

**Modern Monetary Theory and Practice:  
An Introductory Text**

*William Mitchell, L. Randall Wray and Martin Watts*

© William Mitchell, L. Randall Wray and Martin Watts, 2016

First published in 2016,  
Centre of Full Employment and Equity (CofFEE)  
The University of Newcastle  
Callaghan, NSW 2308, Australia

First Printed by CreateSpace 2016  
An Amazon company  
[www.createspace.com](http://www.createspace.com)

# Table of Contents

## **Chapter 1: Introduction**

- 1.1 What Is Economics? Two Views
- 1.2 What is Macroeconomics?
- 1.3 Macro and the Public Purpose

Appendix: Advanced Material

## **Chapter 2: How to Think and Do Macroeconomics**

- 2.1 Introduction
- 2.2 Thinking in a Macroeconomic Way
- 2.3 What Should a Macroeconomic Theory be Able to Explain?

Appendix

## **Chapter 3: A Brief Overview of Economic History and the Rise of Capitalism**

- 3.1 Introduction
- 3.2 An Introduction to Monetary Capitalism
- 3.3 Tribal Society
- 3.4 Slavery
- 3.5 Feudalism
- 3.6 Revolts and the Transition to Capitalism
- 3.7 Capitalism
- 3.8 Monetary Capitalism
- 3.9 Global Capitalism
- 3.10 Economic Systems of the Future?

## **Chapter 4: The System of National Income and Product Accounts**

- 4.1 Measuring National Output
- 4.2 Components of GDP
- 4.3 Equivalence of Three Measures of GDP
- 4.4 GDP versus GNP
- 4.5 Measuring Gross National Income
- 4.6 GDP Growth and The Price Deflator
- 4.7 Measuring CPI inflation
- 4.8 Measuring National Inequality

Appendix: Advanced Material

## **Chapter 5: Sectoral Accounting and the Flow of Funds**

- 5.1 Introduction
- 5.2 The Sectoral Balances View of the National Accounts
- 5.3 Revisiting Stocks and Flows
- 5.4 Integrating NIPA, Stocks, Flows and the Flow of Funds Accounts
- 5.5 Balance Sheets
- 5.6 The Flow of Funds Matrix

## **Appendix**

## **Chapter 6: Introduction to Sovereign Currency: The Government and its Money**

- 6.1 Introduction
- 6.2 The National Currency (Unit of Account)
- 6.3 Floating versus Fixed Exchange Rate Regimes
- 6.4 IOUs Denominated in National Currency: Government and Non-Government
- 6.5 Use of the Term 'Money': Confusion and Precision

## **Chapter 7: The Real Expenditure Model**

- 7.1 Introduction
- 7.2 Aggregate Supply
- 7.3 Aggregate Demand
- 7.4 Private Consumption Expenditure
- 7.5 Private Investment
- 7.6 Government Spending
- 7.7 Net Exports
- 7.8 Total Aggregate Expenditure
- 7.9 Equilibrium National Income
- 7.10 The Expenditure Multiplier

## **Chapter 8: Introduction to Aggregate Supply**

- 8.1 Introduction
- 8.2 Some Introductory Concepts
- 8.3 Price Determination
- 8.4 The General Aggregate Supply Function
- 8.5 Some Properties of the General Aggregate Supply Function (AS)
- 8.6 Factors Affecting Aggregate Output per Hour

## **Chapter 9: Labour Market Concepts and Measurement**

- 9.1 Introduction
- 9.2 Measurement
- 9.3 Categories of Unemployment
- 9.4 Broad Measures of Labour Underutilisation
- 9.5 Flow Measures of Unemployment
- 9.6 Duration of Unemployment
- 9.7 Hysteresis

Appendix: Advanced Material

## **Chapter 10: Money and Banking**

- 10.1 Introduction
- 10.2 Some Definitions
- 10.3 Financial Assets
- 10.4 What do Banks Do?

Appendix: Advanced Material

## **Chapter 11: Unemployment and Inflation**

- 11.1 Introduction
- 11.2 What is Inflation?
- 11.3 The Quantity Theory of Money
- 11.4 The Phillips Curve
- 11.5 The Accelerationist Hypothesis and the Expectations-Augmented Phillips Curve
- 11.6 Hysteresis and the Phillips Curve Trade-Off
- 11.7 Underemployment and the Phillips Curve
- 11.8 Demand-Pull and Cost-Push Inflation
- 11.9 The Conflict Theory of Inflation
- 11.10 Raw Material Price Increases
- 11.11 Cost-Push and Demand-Pull Inflation Summary
- 11.12 Incomes Policies

Appendix: Advanced Material

## **Chapter 12: Full Employment Policy**

- 12.1 Introduction
- 12.2 Full Employment as a Policy Goal
- 12.3 Unemployment Buffer Stocks and Price Stability
- 12.4 Employment Buffer Stocks and Price Stability
- 12.5 Alternative Policies for Promotion of Employment

Appendix: Advanced Material

## **Chapter 13: Introduction to Monetary and Fiscal Policy Operations**

- 13.1 Introduction
- 13.2 The Central Bank
- 13.3 The Treasury
- 13.4 Coordination of Monetary and Fiscal Operations
- 13.5 Taxes and Sovereign Spending
- 13.6 Currency Sovereignty and Policy Independence

Appendix: Advanced Material

## **Chapter 14: Fiscal Policy in Sovereign Nations**

- 14.1 Introduction
- 14.2 Functional Finance versus Sound Finance
- 14.3 Fiscal Policy Debates: Crowding Out and (Hyper)Inflation
- 14.4 The Debt Sustainability Debate
- 14.5 Conclusion: MMT and Fiscal Policy

## **Chapter 15: Monetary Policy in Sovereign Nations**

- 15.1 Introduction
- 15.2 Modern Banking Operations
- 15.3 Interest Rate Targets versus Monetary Targets
- 15.4 Liquidity Management
- 15.5 Implementation of Monetary Policy
- 15.6 Unconventional Forms of Monetary Policy
- 15.7 Monetary Policy in Practice
- 15.8 The Advantages and Disadvantages of Monetary Policy
- 15.9 Central Bank Independence
- 15.10 Horizontal and Vertical Operations: An Integration

**Chapter 16: Policy in an Open Economy: Exchange Rates, Balance of Payments and Competitiveness**

- 16.1 Introduction
- 16.2 The Balance of Payments
- 16.3 Essential Concepts
- 16.4 Aggregate Demand and the External Sector – Revisited
- 16.5 Trade in Goods and Services, Product Market Equilibrium and the Trade Balance
- 16.6 Currency Crises
- 16.7 Capital Controls

Appendix: Advanced Material

**Appendix: Methods, Tools and Techniques**

- A.1 Overview
- A.2 Basic Rules of Algebra
- A.3 A Simple Macroeconomic Model
- A.4 Graphical Depiction of a Macroeconomic Model
- A.5 Power Series Algebra and The Expenditure Multiplier
- A.6 Index Numbers
- A.7 Annual Average Growth Rates
- A.8 Textbook Policy Regarding Formalism

# Table of Tables

<b>Number</b>	<b>Heading</b>
Table 2.1	Average annual real GDP growth by decades, per cent
Table 2.2	Average unemployment rates by decade, per cent
Table 4.1	Items in Australian CPI, September 2013
Table 4.2	Hypothetical data for basket of goods and services
Table 4.3	Gini coefficients for several OECD nations, 2004 and 2012
Table 7.1	Consumption ratios, OECD nations, 2005-2010, per cent
Table 7.2	Expenditure chain volume measures in national accounts (Australia)
Table 7.3	The expenditure multiplier process
Table 7.4	Simulating changes in the multiplier components
Table 9.1	OECD underemployment, per cent of labour force, 1990 to 2014
Table 9.2	Labour market flows matrix
Table 9.3	Gross flows in the US labour market, December 2015–January 2016, millions
Table 9.4	Total inflow and outflow from labour force states, US, December 2015 to January 2016, millions
Table 9.5	Labour market state transition probabilities, US, December 2015 to January 2016
Table 15.1	Target interbank rates for developed economies
Table 16.1	Australian balance of payments, various years, current prices
Table 16.2	Comparison of international prices
Table A.1	Employment in Australia, 2000-12, thousands
Table A.2	Employment indices for Australia, 2000-12
Table A.3	Compound growth rate calculations



# Table of Figures

<b>Number</b>	<b>Heading</b>
Figure 2.1	Comparative unemployment rates, per cent, 1960 to 2015
Figure 2.2	Real wage and productivity indexes, Australia, 1971 to 2015, March 1982=100
Figure 2.3	Household debt to disposable income ratio, Australia, per cent, 1998 to 2015
Figure 2.4	US Federal Reserve Bank monetary base, 1959 to 2015, US dollar millions
Figure 2.5	Government fiscal balance as % of GDP, Japan, 1980 to 2015
Figure 2.6	Gross and net public debt as % of GDP, Japan, 1980 to 2015
Figure 2.7	Japan overnight interest rate, per cent, July 1985 to December 2015
Figure 2.8	Japan government 10-year government bond yield, per cent, 1990 to 2015
Figure 2.9	Inflation and deflation in Japan, per cent, 1980 to 2015
Figure 4.1	The Lorenz curve
Figure 5.1	UK sectoral balances, 1960 to 2014
Figure 5.2	A stylised sectoral balance sheet
Figure 5.3	A uses-and-sources-of-funds statement
Figure 5.4	A complete sector uses-and-sources-of-funds statement
Figure 5.5	A stylised three sector flow-of-funds matrix
Figure 5.6	A graphical sectoral balances framework
Figure 5.6a	Deriving the private domestic sector balances
Figure 5.6b	Private domestic surpluses and deficits
Figure 5.7	Sustainable space for sovereign governments
Figure 5.8	Sustainable space for governments constrained by fiscal rules
Figure 6.1	The Minsky – Foley pyramid
Figure 7.1	Aggregate supply
Figure 7.2	The consumption function
Figure 7.3	The aggregate demand function
Figure 7.4	An increase in the intercept of the aggregate demand function
Figure 7.5	Changing slope of the aggregate demand function
Figure 7.6	Planned expenditure and equilibrium income
Figure 7.7	The multiplier process
Figure 7.8	Impact of a change in government spending on equilibrium expenditure and income

