if you said the government could consider requiring a small co-payment by the individual at the point-of-service to reduce the possibility of overuse then YOU ARE THINKING LIKE AN ECONOMIST.

Can nominal interest rates ever go below zero?

Not too long ago most folks would probably have said that zero would be as low as nominal interest rates could/would go. After all, why would one person willingly pay another for the privilege of lending money to them for some period of time? Why not just keep the money themselves rather than lend it? Well, the GFEC has turned this view on its head. For example, from January 2016 the Bank of Japan's target rate was reduced to –0.1 per cent. In other words, Japanese banks not only received no interest on deposits they willingly chose to keep with the Bank of Japan but they actually had to pay it a fee for the privilege of keeping deposits with it! Furthermore, both Japan's and Germany's 10-year government bonds have now been traded in the market at prices that imply negative yields. This means, if the buyers of those bonds hold them to maturity, in 10 years' time they will receive back less money than they paid for them! Their motivation in doing this is that they at least have the security of knowing their funds are safe. If you said that, with the increased global financial market uncertainty ushered in by the GFEC, some lenders will be satisfied to lend funds to safe-haven 'gilt' borrowers at negative nominal interest rates then YOU ARE THINKING LIKE AN ECONOMIST.

Multiple-choice questions

1 Referring to **Exhibit 19.2**, how many synchronised international recessions have there been since the Second World War?
- **a** 2
- **b** 4
- **c** 6
- **d** 1

2 Again referring to **Exhibit 19.2**, which statement(s) is/are correct? (There may be more than one answer.)
- **a** The GFEC was a synchronised international recession with duration from April 2008 to June 2009.
- **b** During the GFEC, the US was in recession commencing in January 2008, after a peak in its business cycle in December 2007.
- **c** China and India were unable to withstand the recessionary force of the GFEC and also sank into recession in 2008–09.
- **d** All of the US, UK, Japan and Germany had entered into recession by the middle of 2008.

3 Which of the following countries is not one of the 20th century 'locomotor countries'?
- **a** Japan
- **b** China
- **c** US
- **d** Germany

4 Securitisation is defined as:
- **a** the process by which policymakers endeavour to make an economy secure from recession.
- **b** securing reform measures from countries with sovereign debt problems in return for bailout funds.
- **c** bundling together a collection of similar risky financial assets to create a more secure, less risky financial security.
- **d** repackaging a large number of similar financial assets into a new derivative financial security.

5 A mortgage-backed security is:
- **a** the security that a bank requires a mortgagee to demonstrate in order to extend a loan for a residential property.
- **b** created when a lender makes a loan and holds a mortgage over the property as collateral.
- **c** a derivative security resulting from securitisation when the underlying assets being bundled together are property mortgages.
- **d** a financial security that is written on the reverse side of a residential mortgage.

6 Moral hazard occurs when:
- **a** one person takes on more risk because someone else bears the costs.
- **b** people engage in buying morally questionable products.
- **c** companies fail to properly disclose potentially hazardous facts about their products.
- **d** any one of the above situations occurs.

7 A major cause of the GFEC is considered to be:
- **a** poor co-ordination and exceedingly slow response in implementing countercyclical monetary policy by the world's major central banks in 2008–09.
- **b** the existence and unpredictability of the policy information, determination and effectiveness lags inherent in countercyclical macroeconomic policy.
- **c** the purchase by many financial institutions of what turned out to be very poor quality sub-prime mortgage-backed derivative securities.
- **d** the financial contamination that spread around the world due to the sovereign debt problems of a number of countries in the Eurozone.

8 Quantitative easing is defined as:
- **a** the central bank easing the quantitative lending restrictions applying to banks.
- **b** an easing (i.e., a reduction) in the quantity of bonds on offer for sale by the federal government.
- **c** a central bank expanding its balance sheet by buying long-dated government and corporate bonds from the financial sector.
- **d** an easing in the aggregate quantity of goods and services demanded in the economy as often happens during a recession.

9 Which of the following countries is not a member of the so-called 'PIIGS' group of Eurozone countries?
- **a** Greece
- **b** Ireland
- **c** India
- **d** Spain

10 What are the policy actions taken by countries to combat the economic crisis due to the global COVID-19 pandemic?
- **a** Raise the cash rate and implement tight monetary policy.
- **b** Decrease the cash rates and provide tax reliefs to the people.
- **c** Reduce the amount of money available and discourage government spending.
- **d** Increase the tax rates and cut down the welfare payments offered by the government.

Glossary

Abatement subsidies government payments to polluters in exchange for a reduction in their emissions of pollutants. (Chapter 10)

Absolute advantage The ability of a country to produce a product using fewer resources than another country. (Chapter 18)

Aggregate demand (AD) The total of the four categories of demand, namely $C + I + G + (X - M)$. (Chapter 14)

Aggregate demand (AD) curve A curve showing the level of total real GDP that households, businesses, government and foreigners (net exports) would be willing to purchase at different possible average price levels during a time period, *ceteris paribus*. (Chapter 14)

Aggregate demand–output model Also called the Keynesian model; the model that determines the equilibrium level of real GDP by equating output produced to aggregate demand. (Chapter 14)

Aggregate supply (AS) curve The curve that represents the level of real GDP that firms would be willing to produce at different possible price levels during a particular time period, *ceteris paribus*. (Chapter 14)

Anchoring heuristic The tendency to use the first or most recognisable piece of information to make decision or judgement when faced with uncertainty or lack of any information. (Chapter 10)

Appreciation of a currency A rise in the price of one currency relative to another. (Chapter 18)

Australian Securities Exchange (ASX) Where shares in publicly listed companies are floated and traded, along with the buying and selling of a wide range of financial derivative products. (Chapter 15)

Automatic stabilisers Federal expenditures and tax revenues that automatically change over the course of the business cycle in such a way as to help stabilise an economic expansion or contraction. (Chapter 17)

Autonomous expenditure Spending that does not vary in any systematic way, with identifiable economic determinants, such as, the current level of disposable income or interest rates. (Chapter 14)

Availability heuristic The tendency to overestimate (underestimate) the frequency of an event or phenomenon if the examples of the event are familiar (unfamiliar). (Chapter 10)

Average fixed cost Total fixed cost divided by the quantity of output produced. (Chapter 6)

Average total cost Total cost divided by the quantity of output produced. (Chapter 6)

Average variable cost Total variable cost divided by the quantity of output produced. (Chapter 6)

Balance of payments A summary record of the international transactions between a country and other countries during a given period of time. (Chapter 18)

Balanced budget multiplier An equal change in government spending and taxes that changes aggregate demand by the amount of the change in government spending. (Chapter 17)

Barter The direct exchange of one good for another good, rather than for money. (Chapter 15)

Base year A year chosen as a reference point for comparison with some earlier or later year. (Chapter 13)

Behavioural economics Study of human behaviour by combining both economic and psychological factors. (Chapter 1)

Bond A document stipulating a fixed annual interest (the coupon) for a defined number of years (its maturity), and on maturity, the amount to be paid out (its face value). (Chapter 15)

Broad money Equals M3 plus the non-deposit borrowings by AFIs. (Chapter 15)

Budget deficit A budget situation in which government expenditures exceed government revenues in a given time period. (Chapter 17)

Budget surplus A budget situation in which government revenues exceed government expenditures in a given time period. (Chapter 17)

Burden of the debt The possibility that existing national debt may represent a burden on current and future generations, namely, to pay the future interest payments on it and to eventually repay the principal. (Chapter 17)

Business cycle Alternating periods of economic growth and contraction, which can be dated by changes in output, income, sales and employment measures. (Chapter 12)

Cap and trade scheme An alternative name for an emissions trading scheme. (Chapter 10)

Capital The physical plant, machinery and equipment used to produce other goods. Capital goods are human-made goods that do not directly satisfy human wants. (Chapter 1)

Capital adequacy Requirement to maintain a certain minimum percentage of its risk-weighted assets in the form of shareholders' equity. (Chapter 15)

Carbon tax A tax levied per tonne of carbon emitted by producers. (Chapter 10)

Cartel A group of firms formally agreeing to control the price and the output of a product. (Chapter 9)

Ceteris paribus A Latin phrase that means that while certain variables change, all other things remain unchanged. (Chapter 1)

Chain volume measure of GDP The chain volume measure of GDP is a measure of GDP that provides a measure of all the final goods and services produced during a given period using a set of relevant prices prevailing in some selected reference year. (Chapter 11)

Change in demand An increase or decrease in the demand at each possible price. An increase in demand is a rightward shift in the entire demand curve. A decrease in demand is a leftward shift in the entire demand curve. (Chapter 3)

Change in quantity demanded A movement between points along a stationary demand curve. (Chapter 3)

Change in quantity supplied A movement between points along a stationary supply curve, *ceteris paribus.* (Chapter 3)

Change in supply An increase or decrease in the supply at each possible price. An increase in supply is a rightward shift in the entire supply curve. A decrease in supply is a leftward shift in the entire supply curve. (Chapter 3)

Cheque account deposits Convertible to currency 'on demand' by writing a cheque. (Chapter 15)

Circular flow model A diagram showing the flow of products from businesses to households and the flow of resources from households to businesses. In exchange for these products and resources, money payments flow between businesses and households. (Chapter 11)

Civilian labour force Those people 15 years of age and older who are employed or who are actively seeking a job, excluding those in the armed forces, homemakers, students, discouraged workers and other persons not in the labour force. (Chapter 13)

Climate change Long-term changes in the world's climate that are widely believed to involve global warming, and which are also widely believed to be caused by human activity. (Chapter 10)

Coincident indicators Variables that change at about the same time as the economy shifts from one business cycle phase into another. (Chapter 12)

Commercial bill Bill created as evidence of a loan, stipulating a face value to be repaid at maturity and the time to maturity (usually 90 or 180 days). (Chapter 15)

Commodity money Anything that serves as money while having market value in other uses. (Chapter 15)

Commonwealth Government Securities (CGS) Commonwealth government bonds (maturities of more than one year) and Treasury Notes (13- and 26-week). (Chapter 15)

Comparative advantage The ability of a country to produce a product at a lower opportunity cost than another country. (Chapter 18)

Competition policy Government policy that has the objective of increasing competition in the economy or of encouraging firms that are not competitive to behave as if they were. (Chapter 7)

Competitive firm's short-run supply curve The firm's marginal cost curve above the minimum point on its average variable cost curve. (Chapter 7)

Competitive industry's long-run supply curve The curve that shows the quantities supplied by the industry at different equilibrium prices after firms complete their entry and exit. (Chapter 7)

Complementary goods Goods that are jointly consumed. As a result, there is an inverse relationship between a price change for one good and the demand for its 'complementary' good. (Chapter 3)

Concentration ratio A measure that indicates the percentage of total sales in the industry generated by its largest firms. (Chapter 9)

Conditional projection The implementation of monetary policy in Australia during the period 1976–85 consisted of a form of monetary targeting. Each year a 'conditional projection' for the growth of the monetary aggregate, M3, was announced in the federal budget by the Treasurer. (Chapter 16)

Constant returns to scale A situation in which the long-run average cost curve is horizontal as the firm increases its scale of operation. (Chapter 6)