- Figure 7.9 Impact of a change in the marginal propensity to consume on equilibrium expenditure and income
- Figure 8.1 The employment-output function
- Figure 8.2 Output, sales and national income
- Figure 8.3 The general aggregate supply function (AS)
- Figure 8.4 US manufacturing output per person employed
- Figure 8.5 US manufacturing output per hour of all persons
- Figure 9.1 The labour force framework
- Figure 9.2 Labour force participation rate and annual real GDP growth – Australia, 1980 to 2015, per cent
- Figure 9.3 Unemployment rate and average duration of unemployment (weeks), US, February 2008 to October 2012
- Figure 10.1 US Treasury yield curve (February 3, 2016)
- Figure 10.2 A typical bank balance sheet
- Figures 10.3, 10.4, 10.5, 10.7, 10.9 & 10.11  
Bank A balance sheets
- Figures 10.6, 10.10 & 10.12  
Bank B balance sheet
- Figures 10.8 & 10.13  
Central bank balance sheet
- Figure 11.1 Velocity of M2 money stock (M2V), US, 1950-2012
- Figure 11.2 The basic Phillips curve
- Figure 11.3 The unemployment-inflation choice set
- Figure 11.4 The shifting US Phillips curve - 1948-2012
- Figure 11.5 The expectations-augmented long-run Phillips curve
- Figure 11.6 Annual Australian unemployment rate, Treasury and OECD NAIRU estimates, 1960-2015
- Figure 11.7 Inflation and unemployment, Australia, quarterly data, 1978-2015
- Figure 11.8 Inflation and unemployment and underemployment, Australia, quarterly data, 1978-2013
- Figure 11.9 Short- and long-run Phillips curves
- Figure 12.1 The sacrifice ratio and disinflation episode
- Figure 12.2 Sacrifice ratios with persistence and hysteresis
- Figure 12.3 The Job Guarantee and the Phillips curve

- Figure 13.1 Balance sheets associated with net government spending
- Figure 13.2 Balance sheets associated with net government spending
- Figure 15.1 Unemployment rate and inflation rates for Australia, 1980 – 2015, per cent
- Figure 15.2 Vertical and horizontal macroeconomic relations
- Figure 15.3 Horizontal and vertical components of the money supply
- Figure 16.1 A simple bi-lateral foreign exchange market
- Figure 16.2 Net exports as a function of real national income
- Figure 16.3 Equilibrium national income with a change in world income
- Figure 16.4 Mexican peso depreciation, December 1994
- Figure 16.5 Real GDP growth in South East Asia, 1980-2005, per cent per annum
- Figure A.1 Consumption function
- Figure A.2 Slope of consumption function
- Figure A.3 Employment index numbers, Australia, 2000-2012, 2000=100
- Figure A.4 Real wages and productivity, 1978- 2012, March-quarter 1978=100

# Common Abbreviated Terms

## Abbreviation Term

BER	Buffer Employment Ratio
BLS	Bureau of Labor Statistics
CD	Certificate of Deposits
CPI	Consumer Price Index
<i>C</i>	Consumption
CAB	Current Account Balance
EMU	Economic and Monetary Union (EMU) of the European Union
ERM	Exchange Rate Mechanism (European)
EAPC	Expectations-Augmented Phillips Curve
<i>X</i>	Exports
FOMC	Federal Open Market Committee (US)
GE	General Equilibrium
GT	General Theory (Keynes)
GFC	Global Financial Crisis (2008)
GBC	Government Budget Constraint
<i>G</i>	Government Spending
<i>GDP or Y</i>	Gross Domestic Product
GNI	Gross National Income
GNP	Gross National Product
<i>S</i>	Gross Saving
<i>M</i>	Imports
$M_h$	Stock of High Powered Money
$M_s$	Stock of Money
<i>Y</i>	Income (if not used for Gross Domestic Product)
ISLM	Interest-Saving Liquidity Preference-Money Supply Model (Hicks)
<i>I</i>	Investment
JGBs	Japanese Government Bonds
JG	Job Guarantee also known as Employer of Last Resort program
MRU	Macro-equilibrium unemployment rate
MPC	Marginal Propensity to Consume

<i>m</i>	Marginal Propensity to Import
MMT	Modern Monetary Theory
NIPA	National Income and Product Accounts
NRU	Natural Rate of Unemployment
NAFA	Net Acquisition of Financial Assets
<i>NX</i>	Net Exports
FNI	Net external income flows
NI	Net National Income
NNP	Net National Product
NC	New Classical
<i>e</i>	Nominal exchange rate
NAIRU	Non-Accelerating Inflation Rate of Unemployment
NGO	Non-Governmental Organisation
OMO	Open Market Operations
OECD	Organisation for Economic Co-operation and Development
OPEC	Organisation of Petroleum Exporting Countries
OER	Owners' Equivalent Rent
PDI	Person Disposable Income
<i>PI</i>	Personal Income
<i>PS</i>	Personal Saving
<i>P</i>	Profits
QE	Quantitative Easing
QTM	Quantity Theory of Money
RATEX	Rational Expectations hypothesis
RBC	Real Business Cycle
<i>R</i>	Real exchange rate
ROW	Rest of World
SM	Scandinavian Model
SRPC	Short-Run Phillips Curve
<i>T</i>	Taxes
<i>W</i>	Wages
YC	Yield Curve

# About the Authors

## ***Professor William Mitchell***

William Mitchell is a Professor of Economics at the University of Newcastle, NSW, Australia and inaugural Director of the Centre of Full Employment and Equity (CofFEE), an official research centre at the University. He has held that position since the centre's inception in 1998. He is the co-founder of CofFEE-Europe, a sister centre at Maastricht University, the Netherlands, where has been a visiting professor since 1992.

He has published widely in refereed academic journals and books in the areas of macroeconomics, labour economics and spatial econometrics and regional analysis. He has received regular research grant support from national competitive grants schemes operated through the highly competitive Australian Research Council. He also has extensive experience as a consultant to the Australian government, trade unions and community organisations, and several international organisations (including the European Commission; the International Labour Organisation and the Asian Development Bank).

His latest book *European Dystopia: Groupthink and Denial on a Grand Scale*, was published by Edward Elgar (UK) in May 2015. The Spanish translation will be published in May 2016 and a Finnish translation in November 2016.

He is on the editorial board of the *Journal of Post Keynesian Economics*, the *Economic and Labour Relations Review*, and the *Australian Journal of Labour Economics*.

He writes a daily economics blog covering international and national issues, which enjoys a high profile. It was ranked 43rd among the top 200 Top 200 Influential Economics Blogs in the World in 2013, when the index was last compiled.

## ***Professor L. Randall Wray***

L. Randall Wray is Levy Economics Institute Senior Scholar and a professor of economics at Bard College. His current research focuses on providing a critique of orthodox monetary theory and policy, and the development of an alternative approach. Wray is the author of *Money and Credit in Capitalist Economies*, 1990; *Understanding Modern Money: The Key to Full Employment and Price Stability*, 1998; and *Modern Money Theory: A Primer on Macroeconomics for Sovereign Monetary Systems*, 2012 (2nd rev. ed., 2015). He is also co-editor of, and a contributor to, *Money, Financial Instability, and Stabilization Policy*, 2006, and *Keynes for the 21st Century: The Continuing Relevance of The General Theory*, 2008. He is the Co-editor of the *Journal of Post Keynesian Economics* and his most recent book is *Why Minsky Matters: An Introduction to the work of a maverick economist*, Princeton University Press. Wray taught at the University of Missouri–Kansas City from 1999 to 2016 and at the University of Denver from 1987 to 1999, and has been a visiting professor at the Universities of Paris, Bergamo, Bologna, and Rome (La Sapienza). He holds a BA from the University of the Pacific and an MA and a Ph.D. from Washington University, where he was a student of Hyman Minsky.

***Professor Martin Watts***

Martin Watts is Research Associate of the Centre of Full Employment and Equity (CofFEE) and Professor of Economics at the University of Newcastle, Australia. He has had a long term interest in contemporary macroeconomic theory and policy, and the development of Modern Monetary Theory as an alternative perspective. In recent years has published in a number of heterodox economics journals, including the *Journal of Post Keynesian Economics*, *Economic Issues*, *Journal of Australian Political Economy* and *The Economic and Labour Relations Review*. He is on the Editorial Board of the *Journal of Post Keynesian Economics* and *Gender, Work and Inequality*. His other areas of research include the conceptualisation and measurement of segregation and spatial interaction models of commuting behaviour. Watts taught at Monash University (1975-1990) and commenced employment at Newcastle University in 1991. He holds a B.A. from Essex University, an M.A. from Manchester University and a PhD from the University of British Columbia.