Constant-cost industry An industry in which the expansion of industry output by the entry of new firms has no effect on the firms' cost curves. (Chapter 7)

Consumer price index (CPI) An index that measures changes in the average prices of consumer goods and services. (Chapter 13)

Consumer sovereignty The freedom of consumers to make their own choices about which goods and services to buy. (Chapter 3)

Consumer surplus The gain by the consumer in an exchange. At the equilibrium market price, the consumer surplus is represented in the area under the demand curve and above the market equilibrium price. (Chapter 3)

Consumption demand Consumption is the most significant component of aggregate expenditure (around 60%), and the most important determinant of a household's demand for consumer goods and services is disposable income. (Chapter 14)

Cost-push inflation A rise in the general price level resulting from an increase in the cost of production, irrespective of demand conditions. (Chapter 13)

Countercyclical macroeconomic policy Policy that aims to smooth out the fluctuations in the business cycle. (Chapter 16)

Credit creation The process by which money is created by banks extending new loans to their customers in the form of newly created bank deposits. (Chapter 15)

Cross-elasticity of demand The ratio of the percentage change in the quantity demanded of a good or service to a given percentage change in the price of a related good or service. (Chapter 5)

Crowding-out effect A reduction in private-sector spending as a result of federal budget deficits financed by government borrowing from the private sector. (Chapter 17)

Currency Money, including coins and paper money. (Chapter 15)

Current account balance on goods and services This represents the nation's net position in respect of its exports and imports of both goods and services. (Chapter 18)

Current account balance on merchandise trade The value of a nation's merchandise imports subtracted from its merchandise exports. (Chapter 18)

Cyclical unemployment Unemployment caused by the lack of a sufficient number of jobs during a recession. (Chapter 13)

Cyclically adjusted budget deficit The actual recorded budget deficit adjusted for the stage of the economy's business cycle. (Chapter 17)

Debt financing of the deficit When a budget deficit is financed by selling government securities to the general public– that is, by the government borrowing from the public to fund its expenditure. (Chapter 17)

Decision heuristic Mental short cut or rule of thumb to make a quick decision or form a judgement when dealing with a complex or obscure task. (Chapter 10)

Decreasing-cost industry An industry in which the expansion of industry output by the entry of new firms decreases the firms' input costs. (Chapter 7)

Deflation A decrease in the general (average) price level of goods and services in the economy. (Chapter 13)

Demand curve The line connecting the possible price and quantity purchasing responses of an individual consumer. (Chapter 3)

Demand for money A downward-sloping schedule – or inverse relationship – representing the overall quantity of money that people wish to hold at different interest rates, *ceteris paribus*. (Chapter 15)

Demand-pull inflation A rise in the general price level resulting from an excess of total spending (demand) over supply. (Chapter 13)

Depreciation of a currency A fall in the price of one currency relative to another. (Chapter 18)

Direct relationship A positive association between two variables. When one variable increases, the other variable increases, and when one variable decreases, the other variable decreases. (Chapter 1)

Discouraged worker A person who wants to work, but who has given up searching for work because he or she believes there will be no job offers. (Chapter 13)

Discretionary fiscal policy The deliberate use of changes in government spending and/or taxes to alter aggregate demand and stabilise the economy's business cycle. (Chapter 17)

Diseconomies of scale A situation in which the long-run average cost curve rises as the firm increases output. (Chapter 6)

Disinflation A reduction in the rate of inflation. (Chapter 13)

Domestic deficit The difference between domestically sourced government revenue and domestic government expenditure. (Chapter 17)

Dominant strategy An action the player will always take in game theory regardless of the other player's action. (Chapter 9)

Duopoly An oligopoly market structure characterised by there being just two sellers. (Chapter 9)

Dutch Disease A term referring to the possible relationship between a significant increase in natural resource exports and a consequential decline in the manufacturing sector. The proposed mechanism is the increase in natural resource export revenues appreciating the currency, resulting in other exports becoming less profitable, and import-competing businesses less able to compete with imports made cheaper by the appreciated currency. (Chapter 18)

Economic growth The ability of an economy to produce greater levels of output, represented by an outward shift of its production possibilities frontier. (Chapter 2)

Economic profit Total revenue minus explicit and implicit costs. (Chapter 6)

Economics The study of how human beings deal with the problem of scarcity to satisfy unlimited wants. (Chapter 1)

Economies of scale A situation in which the long-run average cost curve declines as the firm increases output. (Chapter 6)

Efficiency A condition where the scarce resources are utilised to maximise the social surplus (sum of consumer surplus and producer surplus). Efficiency occurs at the competitive market equilibrium. (Chapter 3)

Elastic demand A condition in which the percentage change in quantity demanded is greater than the percentage change in price. (Chapter 5)

Embargo A law that bars trade with another country in a range of products. (Chapter 18)

Emissions trading scheme (ETS) A government scheme requiring firms that produce emissions of pollutants to purchase tradeable permits giving them the right to do so. (Chapter 10)

Employment to population ratio The number of people aged 15 years and over that are employed as a percentage of the population that are 15 years and above. (Chapter 13)

Endogenous growth model The view that the process of technological change is significantly endogenous to economic growth rather than being exogenous, as in earlier economic growth models. (Chapter 12)

Enlightened self-interest This involves members of the community respecting the laws and social mores of society while they pursue their own individual goals. (Chapter 1)

Entrepreneurship The creative ability of individuals to seek profits by combining resources to produce new or existing products. (Chapter 1)

Equation of exchange An accounting identity that states that the money supply times the velocity of money equals total spending. (Chapter 16)

Equilibrium A market condition that occurs at any price for which the quantity demanded and the quantity supplied are equal. (Chapter 3)

European Sovereign Debt Crisis Within the 19-country Eurozone area there arose in 2009 widespread fears that a particular five-country sub-group of countries – namely, Portugal, Ireland, Italy, Greece and Spain – may default on their government debt. (Chapter 19)

Excess demand A market condition existing at any price where the quantity supplied is less than the quantity demanded. (Chapter 3)

Excess supply A market condition existing at any price where the quantity supplied is greater than the quantity demanded. (Chapter 3)

Exchange rate The number of units of one nation's currency that equals one unit of another nation's currency. (Chapter 18)

Exchange settlement account Each bank maintains an exchange settlement account with the RBA and they must always have a positive balance by the end of each trading day. (Chapter 15)

Exogenous technological progress Technological progress that occurs independent of, or outside of, the economic growth process. (Chapter 12)

Expansion An upturn in the business cycle during which real GDP, employment and other measures of aggregate economic activity rise. (Chapter 12)

Expenditure approach The national income accounting method that measures GDP by adding all the spending on final products during a period of time. (Chapter 11)

Experimental economics A branch of economics in which experiments designed to mimic real-world scenarios are used to explore aspects of human behaviour relevant to economics. (Chapter 1)

Explicit costs Payments to non-owners of a firm for their resources. (Chapter 6)

Externality A cost or benefit imposed on people other than the consumers and producers of a good or service. (Chapter 4)

Factors of production Inputs used to produce goods and services. Economists divide factors of production into three categories: land, labour and capital. (Chapter 1)

Fair trade Term used to represent the view that a country should only reduce its barriers to imports from another country if the other country does not have some sort of 'unfair' competitive advantage over it and the other country is also willing to reduce its import barriers reciprocally. (Chapter 18)

Federal Funds Rate (FFR) The key monetary policy interest rate instrument of the Federal Reserve System, or Fed, the US central bank. It is the equivalent to the 'overnight cash rate' in Australia. (Chapter 19)

Fiat money Money accepted by law and not because of redeemability or intrinsic value. (Chapter 15)

Final products Finished goods and services produced for the ultimate user. (Chapter 11)

Financial system Facilitates the transfer of resources from savers to borrowers. (Chapter 15)

Fiscal policy The use of government spending and taxes and their influence on the nation's economic growth, employment and price level. (Chapter 17)

Fiscal stance The term used to represent the extent to which the authorities are actively trying to expand or retard the growth of the economy through the settings of government revenue-raising and spending policies. (Chapter 17)

Fixed input Any resource for which the quantity cannot change during the period of time under consideration. (Chapter 6)

Flexible exchange rate system A system in which countries allow their currencies to adjust continuously according to the forces of supply and demand. Most developed countries in the world today employ this type of exchange rate system. (Chapter 18)

Floating exchange rate In a floating exchange rate, market demand and supply determine the foreign price of the currency. (Chapter 16)

Flow A rate of change in a quantity during a given time period, measured in units per time period, such as dollars per year. For example, income and consumption are flows that occur per week, per month or per year. (Chapter 11)

Foreign exchange (Forex) market Where individuals and companies exchange Australian dollars for foreign currencies or vice versa. (Chapter 15)

Fractional reserve banking A system in which banks keep only a small percentage of their deposits on reserve as vault cash or as deposits at the central bank. (Chapter 15)

Free trade The flow of goods and services between countries without restrictions or special taxes or subsidies applied to their flow. (Chapter 18)

Free trade agreement (FTA) Parties to the agreement agree to eliminate or significantly reduce tariffs and other trade barriers among themselves. However, each individual country maintains its own tariff policy against non-parties to the agreement. (Chapter 18)

Frictional unemployment Unemployment caused by the normal search time required by workers with marketable skills who are changing jobs, initially entering the labour force, or re-entering the labour force. (Chapter 13)

Full employment The situation in which an economy operates at an unemployment rate equal to the sum of the seasonal, frictional and structural unemployment rates. (Chapter 13)

Game theory The study of strategic decision-making in situations where the players are inter-dependent. (Chapter 9)

GDP at factor cost A measure of GDP arrived at by adding all the incomes of all factors of production (including labour, land and capital). (Chapter 11)

GDP at market prices A measure of GDP arrived at by valuing GDP produced at the prices at which the goods and services sell. The difference between this and GDP at factor cost is the existence of government indirect taxes less subsidies. (Chapter 11)

GDP gap The difference between full-employment real GDP and actual real GDP. (Chapter 12)

GDP implicit price deflator A price index derived implicitly from the ratio of a set of nominal or current price GDP data to the corresponding chain volume measures of GDP for the same time periods. (Chapter 11)

Global financial and economic crisis (GFEC) Also referred to as the Global Financial Crisis (GFC), the GFEC was a period of tremendous international financial system instability, which began in 2007, intensified in 2008, and ushered in a synchronised international recession dating from April 2008 to June 2009. In the US, the recession in that

country associated with the GFEC is often referred to as the 'Great Recession'. (Chapter 19)

Goal of monetary policy In Australia, the goal of monetary policy is to keep consumer price inflation between 2 and 3 per cent, on average, over the medium term. (Chapter 16)

Goals of macroeconomic policy These are: first, to reduce the severity of the ups and downs of the business cycle; and second, to assist with the achievement of the highest possible sustainable economic growth trajectory consistent with the economic, social and cultural aspirations of the community. (Chapter 12)

Gross disposable income Obtained after subtracting net transfers overseas (mainly foreign aid and pensions paid to recipients living abroad) from gross national income. (Chapter 11)