# Chapter 1: Introduction

## Chapter Outline

### 1.1 What is Economics? Two Views

- Orthodox, neoclassical approach
- Heterodox; Keynesian – institutionalist – Marxist approach
- What do economists do?
- Implications for research and policy

### 1.2 What is Macroeconomics?

- The MMT approach to macroeconomics
- The macro model
- Fiscal and monetary policy
- Policy implications for sovereign nations

### 1.3 Macro and the Public Purpose

- Concluding thoughts on the public purpose

### Appendix: Advanced Material

- 1. The minimum wage debate
- 2. The structure of scientific revolutions

## Learning Objectives

1. Recognise that macroeconomics is a contested discipline with two broad schools of thought, which differ in terms of their perspectives on the effectiveness of markets and the role of Government.
2. Understand that macroeconomics analyses the behaviour of aggregates, such as employment, unemployment, GDP and inflation, whereas microeconomics studies the behaviour of individual economic agents, notably households and firms.
3. Acknowledge that social science disciplines (e.g economics and politics) and physical science disciplines (e.g. physics and chemistry) each have their own language in the form of concepts and theories, which provide the basis for understanding, not merely describing, relevant phenomena.



## 1.1 What Is Economics? Two Views

US President Harry Truman is said to have sought a one-armed economist because he was so frustrated by the propensity of economists to provide policy advice framed as ‘Well, on the one hand, X, but on the other hand, Y’, where ‘Y’ typically would be the precise opposite policy path to ‘X’.

The story is, of course, funny but it does bring up a problem that is ubiquitous to all social sciences. Even if we know the result we would like to achieve (say, smarter and happier kids), we do not know with certainty which policy choices would produce the desired outcome. Since the main topic of the social sciences—human behaviour—is complex, we often do not understand its causes, or even its nature, and much less do we know how to influence it in a desired manner. Economics is as difficult as the other social sciences, such as psychology and political science as it concerns human behaviour in a social sphere that we designate as ‘the economy’, which itself is hard to define and to delineate from other spheres of social interaction.

Unfortunately, economics is sometimes equated to something like the ‘study of business decision-making’, or even relegated to a narrow sub-discipline as a ‘decision science’ in a highly artificial hypothesised world of hyper-rational automatons that maximise pleasure and avoid pain. Some even see economics as just a branch of mathematics, a view fuelled in part by the heavy use of mathematics and models in much of the discipline.

This textbook will take a broader perspective of the economics discipline, including it within the social sciences. While we do think it is useful to carve off ‘the economy’ from the rest of social life, and to apply ‘economics’ to the study of that area of life, we recognise that the division is necessarily arbitrary. In truth, there is no completely separate sphere of ‘economic life’—so economics is linked to, and incorporates findings from, the other social science disciplines.

Further, we want to stress that there is no single ‘right’ way to do economics. In this textbook we will use a variety of methods and approaches to build our understanding of ‘the economy’. We will occasionally bring in research and methods from other disciplines. We will use some mathematics and modelling. As we believe that economic history as well as history of economic thought helps us to understand our economy today, we will look back in time, both in terms of economic events, but also to examine the insights of the great thinkers of the past.

In the rest of this section we will briefly outline the two main approaches to economics taken by those thinkers, as well as by today’s economists. It is always risky to pigeon-hole individuals and their theories into categories. Just as a politician in a particular political party (say, Labor in Australia, or the Republican Party in America) will hold many views shared by most members of that party, they will likely also hold some views more consistent with those of a rival party. This is true of economists, too. Still, it is useful to identify two approaches to economics that have dominated much of the debate over the past two centuries.

Recalling the story about President Truman’s frustration, we can think of the ‘two hands’ of economics as the orthodox, or neoclassical approach and the Heterodox or Keynesian/Institutionalist/Marxist approach. Let us examine each in turn, while recognising that we must generalise.

## Orthodox, neoclassical approach

In the neoclassical approach, there is a presumed, natural human nature: individuals maximise pleasure and avoid pain. Pleasure is defined as ‘utility’, so individuals pursue utility maximising behaviour, avoiding the ‘disutility’ of pain. Further, rational individuals are self-interested - seeking to maximise their own utility, and they do not receive either utility or disutility from the experiences of others. Neoclassical economics presumes individuals are ‘rational’, meaning they maximise utility given constraints. If there were no constraints, individuals would maximise with infinite utility—but they are constrained by their resources that they have at their disposal, which are referred to as ‘individual resource endowments’. Mutually beneficial exchange redistributes resources according to preferences, increasing the utility of both parties to the trade.

In the hypothesised free market, exchanges take place at competitively determined relative prices. (Relative prices are ratios; for example: 1 deer = 3 beavers = 6 rabbits = 2 bushels of wheat = 10 hours of labour services.) Participants in markets take relative prices as signals. Relative scarcity will cause the price to rise, inducing suppliers to produce more of a particular traded commodity, and buyers to demand less. For example, if the supply of students trained in economics is insufficient to meet the demand for economists, the relative wage of economists to (say) that of historians, rises. This signals to students that they ought to switch from the study of history to the study of economics. At the same time, employers try to find close but cheaper substitutes—say, political science students. As the supply of economists increases, the relative wage advantage for students trained in economics falls. Of course, other factors enter into decisions, but the important point is that relative prices function as signals to both suppliers (economics students) and demanders (employers of economists).

Equilibrium is defined as the set of relative prices that ‘clear’ markets; a ‘general equilibrium’ is a complete set of prices to clear all markets. One interpretation of Adam Smith’s famous ‘invisible hand’ analogy is that by producing market-clearing prices, the market provides the signals that guide individuals to maximise their utility while also providing the social or public good of ensuring that demand and supply are equilibrated. The hand is ‘invisible’, guiding individuals and the economy as a whole toward equilibrium, with no need of an authority. For that reason, there is little need for government management of the economy.

Certainly government has some role to play in setting and enforcing rules, in providing national security, and (perhaps) for providing a social safety net. But according to this interpretation of Smith, there is no need for the government to direct individuals to serve the public interest because by reacting to price signals and pursuing their own interests, individuals actually act in the public interest.

There is one more important conclusion reached by neoclassical economics: ‘you deserve what you get’. If we all come to the free market to make mutually beneficial exchanges, all seeking to maximise our own individual utility subject to our resource constraints, then the equilibrium allocation is in an important sense ‘fair’. That does not mean that the allocation is equal—some will have more (and achieve greater utility) and others will have less. However, that is because some start with greater endowments (of resources, ability, and drive).

Technically, the idea is that one receives an allocation of resources based on one’s own contribution to the market. If your final allocation is low, it is because you did not bring enough to market: maybe you were born with few resources, you made a constrained choice to obtain

little education, and you prefer leisure over work. In other words, you have no one to blame for your meagre allocation but yourself.

To be sure, neoclassical economics also allows for bad luck, congenital disabilities, and so on. Hence, there is a role for social policy to get involved in altering the allocation in order to protect the poorest and least advantaged. However, generally speaking, allocations ought to be left to the market because it will reward each participant according to productive contributions to the market—a dimension of fairness.

In recent years, the neoclassical approach to economics has been invoked in support of the conservative backlash against post-war economic and social reforms in Western nations (this is generally called neoliberal outside the USA or neoconservative within the USA). This ‘anti-government’ movement is closely associated with the terms in office of President Ronald Reagan in the USA and Prime Minister Margaret Thatcher in the UK. When running for President in 1980, President Reagan promised to “get the government off the backs of the people”; Prime Minister Thatcher was famous for arguing that there is no such thing as society, reflecting the individualistic framework shared by neoclassical economics.

Downsizing government and especially reducing the social safety net, is consistent with the view that government only needs to ‘get the incentives right’, and then the ‘free market’ would maximise individual welfare while the invisible hand will ensure that signals coming from markets guide individuals to do what is best for the economy as a whole.

While the neoliberal/neoconservative policies are most closely associated with conservative political parties, even the moderate parties continued the policies throughout the 1990s and 2000s. For example, President Clinton (a Democrat) echoed President Reagan’s distaste for social welfare programs when he promised to “end welfare as we know it” in his 1992 election campaign, eliminating the biggest USA anti-poverty program (Aid to Families with Dependent Children) and replacing it with a term-limited program that tries to force aid recipients to work for their benefits (‘workfare’ rather than ‘welfare’).

Outside the USA the more left wing parties such as the Labour Party in the UK pursued similar strategies (such as ‘work for the dole’). Neoclassical economic theory provided a strong justification for these economic and social ‘reforms’ as policy would rely more heavily on ‘market outcomes’ while reducing ‘government interference’ into the workings of the ‘invisible hand’.

Finally, let us turn to the neoclassical definition of economics, as it provides a very nice summary of the approach taken.

**Neoclassical Definition of Economics:** the study of the allocation of scarce resources among unlimited wants.

This is often framed as ‘the economic problem’: while resources are scarce, our wants are unlimited. The ‘problem’ is that we cannot ever satisfy our wants. Many call economics ‘the dismal science’, which comes from this statement of ‘the problem’. While we all try to ‘maximise utility’, resource constraints prevent us from ever achieving maximal bliss.

Another common statement attributed to economists is that ‘there’s no such thing as a free lunch’, which also derives from the definition. Since resources are scarce, there is always a trade-off: if we move resources from one use to another, we necessarily reduce enjoyment in the first use in favour of enjoyment in the second. For example, if we want to have more ‘guns’, we must have less ‘butter’. Or if we want to improve the standard of living enjoyed by ‘Bob’, we must reduce the living standard of ‘Jill’.

Strictly speaking, this would be true only at full employment of all resources. However, with the invisible hand guiding the allocation of resources, flexible relative prices ensure that all scarce resources are fully employed. The idea is that prices will always fall until supply equals demand so that no resource is left idle.