Gross domestic product (GDP) The market value of all final goods and services produced within a nation's geographic borders during a period of time, usually a quarter or a year. (Chapter 11)

Gross investment Any spending that maintains or increases the capital stock in the country. (Chapter 11)

Gross national expenditure (GNE) The sum of consumption, investment and government spending in the economy in a quarter or a year (= C + I + G), irrespective of whether the spending was on domestic production or imports. (Chapter 11)

Gross national income (GNI) The income accruing to a country's residents from the production of all final goods and services during a period of time, no matter where in the world the goods and services are produced. (Chapter 11)

Herding heuristic The tendency to make a choice by following others on the belief that the decision by the group should be right (Chapter 10)

Horizontal (or Keynesian) range The horizontal segment of the aggregate supply curve, which represents an economy in a severe recession. (Chapter 14)

Hyperinflation An extremely rapid rise in the general price level, generally defined as prices rising by 50 per cent or more each month. (Chapter 13)

Hysteresis Occurs when the full-employment rate of unemployment increases (decreases) as the actual unemployment rate increases (decreases). (Chapter 13)

Implicit costs The opportunity costs of using resources owned by a firm. (Chapter 6)

Income approach The national income accounting method that measures GDP by adding all incomes, including compensation of employees, rents, net interest and profits. (Chapter 11)

Income elasticity of demand The ratio of the percentage change in the quantity demanded of a good or service to a given percentage change in income. (Chapter 5)

Increasing-cost industry An industry in which the expansion of industry output by the entry of new firms increases the firm's input costs. (Chapter 7)

Independent relationship No association between two variables. When one variable changes, the other variable remains unchanged. (Chapter 1)

Index numbers Indicate changes in the value of a variable over time where the value of the variable in the base (or reference) year is set at 100. (Chapter 1)

Indirect taxes less subsidies Government taxes levied on the production and/or sale of goods and services sold, less any subsidies paid to business from government. Examples include general sales taxes (like Australia's GST), excise taxes and customs duties. (Chapter 11)

Inelastic demand A condition in which the percentage change in quantity demanded is smaller than the percentage change in price. (Chapter 5)

Inferior good Any good or service for which there is an inverse relationship between changes in income and its demand curve. (Chapter 3)

Inflation An increase in the general (average) price level of goods and services in the economy. (Chapter 13)

Information lag The delay (months) before information about the current phase of economic activity in the real world becomes available. (Chapter 16)

Interest-rate effect The impact on total spending (real GDP) of the direct relationship between the price level and the interest rate. (Chapter 14)

Intermediate products Goods and services used as inputs for the production of final products. (Chapter 11)

Intermediate range The upward-sloping segment of the aggregate supply curve, which represents an economy as it approaches full-employment output. (Chapter 14)

Inverse relationship A negative association between two variables. When one variable increases, the other variable decreases, and when one variable decreases, the other variable increases. (Chapter 1)

Investment The process of producing human and physical capitals, such as education, factories and machines. Increase in capital stock will shift a country's production possibility frontier to the right. (Chapter 2)

Investment demand *Ceteris paribus*, the quantity of investment demanded by businesses increases as rates of interest drop. Shifts in investment demand – that is, desired investment spending at any given interest rate – result from changes in profit expectations, technological change, capacity utilisation and business taxes. (Chapter 14)

Labour The mental and physical capacity of workers to produce goods and services. (Chapter 1)

Labour force participation rate The percentage of people aged 15 years and over that are in the labour force. (Chapter 13)

Laffer curve Representation of the relationship between the income tax rate and the amount of income tax revenue collected by the government. (Chapter 17)

Lagging indicators Variables that change direction after a phase change in the economy has occurred, thereby confirming its occurrence. (Chapter 12)

Land A shorthand expression for any resource provided by nature. (Chapter 1)

Law of demand The principle that there is an inverse relationship between the price of a good or service and the quantity buyers are willing to purchase in a defined time period, *ceteris paribus*. (Chapter 3)

Law of diminishing marginal utility The principle that the marginal utility obtained a good or service falls when more of it is consumed (Chapter 3)

Law of diminishing returns The principle that, beyond some point, the marginal product decreases as additional units of a variable factor are added to a fixed factor. (Chapter 6)

Law of increasing opportunity costs The principle that the opportunity cost increases as the production of one output expands. (Chapter 2)

Law of supply The principle that there is a direct relationship between the price of a good and the quantity sellers are willing to offer for sale in a defined time period, *ceteris paribus*. (Chapter 3)

Leading indicators Variables that change direction before the economy shifts from one business cycle phase into another. (Chapter 12)

Long run A sufficient period of time to allow all inputs to be varied. (Chapter 6)

Long-run average cost curve The curve that traces the lowest cost per unit at which a firm can produce any level of output when the firm can build any desired plant size. (Chapter 6)

M1 The narrowest definition of the money supply, consisting of currency and cheque account deposits. (Chapter 15)

M3 Equals M1 plus all other bank and other ADI deposits. (Chapter 15)

Macroeconomics The branch of economics that studies decision-making for the economy as a whole. (Chapter 1)

Marginal analysis An examination of the effects of additions to or subtractions from a current situation. (Chapter 2)

Marginal cost The increase in cost of production when an additional unit of output is produced. (Chapter 3)

Marginal product The change in total output produced by adding one unit of a variable input, with all other inputs used being held constant. (Chapter 6)

Marginal propensity to consume (MPC) The change in consumption resulting from a given change in disposable income. (Chapter 14)

Marginal propensity to save (MPS) The change in savings resulting from a given change in disposable income, namely 1 – MPC. (Chapter 14)

Marginal revenue The change in total revenue from the sale of one additional unit of output. (Chapter 7)

Marginal-average rule When applied to cost relationships, the rule stating that when marginal cost is below average cost, average cost falls. When marginal cost is above average cost, average cost rises. When marginal cost equals average cost, average cost is at its minimum point. (Chapter 6)

Market Any arrangement in which the buyers and sellers interact to determine the price and quantity of goods and services exchanged. (Chapter 3)

Market failure A situation in which the price system fails to operate efficiently, creating a problem for society. (Chapter 4)

Market structure A classification system for the key characteristics of a market, including the number of firms, the similarity of the products they sell and the ease of entry into and exit from the market. (Chapter 7)

Medium of exchange The primary function of money, which is to be widely accepted in exchange for goods and services. (Chapter 15)

Microeconomics The branch of economics that studies decision-making by a single individual, household, firm or industry. (Chapter 1)

Minimum liquidity requirement Requirement that a bank must always maintain a certain minimal reserve level of highly liquid assets. (Chapter 15)

Model A simplified description of reality used to understand and predict the relationship between variables. (Chapter 1)

Monetarism The theory that changes in the growth of the money supply directly determine changes in prices, real GDP and employment. (Chapter 16)

Monetary Base The primary liquidity of the financial system, consisting of currency in circulation in the private sector, plus the deposits of banks with the RBA, plus other RBA liabilities to the private non-bank sector. (Chapter 15)

Monetary policy Actions taken by the central bank to influence interest rates, aggregate demand, economic activity and prices in the economy to achieve macroeconomic goals. (Chapter 15)

Monetary policy transmission mechanism The central bank operates in the financial system to influence interest rates in the desired direction, which in turn influence aggregate demand, thereby leading to changes in prices, real GDP and employment. (Chapter 16)

Money Anything that serves as a medium of exchange, unit of account and store of value. (Chapter 15)

Money financing of the deficit When a budget deficit is financed by the central bank purchasing the securities issued by the government to fund its deficit. (Chapter 17)

Money multiplier The extent to which banks can multiply an initial increase in their reserves. It may be thought of as the reciprocal of the minimum reserve liquidity ratio that banks maintain to satisfy regulatory requirements; eg., if the minimum reserve liquidity ratio is 10 per cent = 0.1, then the money multiplier would be 10 (= 1/0.1) (Chapter 15)

Monopolistic competition A market structure characterised by (1) many small sellers, (2) a differentiated product and (3) easy market entry and exit. (Chapter 9)

Monopoly A market structure characterised by (1) a single seller, (2) a unique product and (3) extremely difficult or impossible entry into the market. (Chapter 8)

Moral hazard A moral hazard occurs when one person takes on more risk because someone else bears the costs. (Chapter 19)

Mortgage-backed security (MBS) In securitisation, when the underlying assets are mortgages that are being bundled together, the resulting derivative security is called a mortgage-backed security. (Chapter 19)

Mutual interdependence A condition in which an action by one firm may cause a reaction on the part of other firms. (Chapter 9)

Nash equilibrium The combination of strategies in a game where neither player has any incentive to change strategies given the strategy of his competitor. (Chapter 9)

National debt The total face value of outstanding Commonwealth Government securities (CGS) on issue. (Chapter 17)

National investment A country's total spending on capital goods by the private sector and government. If national investment is greater than national saving, then the country will have had an inflow of foreign investment. (Chapter 11)

National saving A country's total income accruing to its residents less its consumption. (Chapter 11)

Natural monopoly An industry in which the long-run average cost of production declines throughout a large range of output. As a result, a single firm can supply the entire market demand at a lower cost than two or more smaller firms. (Chapter 8)

Net disposable income Gross disposable income less an allowance for capital depreciation. (Chapter 11)

Net domestic product (NDP) Gross domestic product minus depreciation of the capital worn out in producing output. (Chapter 11)

Net exports effect The impact on total aggregate demand (for real GDP) caused by the inverse relationship between the price level and the net exports of an economy. (Chapter 14)

Net foreign income Also known as 'net foreign primary income'; the difference between aggregate investment income flows into and out of a country. Such income includes interest payments on loans, rent on property and dividends on shares. (Chapter 18)

Net foreign investment The increase (or decrease) in the total of a country's net foreign liabilities (i.e., its net foreign debt plus the net foreign ownership of its domestic factors of production). (Chapter 11)

Net investment Only that spending that actually increases the nation's capital stock. (Chapter 11)

Net national income Also known as net national product; measures the income accruing to a nation's residents from the production of final goods and services net of an allowance for capital depreciation. (Chapter 11)

Nominal (or money) income The actual number of dollars received as income during a period of time. (Chapter 13)

Nominal GDP The value of all final goods and services produced in a period based on the prices existing during the time period of production. (Chapter 11)

Nominal interest rate The actual rate of interest earned or paid over a period of time. (Chapter 13)

Non-price competition The situation in which a firm competes using differences in advertising, packaging, product development, quality and service, rather than lower prices. (Chapter 9)

Normal good Any good or service for which there is a direct relationship between changes in income and its demand. (Chapter 3)

Normal profit The minimum return to own inputs (including risk taking) necessary to keep a firm in operation. A firm that earns normal profit earns total revenue equal to its total opportunity cost (total explicit and implicit costs). (Chapter 7)

Normative economics An analysis based on value judgements. (Chapter 1)

Nudging A choice architecture to influence optimal decision-making without changing the incentives and reducing individual freedom. (Chapter 10)

Oligopoly A market structure characterised by (1) few sellers, (2) either a homogeneous or a differentiated product and (3) barriers to market entry. (Chapter 9)