Note also that the trade-off might only be temporary. For example, if we move resources out of the production of consumption goods and into the production of investment goods that raise productive capacity, then in the future we can have more consumer goods. Through economic growth we can increase the quantity of resources so that both ‘Bob’ and ‘Jill’ can have more. This does not violate the admonition that there is no free lunch, however. If we are to have more production in the future, we need to sacrifice some consumption today—we suffer today with lower consumption, but we are willing to endure the ‘pain’ on the promise that in the future we can enjoy more consumption.

We will have much more to say about the neoclassical approach later in the text. However, it is time to move on to the second approach to economics.

### **Heterodox, Keynesian-institutionalist-Marxist approach**

There is a second, long, tradition in economics that adopts a quite different framework. Unfortunately, there is no strong consensus about what to call it. Sometimes it is called ‘nonorthodox’, which appears to define it in opposition to ‘orthodox’ or neoclassical economics. In recent years, many of those working in this tradition have settled on the term ‘heterodox’, but that, too, is usually defined as ‘not in agreement with accepted beliefs’. Yet at one time, those views now associated with ‘heterodox’ economics were dominant, while the ‘orthodox’ views were considered by most economists as ‘unorthodox’ in the sense that they were not in agreement with the beliefs held by most economists!

Further, unlike neoclassical theory, which is substantially accepted by all orthodox theorists, ‘heterodoxy’ is made up of a number of well-established and coherent economic schools of thought<sup>1</sup>. While these share a common approach, they also deviate from one another in important ways. The three most important of these schools of thought are the Marxist (following the work of Karl Marx), the Institutionalist (following the work of Thorstein Veblen), and the Keynesian (followers of John Maynard Keynes)<sup>2</sup>.

---

<sup>1</sup> Note that the approach taken in this text, Modern Money Theory, falls within the heterodox camp. Indeed, it rests upon the foundations of many of the heterodox traditions.

<sup>2</sup> A caveat is necessary here. Many of those who call themselves ‘Keynesian’, as well as the approach that is often presented in economics textbooks as ‘Keynesian theory’, are not heterodox. They are much closer to the neoclassical approach. Indeed, one of the founders of orthodox macroeconomic theory, Paul Samuelson, called it the “Neoclassical Synthesis” to indicate that its foundations are neoclassical but some of Keynes’s ideas are

What are we to do? In spite of the objections we raised, we will conform to the convention and call this second approach the heterodox or Keynesian/Institutionalist/Marxist approach. Let us examine the shared framework adopted.

First, according to this approach there is no ‘natural’ human behaviour; rather, it is shaped by institutions, culture, and society. There is nothing ‘natural’ about self-interested (or, better, ‘selfish’) behaviour, nor would such behaviour be ‘rational’ in the neoclassical sense. Humans are social animals and in many cultures, selfish behaviour is punished and selfish individuals are ostracised. Since human survival requires cooperation, selfishness would actually be irrational as it would reduce one’s chances of survival. In any event, in all known societies, elaborate rituals and traditions are designed to promote cooperation and even sacrifice for the common good.

Human behaviour varies significantly across societies, and the economic system is one factor that helps to determine appropriate behaviour within any particular society. Self-interested behaviour is more acceptable in some societies than in others. It is not a coincidence that neoclassical economic theory was developed largely in Western capitalist societies—and particularly in England. The ‘rational’ behaviour attributed by neoclassical economists to all humans actually comes reasonably close as a description of the behaviour of early British capitalists. In the social environment in which they operated, pursuit of their own self-interest without regard to the welfare of others (especially that of their employees), may have increased their probability of success as capitalists. Further, they operated in a hostile political climate in which the Crown and their feudal lord cronies wanted to increase their own share of the nation’s rather feeble output. Government ‘intervention’ was almost always a bad thing, from the perspective of the first capitalists because government operated substantially in the interest of the Crown and the feudal lords.

We will not go into economic history now. What we wish to emphasise is that human behaviour is surprisingly malleable and complexly influenced by custom and tradition.

Further, most decisions are not ‘rational’ for another reason: the future is uncertain, and even the present and past are uncertain in the sense that we do not fully understand what happened and what is now happening. Clearly, we do not know the future, and we know that we do not. Hence, we cannot know for certain that any action we take is truly ‘utility maximising’. Should I buy the Renault or the Mazda? Once the decision is made and with the passage of time, I might have a better idea of the best choice, but it is probable that even a decade down the road I will not know which would have been best. Obviously, that choice is relatively unimportant and even simple compared to most economic choices one must make. In truth, we almost never know whether we are ‘maximising’ utility—indeed even with hindsight we often cannot tell if we made the right decision.

According to the heterodox approach, most decisions are not ‘rational’ in the neoclassical sense of the term. Decisions and behaviour depend on a range of other factors, including uncertainty, power, discrimination, prejudice, and segregation. Options available depend on status, social class, race, religion, and gender, for example. These ‘noneconomic’ factors heavily influence and even constrain our choices.

---

‘synthesised’ or grafted onto that base. Heterodox followers of Keynes argue that such integration is not possible. We will revisit these issues later in the text.

Heterodoxy rejects the notion that economic outcomes are arbitrated by an impersonal market that only seeks to equilibrate ‘demand and supply’. Actually, market prices are largely administered by firms with market power that allows them to discriminate. Wages are set not to ‘clear’ the labour market, but rather reflect the outcome of conflicted bargaining processes. Capitalism is a system defined by class conflict. In general, workers want to earn as much as they can for the effort they expend, while bosses want workers to produce as much as they can but pay them as little as possible. And, as will be discussed later, unemployment cannot be eliminated through wage reductions that eliminate relative excess labour supply; indeed wage reductions can actually reduce the demand for labour and thus increase unemployment. More generally, wages and other prices are not simply signals of the invisible hand, but rather determine incomes and thus influence business sales and decisions going forward. For that reason, price and wage determination are not usually left to the invisible hand of the market.

Heterodoxy holds a different view of the so-called ‘economic problem’ of scarce resources and unlimited wants. Wants are largely socially created, and there is nothing natural about humans having ‘unlimited’ wants. While it is true that modern advertising operates to continually expand our desires, this can be countered through education. Further, resources are also largely socially created. While it is true that some natural resources have a limited supply, innovations continually produce substitutes. For example, Western societies faced their first major energy crisis in the 19th century when whalers had significantly reduced the number of whales, the source of whale oil used for lighting and other purposes. However, the production of petroleum and then electricity quickly replaced whale oil.

Moreover, the most important resource in any economy is labour. Ironically, in capitalist economies labour is virtually always in excess supply—that is, many workers are left unemployed. It is ironic that neoclassical economics starts from the presumption that resources are scarce, when the obvious empirical fact is that labour is unemployed. Any theory that begins with the presumption that labour is always fully employed, and hence scarce, is ignoring a glaring inconsistency.

Let us look at the heterodox definition of economics.

**Heterodox Definition of Economics:** the study of social creation and social distribution of society’s resources.

Note that unlike the orthodox definition, this one focuses on the creation of resources. Further, most of that creation is social, rather than individual: people work together to produce society’s resources. Distribution, too, is socially determined, rather than determined by a technical relation (one’s contribution to the production process). For example, labour unions engage in collective bargaining with their employers, who also band together to keep wages low.

The political process is also important in determining distribution; not only does government directly provide income to large segments of society, but it also puts in place minimum wages, benefits, and working conditions that must be met by employers. Government is also a creator of resources; it is not just a user of them. It organises and funds innovative research and development (often in its own labs) that is then used to create resources (frequently by private firms). It also purchases directly from firms, encouraging them to increase hiring and output. Not

only do these government activities increase production, but they also affect distribution. This is well-understood by voters and their representatives in government as policy creates winners and losers—and not usually in a zero-sum manner: some policies can create winners while others might create more losers.

Power, discrimination, collusion, and cooperation all play a role in determining ‘who gets what’. The point is that society does not have to let ‘the market’ decide that women should be paid less than men, for example, or that those with less education should remain jobless and thus poor.

Economics, like all social sciences, is concerned with a society that is complex and continually undergoing change. Since economists study human behaviour in the economic sphere, their task is very difficult. Whatever humans do, they could have done something different. Humans have some degree of free will, and their behaviour is largely based on what they think they ought to do. That in turn depends on their expectations of an unknowable future—they do not know precisely what the outcome of their actions will be, and they do not know what others will do.

Indeed, humans do not know exactly what happened in the past, nor do they fully understand what is happening today. They must interpret the environment in which they live, and realise that they cannot fully understand it. They can never know if they have truly ‘maximised’ their pleasure. They make plans in conditions of existential uncertainty, and do the best they can do given their circumstances. Their actions are almost always taken with consideration given to the impacts on others—humans are above all social animals and that is why economics must be a branch of the social sciences.

### **What do economists do?**

Like sociologists and political scientists, economists are trying to understand particular aspects of human behaviour – for example decisions about levels and patterns of spending, choices about enrolment in post-school education and types of employment to pursue– which we argue above, are influenced by institutions; culture; and society; in addition to economic variables, such as income; the prices of goods; and prospective wage rates for different occupations. In microeconomics our focus is the behaviour of individual consumers and firms, whereas in macroeconomics the focus is the aggregate impacts of these decisions on outcomes, including total output and employment and the rate of inflation. We elaborate on these definitions of microeconomics and macroeconomics, below.

In trying to understand particular forms of economic behaviour, we need to develop theories that require us to decide on those factors that we think influence the particular economic decisions of interest. In other words, we need to make simplifying assumptions (engage in abstraction), which means we necessarily ignore those factors that we consider to be irrelevant. Otherwise we are trying to replicate the complex reality, as we see it, and we are engaging in description rather than theorising. In the development of theory, concepts are formulated, which can be viewed as the building blocks of theory. A model can be viewed as the formalisation of a theory (see below). To understand any theory (model), it is important that students comprehend the underlying concepts.

Social scientists seek to confront their abstract theoretical models, expressed in the form of conjectures about real world behaviour, with the empirical data that the real world provides. For example, we might form the conjecture that if disposable income rises, household consumption will rise. We would then collect the relevant data for disposable income and household

consumption and any other information we thought might bear on the relationship and use various statistical tools (for example, regression analysis) to enumerate the relationship between disposable income and household consumption to see whether our conjecture was data consistent. In engaging in this sort of exercise, the responsible social scientist is not seeking to establish whether the theoretical model is true, for that is an impossible task, given there is no way of knowing what the truth is anyway. Instead, we seek to develop theories or conjectures that provide the best correspondence with the empirical world we live in. This means our current, accepted body of knowledge comprises theories and conjectures that explain the real world data in the most comprehensive way when compared to the competing theories.

Further, we can rarely refute a theory. As President Truman complained, there are two or more sides to the most important economic questions, so there are competing theoretical approaches yielding different conclusions. Even when a researcher resorts to the analysis of relevant data, (which often entails the use of econometrics), they can never refute a theory with 100 per cent confidence. Often the acceptance of a theory is driven by ideology and politics, rather than a balanced assessment of the competing theories and associated evidence.

### **Implications for research and policy**

Many students, like President Truman, find the inability of economists to come up with definitive answers to economic questions to be rather frustrating. Here it is important to emphasise that, like physical sciences and other social sciences, economics is a contested discipline, as is illustrated by our brief discussion of the two schools of thought in Section 1.1. Students will be exposed to some major contemporary debates in macroeconomics later in this textbook, but below we outline a long-standing debate in developed economies, such as the UK, USA and Australia, about the impact of an increase in the minimum wage on unemployment (Advanced Topic 1 in the Appendix to this Chapter).

If there are longstanding debates in economics (and other disciplines), which appear to be unresolved, how can there be progress in our understanding of economic phenomena? This is an important question because decisions made by macroeconomic policymakers have profound effects on the welfare of the population in terms of for example, employment opportunities and wages. Thomas Kuhn developed a way of understanding how progress is made in the social and physical sciences; see Advanced Material 2 in the Appendix to this Chapter.

## **1.2 What is Macroeconomics?**

In macroeconomics we study the aggregate outcomes of economic behaviour. The word ‘macro’ is derived from the Greek word ‘makro’, which means large and so we take an economy-wide perspective.

Macroeconomics is not concerned with analysing how each individual person, household or business firm behaves or what they produce or earn – that is the terrain of the other major branch of economic analysis, microeconomics. Macroeconomics focuses on a selected few outcomes at the aggregate level and is rightly considered to be the study of employment, output and inflation in an international context. A coherent macroeconomic theory will provide consistent insights into how each of these aggregates is determined and change.

In this regard, there are some key macroeconomic questions that we seek to explore:



1. What factors determine the flow of total output produced in the economy over a given period and its growth over time?
2. What factors determine total employment and why does mass unemployment occur?
3. What factors determine the evolution of prices in the economy (inflation)?
4. How does the domestic economy interact with the rest of the world and what are the implications of that interaction?

A central idea in economics whether it is microeconomics or macroeconomics, is efficiency – getting the best out of what you have available. The concept is extremely loaded and is the focus of many disputes, some more arcane than others. But there is a general consensus among economists that at the macroeconomic level, the ‘efficiency frontier’ (which defines the best outcome achievable from an array of possible outcomes) is normally summarised in terms of full employment. The hot debate that has occupied economists for years is the exact meaning of the term – full employment. We will consider that issue in full in Chapters 11 and 12. But definitional disputes aside, it is a fact that the concept of full employment is a central focus of macroeconomic theory. Using the available macroeconomic resources including labour to the limit is a key goal of macroeconomics. The debate is over what the actual limit is. The related macroeconomic challenge is how to maintain full employment but at the same time achieve price stability, which means that prices are growing at a low and stable rate.

The clear point is that if you achieve that goal then you will be contributing to the prosperity and welfare of the population by ensuring real output levels are high within an environment of stable prices.

This book develops a framework for understanding the key determinants of these aggregate outcomes – the level and growth in output; the rate of unemployment; and the rate of inflation – within the context of what we call a monetary system. All economies use currencies as a way to facilitate transactions. The arrangements by which the currency enters the economy and the role that the currency issuer, the national government, has in influencing the outcomes at the aggregate level, is a crucial part of macroeconomics. Modern Monetary Theory (MMT), which is briefly outlined below, develops a macroeconomic framework that incorporates the unique features of the monetary system.

### **The MMT approach to macroeconomics**

Modern Monetary Theory (MMT) is distinguished from other approaches to macroeconomics because it places the monetary arrangements at the centre of the analysis. Learning macroeconomics from an MMT perspective requires you to understand how money ‘works’ in the modern economy and to develop a conceptual structure for analysing the economy as it actually exists.

Most people are unaware that a major historical event occurred in 1971, when US President Nixon abandoned gold convertibility and ended the system of fixed exchange rates. Under that system, which had endured for about 80 years (with breaks for war), currencies were convertible into gold, exchange rates were fixed, and governments could expand their spending only by increasing taxes or borrowing from the private sector. After 1971, most governments issued their own currencies by legislative fiat; the currencies were not convertible into anything of value, and were floated and traded freely in foreign currency markets.

It is thus essential to understand the notion of a currency regime, which can range through a continuum from fixed exchange rate systems to floating exchange rate systems with varying degrees of exchange rate management in between. Understanding the way the exchange rate is set is important because it allows us to appreciate the various policy options that the currency issuer – the government – has in relation to influencing the main objects of our study; employment, output and inflation.

A flexible exchange rate releases monetary policy from defending a fixed parity against a foreign currency. Fiscal and monetary policy can then concentrate on ensuring domestic spending is sufficient to maintain high levels of employment. A consequence of this is that governments that issue their own currencies no longer have to ‘fund’ their spending. They never need to ‘finance’ their spending through taxes or selling debt to the private sector. The reality is that currency-issuing governments such as those of Australia, Britain, Japan and the US can never run out of money. These governments always have the capacity to spend in their own currencies.

Most of the analysis appearing in macroeconomics textbooks, which filters into the public debate and underpins the cult of austerity, is derived from ‘gold standard’ logic and does not apply to modern fiat monetary systems. Economic policy ideas that dominate the current debate are artefacts from the old system, which was abandoned in 1971.

At the heart of macroeconomics is the notion that at the aggregate level, total spending equals total income and total output. In turn, total employment is related to the total output in the economy. So to understand employment and output determination we need to understand what drives total spending and how that generates income, output and the demand for labour.

In this context, we will consider the behaviour and interactions of the two economic sectors – that is, government and non-government. Then we will unpack the non-government into its component sectors – the private domestic sector (consumption and investment) and the external sector (trade and capital flows). In Chapter 4 we analyse in detail the so-called National Accounts, drawing on these broad macroeconomic sectors. This approach is called the sectoral balance approach, which builds on the accounting rule that a deficit in one sector must be offset by surpluses in the other in the case of the government – non-government dichotomy. More generally, the sum of the sectoral balances nets to zero when we consider the government, private domestic and external sectors.

If one sector spends more than its income, at least one of the others must spend less than its income because for the economy as a whole, total spending must equal total receipts or income. While there is no reason why any one sector has to run a balanced budget, the National Accounts framework shows that the system as a whole must. Often though, but not always, the private domestic sector runs a surplus – spending less than its income. This is how it accumulates net financial wealth. Overall private domestic sector saving (or surplus) is a leakage from the overall expenditure cycle that must be matched by an injection of spending from another sector. The current account deficit (the so-called external sector account) is another leakage that drains domestic demand. That is, the domestic economy is spending more overseas than foreigners are spending in the domestic economy. These concepts are developed in full in Chapter 5.

Here it is useful to differentiate between a stock and a flow. The latter is a magnitude per period of time. For example, spending is always a flow of currency per period (for example, households might spend \$100 billion dollars in the first three months of 2016). On the other hand, a stock is measured at a point in time. For example, a student’s financial wealth could consist of a deposit

account at a local bank, with a balance of \$1000 on January 1, 2016. We explain stocks and flows in more detail in Chapters 4 and 5.

The sectoral balances framework, outlined later, shows that a sectoral deficit (a flow, say per year) accumulates, as a matter of accounting to a financial debt (a stock). On the other hand, a sequence of sectoral surpluses accumulate to a financial asset which is also a stock. MMT is thus based on what is known as a stock-flow consistent approach to macroeconomics where all flows and resulting stocks are accounted for in an exhaustive fashion. The failure to adhere to a stock-flow consistent approach can lead to erroneous analytical conclusions and poor policy design.

From the perspective of fiscal policy choices, an important aspect of the stock-flow consistent approach that will be explained in greater detail in Chapter 5, is that one sector's spending flow equals its income flow plus changes to its financial balance (stock of assets).

The textbook will show that a country can only run a current account deficit if the rest of the world wishes to accumulate financial claims on the nation (financial debt). Often these claims are in the form of government debt. The MMT framework shows that for most governments, there is no default risk on government debt, and therefore such a situation is 'sustainable' and should not be interpreted to be necessarily undesirable. Any assessment of the fiscal position of a nation must be taken in the light of the usefulness of the government's spending program in achieving its national socio-economic goals. This is what Abba Lerner (1943) called the 'functional finance' approach. Rather than adopting some desired budgetary outcome, government ought to spend and tax with a view to achieving 'functionally' defined outcomes, such as full employment.