Open-market operations The buying and selling of government securities by the central bank to the private sector financial markets to impact on the Monetary Base and the 'cash rate'. (Chapter 16)

Opportunity cost The best alternative sacrificed for a chosen alternative. (Chapter 2)

Over-confidence bias People exhibiting greater subjective confidence in themselves than the objective accuracy of their judgement, resulting in over-estimation of positive abilities and under-estimation of negative ones. (Chapter 10)

Overnight cash rate (cash rate) The interest rate that large-scale borrowers – including banks – in the financial markets must pay to borrow funds overnight. (Chapter 16)

Peak The point in the business cycle at which the economy reaches its maximum after rising during an expansion. (Chapter 12)

Perfect competition A market structure characterised by: (1) a large number of small firms, (2) a homogeneous product and (3) very easy entry into or exit from the market. (Chapter 7)

Perfectly competitive industry's short-run market supply curve The supply curve derived from the horizontal summation of the short-run supply curves of all firms in the industry. (Chapter 7)

Perfectly elastic demand A condition in which a small percentage change in price brings about an infinite percentage change in quantity demanded. (Chapter 5)

Perfectly inelastic demand A condition in which the quantity demanded does not change as the price changes. (Chapter 5)

Policy determination lag The time taken to decide on an apparently appropriate policy response; may be several months. (Chapter 16)

Policy effectiveness lag The delay between the time when a policy is initiated and the time when it begins to take effect on economic activity. (Chapter 16)

Positive economics An analysis limited to statements that are verifiable. (Chapter 1)

Precautionary motive for the demand for money The explanation for the stock of money people hold to pay unpredictable expenses. (Chapter 15)

Price ceiling A legally established maximum price a seller can charge. (Chapter 4)

Price discrimination This occurs when, for the same product, a seller charges different customers different prices not justified by cost differences. (Chapter 8)

Price elasticity of demand The ratio of the percentage change in the quantity demanded of a product to a percentage change in its price. (Chapter 5)

Price elasticity of supply The ratio of the percentage change in the quantity supplied of a product to the percentage change in its price. (Chapter 5)

Price floor A legally established minimum price a seller can be paid. (Chapter 4)

Price leadership A pricing strategy in which a dominant firm sets the price for an industry and the other firms follow. (Chapter 9)

Price maker A firm that faces a downward-sloping demand curve and can therefore choose among price and output combinations along the demand curve. (Chapter 8)

Price system A mechanism that uses the forces of supply and demand to determine the equilibrium price and output. The price system serves the function of resource allocation and rationing of output. (Chapter 3)

Price-taker A seller that has to accept or take the prevailing market price of its product. (Chapter 7)

Producer surplus The gain by the producer in an exchange. At the equilibrium market price, the producer surplus is represented in the area above the supply curve and below the market equilibrium price. (Chapter 3)

Product differentiation The process of creating real or apparent differences between goods and services. (Chapter 9)

Production function The relationship between the maximum amounts of output a firm can produce and various quantities of inputs. (Chapter 6)

Production possibilities frontier Shows the maximum combinations of two outputs that an economy can produce, given its available resources and technology. (Chapter 2)

Profit maximisation in perfect competition Occurs when $MR = MC$. (Chapter 7)

Protectionism The government's use of embargoes, tariffs, quotas and other restrictions to protect domestic producers from foreign competition. (Chapter 18)

Public good A good or service that is non-rival and non-excludable. (Chapter 4)

Quantitative easing The process whereby the central bank expands its balance sheet by buying long-dated government and corporate bonds from the financial sector. In paying for these purchases the central bank expands the Monetary Base. The primary goal of the strategy is to significantly expand the available liquidity in the financial system. (Chapter 16)

Quantity theory of money The theory that changes in the price level (inflation) are directly related to changes in the growth of the money supply. (Chapter 16)

Quota A limit on the quantity of a good that may be imported in a given time period. (Chapter 18)

Rational choice model A framework of studying behaviour by assuming human beings make consistent, rational choices to advance their self-interest. (Chapter 10)

Real balances or wealth effect The impact on total spending (real GDP) of the inverse relationship between the price level and the real value of financial assets with fixed nominal value. (Chapter 14)

Real GDP The value of all final goods and services produced during a given time period based on the prices existing in a selected reference year. (Chapter 11)

Real income The actual number of dollars received (nominal income) adjusted for changes in the average price level as measured by, for example, the CPI. (Chapter 13)

Real interest rate The nominal rate of interest minus the inflation rate. (Chapter 13)

Recession A downturn in the business cycle in which output, sales and employment decline. (Chapter 12)

Reserve Bank Board The RBA Board determines monetary policy in Australia. It consists of the Governor, the Deputy Governor, the Secretary to the Commonwealth Treasury and six other members appointed by parliament. (Chapter 15)

Reserve Bank of Australia (RBA) The central bank of Australia, responsible for monetary policy, the payments system and financial system stability, and providing banking services to the banks, other financial institutions and the federal government. (Chapter 15)

Scarcity The condition in which human wants are forever greater than the available supply of resources. (Chapter 1)

Seasonal unemployment Unemployment caused by recurring changes in hiring due to recurring changes in weather conditions, demand and/or production patterns. (Chapter 13)

Securitisation The practice of bundling up a large number of similar financial assets – say, housing loans – into a new derivative financial asset (or 'security'). (Chapter 19)

Short run A period of time during which there is at least one fixed input. (Chapter 6)

Slope The ratio of the changes in the variable on the vertical axis (the rise or fall) to the change in the variable on the horizontal axis (the run). (Chapter 1)