On matters of terminology, we avoid using the term 'budget' to describe the spending and taxation outcomes for the currency-issuing government. Instead, we use the term **fiscal balance**. A government fiscal deficit occurs when its spending exceeds its taxation revenue, whereas a fiscal surplus occurs when government spending is less than its taxation revenue.

The use of the term 'budget' to describe the fiscal balance invokes the idea that the currency-issuing government faces the same 'budget' constraints as a household. A careful understanding of the monetary system will make it obvious that the government is not a 'big household'. The government can consistently spend more than its revenue because it creates the currency. Households use the currency issued by the government and must finance their spending. Our access is constrained by the sources of available funds, including income from all sources, asset sales, and borrowings from external parties. Whereas households have to save (spend less than they earn) to spend more in the future, governments can purchase whatever they like, as long as there are goods and services for sale in the currency they issue.

A sovereign government must spend first before it can subsequently tax or borrow. A household cannot spend more than its revenue indefinitely because continuously increasing private debt is unsustainable. The budget choices facing a household are thus limited and prevent permanent deficits. A currency-issuing government can never be revenue constrained in a technical sense and can sustain deficits indefinitely without solvency risk. In other words, our own personal budget experience generates no knowledge relevant to consideration of government matters. The alternative narrative, which we present in this book, highlights the special characteristics of the government's currency monopoly.

Fiscal surpluses provide no greater capacity to governments to meet future needs, nor do fiscal deficits erode that capacity. Governments always have the capacity to spend in their own currencies. The consequences of a fiscal surplus – the government spending less than it is taking out of the economy by way of taxation – when a nation runs an external deficit will also be outlined. In summary, budget surpluses force the non-government sector into deficit and the domestic private sector is forced to accumulate ever-increasing levels of indebtedness to maintain its expenditure. The textbook will explain why this is an unsustainable growth strategy and how eventually the private domestic sector is forced to reduce its risky debt levels by saving more and the resulting drop in non-government spending will reinforce the negative impact of the government fiscal surplus on total spending.

### **The macro model**

To organise the way of thinking in this regard we use a conceptual structure sometimes referred in the economics literature as a model – in this case a macroeconomic model. A model is just an organising framework and is a simplification of the system that is being investigated. In this textbook, we will develop a macroeconomic model, which combines narrative and some algebra to advance your understanding of how the real world economy operates. We will necessarily stylise where complexity hinders clarity, but we will always focus on the real world rather than an assumed world that has no relevance to the actual economy.

All disciplines develop their own language as a way of communicating. One might think that this just makes it harder to understand the ideas and we have sympathy for that view. But we also understand that students of a specific discipline – in this case macroeconomics – should be somewhat conversant with the language of the discipline they are studying.

In the Appendix to this book – *Methods, Tools and Techniques* - we present the essential analytical techniques and terminology that you will find used to specify and solve macroeconomic models throughout this book. These tools and techniques are also deployed in the practical exercises that accompany this text and are to be found on the internet home page for the book. The Appendix should be regularly consulted.

A macroeconomic model draws on concepts and algebraic techniques to advance our understanding of the main economic aggregates (such as output, employment and price level). This textbook design is unique because it specifically develops the MMT macroeconomic model, which will be applicable to the real-world issues including economic policy debates. The application to policy is important because macroeconomics is what might be termed a policy science.

By placing government as the currency issuer at the centre of the monetary system we immediately focus on how it spends and how that spending influences the major macroeconomic aggregates that we seek to explain. The framework will at first, provide a general analysis of government spending that applies to all currency-exchange rate systems before explaining the constraints (policy options) that apply to governments as we move from a flexible exchange rate to a fixed exchange rate system. We will consider how the design of the monetary system impacts on the domestic policy choices open to government and the outcomes of specific policy choices in terms of output, employment and inflation.

## **Fiscal and monetary policy**

The two main policy tools that influence what is termed the demand or spending side of the economy are monetary and fiscal policy.

Fiscal policy is represented by the spending and taxation choices made by the government (the 'treasury'). The net financial accounting outcomes of these decisions are summarised periodically by the government fiscal position. Fiscal policy is one of the major means by which the government seeks to influence overall spending in the economy and achieve its aims.

The textbook will show that a nation will have maximum fiscal space:

- If it operates with a sovereign currency; that is, a currency that is issued by the sovereign government and its value is not pegged to foreign currencies; and
- If it avoids incurring debt in foreign currencies, and avoids guaranteeing the foreign currency debt of domestic entities (firms, households, and state, province, or city debts).

Under these conditions, the national government can always afford to purchase anything that is available for sale in its own currency. This means that if there are unemployed resources, the government can always mobilise them – putting them to productive use – through the use of fiscal policy. Such a government is not revenue-constrained, which means it does not face the financing constraints that a private household or firm faces in framing its expenditure decision.

To put it as simply as possible – this means that if there are unemployed workers who are willing to work, a sovereign government can afford to hire them to perform useful work in the public interest. From a macroeconomic efficiency argument, a primary aim of public policy is to fully utilise available resources.

The central bank in the economy is responsible for the conduct of monetary policy, which typically involves the setting of a short-term policy target interest rate (Fed Funds in the USA, also called bank rate in many countries). In the recent global economic crisis the ambit of monetary policy has broadened considerably and these developments will be considered in Chapter 15.

The typical roles of a central bank include not only the conduct of monetary policy via the overnight interbank lending rate, but also operating the interbank clearing mechanism (so that bank cheques clear among banks), acting as lender of last resort (to stop bank runs), and regulating and supervising the banks.

MMT considers the treasury and central bank functions to be part of what is termed the consolidated government sector. In many textbooks, students are told that the central bank is independent from government. The MMT macroeconomic model will demonstrate how it is impossible for the two parts of government to work independently if the monetary system is to operate smoothly.

## **Policy implications for sovereign nations**

MMT provides a broad theoretical macroeconomic framework based on the recognition that sovereign currency systems are in fact public monopolies *per se*, and that the imposition of taxes coupled with insufficient government spending generates unemployment.

An understanding of this point will be developed to allow the student to appreciate the role that government can play in maintaining its near universal dual mandates of price stability and full employment. The student will learn that there are two broad approaches to control inflation available to government in designing its fiscal policy choices.

Both approaches draw on the concept of a **buffer stock** to control prices. We will examine the differences between the use of:

- a) *Unemployment buffer stocks*: The neoclassical approach, which describes the current policy orthodoxy, seeks to control inflation through the use of high interest rates (tight monetary policy) and restrictive fiscal policy (austerity), which leads to a buffer stock of unemployment. In Chapters 11 and 12, students will learn that this approach is very costly and provides an unreliable target for policy makers to pursue as a means for inflation proofing; and
- b) *Employment buffer stocks*: Under this approach the government exploits its fiscal capacity, inherent in its currency issuing status, to create an employment buffer stock. In MMT, this is called the **Job Guarantee (JG)** approach to full employment and price stability. This model, which is considered by MMT to be the superior buffer stock option, is explained in detail in Chapter 12.

The MMT macroeconomic framework shows that a superior use of the labour slack necessary to achieve price stability is to implement an employment program for those who are otherwise unemployed as an activity floor in the real output sector, which both anchors the general price level to the price of employed labour of this (currently unemployed) buffer and can produce useful output with positive supply side effects.

### 1.3 Macro and the Public Purpose

The households and business firms in a modern capitalist economy make many of the important economic decisions that contribute to determination of the level of employment and output, the composition of that output, the distribution of income, and the prices at which output is sold. Claims are sometimes made that a ‘free market’ economy comprised of individuals seeking only their own self interest can operate ‘harmoniously’ as if guided by an ‘invisible hand’ (see Section 1.1). In fact, economists had rigorously demonstrated by the 1950s that the conditions under which such a stylised economy could reach such a result couldn’t exist in the real world. In other words, there is no scientific basis for the claim that ‘free markets’ are best.

In any case, these claims, even if true for some hypothesised economy, are irrelevant for the modern capitalist economies that actually exist. This is because all modern capitalist economies are ‘mixed’, with huge corporations (including multinational firms), labour organisations, and big government. Individuals and firms operate within socio-political-cultural-economic structures that are constraining but also enabling.

Sometimes the goals of individuals and firms coincide with what might be called the public purpose, while often they do not. In this section we will discuss the public purpose and the role played by government in trying to align private interests with socially progressive goals.

What is the public purpose? It is not easy to define or to identify the public purpose. One of the basic functions of any social organisation is to provide the necessary food, clothing, shelter, education, health care, legal framework, and socialisation for survival of the society.

While the subject of this course is economics, there is no sharp distinction between the sphere of economics and the spheres of other social sciences that study social processes. We usually think of the economy as the part of the social organisation that is responsible for provision of the material means of survival—the food, clothing, shelter, and so on. However, the economy is always embedded in the social organisation as a whole, affecting and affected by culture, politics, and social institutions.

Even if we can agree that any successful economic organisation should be able to produce adequate food for its population, that still leaves open many questions: What kind of food?; How should it be produced?; How should it be distributed?; and even What does adequate mean?

Further, no society is comprised of harmonious individuals and groups. There are always conflicting claims and goals that must be moderated. There is no single, obvious public purpose to which all members of a society wish to strive. Even if we can identify a set of goals that the majority of society would like to work toward, that set will surely change over time as hopes and dreams evolve. The public purpose is an evolving concept.

The position taken in this book is that there is no ‘invisible hand’ that ensures that private interests are consistent with the public purpose. Indeed, the economy is just one component of the social organisation that is necessary to establish the always-evolving public purpose and that is necessary to work towards achievement of the public purpose.

The ‘market’ is just one institution among a wide variety of social institutions working to delineate social goals that comprise the social and private purposes. Other institutions include political organisations, labour unions, manufacturers, and NGOs (non-governmental organisations).

As we noted at the beginning of this Chapter, the national government must play an important role in society as it can help to identify the social purpose and to establish a social structure in which individuals and groups will work toward achieving the social purpose.

While it is admittedly difficult to outline what defines the social purpose, it is possible to identify widely accepted goals. For example, the United Nations Universal Declaration of Human Rights (1948) commits signatory nations to a common set of relatively well-defined goals.

The declaration is outlined on the United Nations Home Page:<sup>3</sup>

Now, Therefore THE GENERAL ASSEMBLY proclaims THIS UNIVERSAL DECLARATION OF HUMAN RIGHTS as a common standard of achievement for all peoples and all nations, to the end that every individual and every organ of society, keeping this Declaration constantly in mind, shall strive by teaching and education to promote respect for these rights and freedoms and by progressive measures, national and international, to secure their universal and effective recognition and observance, both among the peoples of Member States themselves and among the peoples of territories under their jurisdiction.

The Articles that define the Declaration include:

- Everyone has the right to life, liberty and security of person.
- No one shall be held in slavery or servitude; slavery and the slave trade shall be prohibited in all their forms.

---

<sup>3</sup> <http://www.un.org/en/universal-declaration-human-rights/> accessed January 15, 2016.

- Everyone has the right to an effective remedy by the competent national tribunals for acts violating the fundamental rights granted them by the constitution or by law.
- Everyone has the right to freedom of movement and residence within the borders of each state.
- Everyone has the right to a nationality.
- Men and women of full age, without any limitation due to race, nationality or religion, have the right to marry and to found a family. They are entitled to equal rights as to marriage, during marriage and at its dissolution.
- Everyone has the right to own property alone as well as in association with others.
- Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change their religion or belief, and freedom, either alone or in community with others and in public or private, to manifest their religion or belief in teaching, practice, worship and observance.
- Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.
- Everyone has the right to freedom of peaceful assembly and association.
- Everyone has the right to take part in the government of their country, directly or through freely chosen representatives.
- Everyone has the right of equal access to public service in their country.
- Everyone, as a member of society, has the right to social security and is entitled to realisation, through national effort and international co-operation and in accordance with the organisation and resources of each State, of the economic, social and cultural rights indispensable for their dignity and the free development of their personality.
- Everyone has the right to work, to free choice of employment, to just and favourable conditions of work and to protection against unemployment.
- Everyone, without any discrimination, has the right to equal pay for equal work.
- Everyone who works has the right to just and favourable remuneration ensuring for themselves and their family an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection.
- Everyone has the right to form and to join trade unions for the protection of their interests.
- Everyone has the right to rest and leisure, including reasonable limitation of working hours and periodic holidays with pay.
- Everyone has the right to a standard of living adequate for the health and well-being of themselves and of their family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond their control.



- Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages. Elementary education shall be compulsory. Technical and professional education shall be made generally available and higher education shall be equally accessible to all on the basis of merit.
- Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits.

It is obvious that many of these identified human rights, especially near to the end of this list, are connected to the operation of the economy. For example, we argued above that any successful economy should provide adequate food, clothing, and shelter, and many of the human rights listed in the UN Charter address the material well-being of a nation's population.

Further, other human rights that superficially appear to be unrelated to economic performance actually presuppose fulfilment of other human rights that are directly related to material well-being.

For example, in a modern capitalist economy access to employment (one of the recognised rights) is necessary for full participation in society. Not only does a job provide income that allows one to purchase food, clothing, and shelter, but it also provides access to social networks, generates feelings of self-worth as one contributes to social production, enhances social prestige, and helps to provide for retirement in old age.

Indeed, employment has been shown to have a wide range of other benefits to individuals and to society including better physical and psychological health, reduced crime and drug abuse, lower child and spouse abuse, and greater participation in other social and political activities.

To be sure, this list (which is itself only a partial listing of the agreed universal rights) includes many rights that have not been fully achieved even in the wealthiest and most democratic nations. In that sense, these rights are 'aspirational', with the signatory nations committing to striving toward achieving them. Again, if we look at the example of the right to work and to an adequate standard of living, those are rights that are routinely violated even in the best of times in the wealthiest of nations. Still, these universally recognised rights provide a measure against which nations can measure their progress.

### **Concluding thoughts on the public purpose**

We conclude with three important points.

First, this reason the public purpose is broad and evolving over time, and for these reasons it varies across time and place. It should include rising living standards, particularly for those at the bottom of society. Environmental sustainability must be included. Reduction of racial, ethnic, and gender inequalities across the full socio-political-economic spectrum is an important component of the public purpose. This must go beyond simple economic measures such as family income to include full participation in the life of the community. The public purpose also should include reductions of crime, corruption, cronyism, invidious distinction, conspicuous consumption, and other social pathologies.

Second the UN Charter lays out what it sees as 'universal' human rights. This is a useful, but not wholly satisfactory list to be included in a statement of the public purpose. What is considered to be a human right today might have appeared to be radically utopian a century ago; and today's list will appear far too cautiously conservative at some date in the future.

The public purpose is inherently a progressive agenda that strives to continually improve the material, social, physical, cultural, and psychological well-being of all members of society. It is inherently 'aspirational' in the sense that there is no end point as the frontiers of the public purpose will continually expand.

Third, the national government as well as international organisations (such as the United Nations) must play important roles in shaping our vision regarding the types of societies to which we aspire. And beyond setting these goals, governments at all levels must take the lead in developing sets of institutions, rules of behaviour, and sanctions for undesirable behaviour in order to move toward reaching the goals set as the public purpose.

As an example that demonstrates these points, a half century ago national governments and international organisations set about to eliminate the devastating disease known as smallpox. While markets and for-profit production played a role in helping to develop vaccines, in distributing the vaccines, and in formulating information campaigns, private initiative alone would never have eliminated the disease.

The task was too big, it was not completely consistent with the self-interest of profit seeking behaviour, and it required international cooperation beyond the reach of even the largest firms.

Hence, governmental organisations had to play a role.

With respect to the aspirational nature of the public purpose, successful elimination of smallpox would not be the end, but rather would serve as the beginning of a new campaign, to eliminate another disease, and then another and yet another.

Perhaps in a long-distant future, a human right to a disease-free life would be recognised, adding to an ever-increasing list of established rights that all nations would be expected to protect.

While we cannot, of course, imagine such a future, it was not so long ago that the Congress of the US did not recognise the voting rights of women and African Americans. Today, any nation that denies the vote to members of society on the basis of gender, religion, race or ethnicity, or national origin is considered to be in violation of human rights, and thus, to be an international pariah—even though such restrictions were considered acceptable just a few generations ago. For example, white US women over the age of 21 did not secure the vote until the 1920 Presidential election, whereas in the UK suffrage was extended to all women over the age of 21 in 1928. In Australia aborigines were granted the right to enrol and vote in Federal elections in 1962. Many developed countries did not give women or minorities the vote until the 20th Century.

The public purpose is inherently progressive; it can never be finished.

## Appendix: Advanced Material

### 1. The minimum wage debate

Economic theory gives (at least) two answers to the question of the effect of raising minimum wages on unemployment.

1. Raising wages increases business costs that beyond some point will increase the price of output. If we hold the income and purchasing power of consumers constant, it would seem that the higher prices must lead to fewer sales, and hence to lower employment as employers downsize. (There are other effects that could strengthen this impact, such as higher imports from abroad where labour is cheaper, and also substitution of machines for labour whose price has gone up.) Thus, neoliberals argue that raising the minimum wage must lead to higher unemployment.
2. Not so fast, says our two-armed economist. If wages rise, then it is not necessarily true that consumer income and hence purchasing power is constant. After all, most consumption is financed by wages, and the incomes of those employed at the lowest wages have increased. Those workers buy more goods and services. Firms selling them might decide to hire more workers. Those workers buy more, too. Even if some employers decide that at the higher minimum wage they prefer to buy robotic machines to replace workers that means more jobs making machines. We cannot say for sure that the net result of this complex chain reaction will be more jobs or fewer jobs.

The frustrated student asks ‘But why can’t we just look at real world evidence to settle the question?’ Economists do, of course, try to do just that—and the tool of choice is econometrics. We can look at a number of cases where minimum wages have been raised (in the USA, for example, the 50 US states have their own minimum wage laws so it is possible to compare employment effects in one state when the minimum wage is raised while it is held constant in a neighbouring state with otherwise similar conditions). What the most careful studies in the USA find is that raising wages does not tend to reduce employment and raise unemployment—indeed, it looks like the correlation goes the other way, with employment rising.

Does that settle the case? No. Even leaving aside clearly ideologically biased claims by opponents of minimum wage hikes, such empirical studies cannot be decisive. Even the most carefully controlled tests cannot control for all possible factors that might affect employment. We cannot be sure raising wages caused employment to rise. There could well be an uncontrolled factor that coincidentally increased employment even though the wage hike by itself would have reduced the number employed.

Economists are well aware of this conundrum: empirical correlation never proves causation. Causation, itself, is a deeply complex topic. While we can put together theory and models and data to make a case, we probably will not be able to prove that ‘X causes Y’ when it comes to the most significant questions in economics.