Smoothing and testing The RBA regularly operates in the market if it considers the AUD is coming under undue short-run speculative pressure ('testing'), or if Forex market conditions are excessively volatile 'smoothing'). (Chapter 16)

social surplus The total gain by the society in an exchange. Social surplus is the sum of consumer surplus and producer surplus. (Chapter 3)

Solow model of economic growth An early growth model that sought to explain how consumption, saving, capital, labour and technological change combine in the longer term to determine a nation's economic growth. (Chapter 12)

Speculative motive for the demand for money The explanation for the stock of money people hold to take advantage of expected future financial asset price changes. (Chapter 15)

Spending multiplier The induced rounds of spending that occur in an economy after some initial stimulus to spending. (Chapter 14)

Stagflation The condition that occurs when an economy experiences the twin maladies of increasing unemployment and rising inflation simultaneously. (Chapter 14)

Statement on the Conduct of Monetary Policy Signed in 1996, it encapsulates the Australian government's agreement that the RBA should have the goal of keeping inflation on average to between 2 and 3 per cent per annum over the course of the business cycle. In 2016, the language was adjusted a little to that of the goal being to keep consumer price inflation between 2 and 3 per cent, on average, over the medium term. (Chapter 16)

Status quo bias The preference to continue with the current activity (status quo) rather than making a change because of myopia or future uncertainty. (Chapter 10)

Stock A quantity measured at one point in time, such as a company's assets or the amount of money in a cheque account. (Chapter 11)

Store of value The ability of money to hold value over time. (Chapter 15)

Strategic trade theory The idea that governments should seek to be strategic in their use of their spending and taxing powers in order to actively facilitate some sectors of the economy that they may feel have the potential to be strong export earners for their nations. (Chapter 18)

Structural unemployment Unemployment caused by a mismatch of the skills of workers out of work and the skills required for existing job opportunities. (Chapter 13)

Sub-prime mortgage A housing loan extended to a borrower assessed by the originator of the loan as having a relatively higher level of difficulty in paying it off. (Chapter 19)

Substitute goods Goods that compete with each other for consumer purchases. As a result, there is a direct relationship between a price change for one good and the demand for its 'competitor' good. (Chapter 3)

Sunk cost bias The choice to continue with an activity because of investment in the past even though the invested amount is not recoverable. (Chapter 10)

Supply-side fiscal policy A fiscal policy that emphasises government policies that increase aggregate supply in order to achieve long-run growth in real output and full employment (along with a possibly lower price level). (Chapter 17)

Tacit collusion An informal arrangement for firms to fix prices. (Chapter 9)

Tariff A tax on an import that increases the domestic price of the imported product. (Chapter 18)

Tax incidence The share of a tax ultimately paid by consumers or by sellers. (Chapter 5)

Technology The body of knowledge applied to how goods and services are produced. (Chapter 2)

Terms of trade (TOT) Defined as the ratio of an index of a country's export prices to an index of its import prices. (Chapter 18)

Total cost The sum of total fixed cost and total variable cost at each level of output. (Chapter 6)

Total fixed cost Costs that do not vary as output varies and that must be paid even if output is zero. These are payments that the firm must make in the short run, regardless of the level of output. (Chapter 6)

Total revenue The total number of dollars a firm earns from sales of a good or service; equal to its price multiplied by the quantity demanded. (Chapter 5)

Total variable costs Costs that vary as output varies. (Chapter 6)

Transactions motive for the demand for money The explanation for the stock of money people hold to pay everyday predictable expenses. (Chapter 15)

Transfer payment A government payment to individuals that does not represent a payment in exchange for currently produced goods or services. (Chapter 11)

Treasury note (T-note) A T-note is sold by the Treasury on behalf of the federal government and is considered riskless. (Chapter 15)

Troubled Asset Relief Program (TARP) Announced in October 2008 by the US Treasury announced, costing US$700 billion; aimed at trying to bail out a range of troubled financial companies whose problems it deemed to be relatively short-term in nature. (Chapter 19)

Trough The point in the business cycle at which the economy reaches its minimum after falling during a recession. (Chapter 12)

Underemployment Another reason why some argue the measured unemployment rate understates the degree of labour unemployment in the economy. It is defined as those employed who would nonetheless like to work more hours if they could, expressed as a percentage of the labour force. (Chapter 13)

Unemployment rate The percentage of people in the labour force who are without jobs and are actively seeking jobs. (Chapter 13)

Unit of account The function of money to provide a common measurement of the relative value of goods and services. (Chapter 15)

Unitary elastic demand A condition in which the percentage change in quantity demanded is equal to the percentage change in price. (Chapter 5)

Variable input Any resource for which the quantity can change during the period of time under consideration. (Chapter 6)

Velocity of money The average number of times per period a dollar of the money supply is spent on final goods and services. (Chapter 16)

Vertical (classical) range The vertical segment of the aggregate supply curve, which represents an economy at full-employment output. (Chapter 14)

Wage–price spiral A situation that occurs when increases in nominal wage rates are passed on in higher prices, which in turn result in even higher nominal wage rates and prices. (Chapter 13)

Wealth The value of the stock of assets owned at some point in time. (Chapter 13)

What, how and for whom questions Fundamental questions about the composition of a nation's output, the ways in which this output should be produced and the way in which the benefits of this output should be distributed among the nation's citizens. (Chapter 2)

World Coincident Index An index of the current state of 'world' economic activity constructed from the coincident indexes for the US, UK, Japan, Germany, India and China maintained by the Economic Cycle Research Institute (ECRI). (Chapter 19)

World Leading Index Constructed from the leading indexes of 19 countries covered by the Economic Cycle Research Institute (ECRI). (Chapter 19)

Index

C

D

E

F

G

H

N

O

P

Q

R

S

T

U

V

W

X

Y

Z