John Maynard Keynes argued that the best one can do is to convince by the weight of one’s argument. Certainly one needs theory and probably evidence and maybe even a math model, but even that will not convince an opponent unless the case is made through a persuasive argument. Keynes was a master of argument—but even he did not always win. More recently Deidre McCloskey has made much the same argument in her book *The Rhetoric of Economics* (1985).

Again, her point is that evidence alone is not decisive; ‘rhetoric’, or the art of discourse, is also important.

## 2. The structure of scientific revolutions

In his influential book, *The Structure of Scientific Revolutions*, Thomas Kuhn (1970) advanced a thesis in which he distinguishes between ‘normal science’ that works within a ‘paradigm’ and a ‘scientific revolution’ that breaks free of the paradigm. For our purposes, we can think of the neoclassical approach as a ‘paradigm’ that works within the framework of utility maximisation and rationality, with Keynesian (/Institutionalist/Marxist) as the paradigm-breaking scientific revolution.

Returning to the debate about the minimum wage, the neoliberal conclusion that raising wages causes unemployment is the correct answer if one views the question from within the neoclassical paradigm. In that paradigm, prices ration resources, and at higher prices there will be less demand. As wages rise, employers want fewer workers. It makes sense to argue that unemployment rises.

However, within the heterodox paradigm, what matters is aggregate effective demand (a topic we will turn to later). Higher wages mean more income and more sales, hence firms want more workers. The net effect of a wage hike could be more employment.

Kuhn’s breakthrough was the realisation that most of the time scientists (including economists) work within a paradigm, asking and attempting to answer questions in a manner that is consistent with the paradigm. He calls this ‘normal science’ and the research process as largely one of ‘puzzle solving’.

The ‘normal scientist’ comes across ‘anomalies’ that are hard to resolve within the constraints of the paradigm within which they work. An example commonly used is the convention of tipping in a restaurant. If we assume that the diner and the server are both completely rational in the neoclassical sense (that is, selfish), then the tip typically should be negotiated before the meal to induce good service—except in the case where the diner is a local who frequents the restaurant. The local diner can wait until after the meal to tip for good service. The server will provide good service in advance of the tip expecting the diner will reward good service. If the diner disappoints, they can expect not only bad service but perhaps even worse on the next visit (servers have been known to spit into food, after all).

The tourist or business visitor, however, might never expect to return to the restaurant. A tip before service could be negotiated depending on the level of service the diner wants. A contract is made and then if the server provides the service contracted the payment is made at dinner’s end. The contract might include an external opinion and enforcement mechanism. In practice we do not observe such contracts. Rather, the diner pays a tip at the end of dinner, based on assessment of services rendered. However, a rational one-time visitor would never pay a tip after service. Why bother? It is too late for the server to spit in the food. And the diner never expects to return. Such behaviour is an anomaly for the neoclassical paradigm.

Kuhn’s argument was that over time, as researchers pursue normal science working within their paradigm they come up against more and more anomalies that cannot be explained. Another example would be the flat earth theory. Early scientists could come up with increasingly complicated explanations for apparent anomalies. For example, as ships approach shore from a

distant horizon, only the tops of the masts are first visible—due to the earth’s curvature. However, if light travels in a curved path that phenomenon could be explained within the flat earth paradigm. Yet, other tests would find that light apparently travels in a straight line. An anomaly.

According to Kuhn as the anomalies build, some researchers begin to think outside the box of the paradigm. Well, maybe the earth is not flat. Maybe guests and servers are not ‘rational’ in the narrow neoclassical sense. They begin to develop a new paradigm. Kuhn calls this a ‘scientific revolution’ and it has been likened to taking off distorting glasses and putting on prescription lenses that correct vision. The world never looks the same again as the new paradigm changes one’s view completely. What were thought to be anomalies are easily explained within the new paradigm. It isn’t a coincidence that the new paradigm is developed by younger researchers or by those outside officialdom of the profession as it is easier for them to cast off the old ideas.

Within the new paradigm, normal science advances by puzzle solving, and eventually comes up against new anomalies. Eventually yet another scientific revolution will be needed. Note that no disparagement of ‘normal’ science is intended. Most of the advance of science comes through puzzle solving. Indeed, one cannot do research or even attempt to understand the world without a paradigm. But puzzle solving, by itself, is not enough. Scientific revolutions are needed because paradigms are also constraining—they limit the conception of what is possible.

John Maynard Keynes wrote a friend, George Bernard Shaw, when he had finished drafts of the General Theory proclaiming that his new book would revolutionise economic theory—if not at once, then at least eventually. That is quite a claim to make, of course, but Keynes was confident and brilliant. The immediate reaction to his book seemed to validate his expectation. While not everyone was about to jump aboard Keynesian theory, it is not an exaggeration to say that many recognised the revolutionary nature of his theory. By the 1960s, most macroeconomists considered themselves to be Keynesian.

And yet Keynesian theory soon fell out of favour. Mainstream macroeconomics began to shed Keynesian ideas from the early 1970s, and they were almost completely gone by the 1990s. It would be as if we returned to flat earth theory after once embracing round earth theory.

Note that part of the difference is that economics is a social science that studies human behaviour and proposes policy that directly affects human lives. It concerns topics that are contentious and where policy benefits some and can hurt the interests of others. All of the policies that came out of the Keynesian revolution were opposed by some groups—whether it is social welfare for the poor, social security for the aged, or jobs for the unemployed. Opponents inevitably regroup and attempt a counterrevolution.

In all these respects, the other social sciences also experience reversals. Social theories from the past are thrust into the limelight again. Indeed, even in the ‘hard’ sciences, old ideas sometimes come back. In the USA, for example, the well-established theory of evolution is again under attack. Kuhn had warned that we should not see science as steadily progressing in a linear fashion from myth to truth. There is a tendency to write the textbooks in that manner, but reality is messy.

In any event, the authors of this textbook do view Keynes’s General Theory as a scientific revolution in Kuhn’s sense, as was Karl Marx’s theories presented in his 1867 book Capital. In both cases, orthodoxy mounted counter revolutions to restore neoclassical thought. Already by

the 1870s three orthodox economists had published books to not only defend but to strengthen the arguments of neoclassical economics against Marx's economics: Jevons, Walras and Menger published their great contributions between 1871 to 1873 in response to Marx (Henry, 2012).

In the case of the General Theory, the Keynesian revolution was gradually aborted as a few of Keynes's ideas were wedded to the neoclassical approach, forming the 'Neoclassical Synthesis' of textbooks (with Paul Samuelson (1947) at the lead). All the revolutionary insights of Keynes (and Veblen and Marx) were dropped in order to make Keynes more-or-less consistent with neoclassical economics. Unlike the case of Marx's Capital, which was openly disparaged, Keynes's book was celebrated. Most macroeconomists became 'Keynesian' even though few understood the book and few of Keynes's ideas were actually adopted in the 'Synthesis'.

Heterodox economists insist that this was a mistake. Instead, neoclassical theory should have been dropped, and the revolutionary insights of heterodoxy (stretching all the way back to Marx) should have led to a new paradigm.

Let us turn to a heterodox framework for macroeconomics. While we will throughout this book present the neoclassical approach, our main purpose is to develop the coherent heterodox alternative.

## References

- Henry, J.F. (2012) *The Making of Neoclassical Economics*, Routledge Revivals, Taylor and Francis.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, Macmillan, London.
- Kuhn, T.S. (1970) *The Structure of Scientific Revolutions*, 2nd edition, University of Chicago Press, Chicago.
- Lerner, A. (1943) 'Functional Finance and the Federal Debt', *Social Research*, 10(1), 38-51.
- Marx, K. (1867) *Capital*, Volume I, Everyman's library, London.
- McCloskey, D. (1985) *The Rhetoric of Economics*, University of Wisconsin Press, Madison.
- Samuelson, P. (1947) *Foundations of Economic Analysis*, Harvard University Press, USA.
- United Nations (1948) *United Nations Declaration of Human Rights*, United Nations General Assembly, 10 December 1948, available at: <http://www.un.org/en/universal-declaration-human-rights/> accessed January 15, 2016.

# Chapter 2: How to Think and Do Macroeconomics

## Chapter Outline

- 2.1 Introduction
  - 2.2 Thinking in a Macroeconomic Way
  - 2.3 What Should a Macroeconomic Theory be Able to Explain?
    - Real GDP growth
    - Unemployment
    - Real wages and productivity
    - Private sector indebtedness
    - Central bank balance sheets
    - Japan case study
    - Summary
- Appendix
- The Buckeroos model
  - Implications of the Buckeroos model

## Learning Objectives

1. Recognise the importance of the fallacy of composition in understanding macroeconomics.
2. Gain an awareness that macroeconomics is a highly contested discipline in terms of theory and policy prescription.
3. Note the importance of referring to the stylised facts in analysing theory and policy prescription.
4. Develop critical thinking skills about the working of a macroeconomy with its own sovereign currency

## 2.1 Introduction

In Chapter 1, we noted that any science, whether physical or social, develops theories to gain an understanding of the specific phenomena that it is trying to explain. This necessitates abstraction. In economics there are two broad schools of thought, which means that economics is a contested discipline, with ongoing debates about both theory and policy. We outlined the subject matter of macroeconomics and highlighted the distinct features of Modern Monetary Theory (MMT). Finally we provided a discussion of macroeconomic policy objectives, by introducing the concept of public purpose.

All disciplines have their own language and way of thinking. In the next Section of this Chapter, we argue that thinking as a macroeconomist is particularly challenging, because the discipline is highly contested with self-styled experts offering diverse views. An important contemporary example is the MMT rejection of the neoclassical claim that a currency-issuing national government is like a household and subject to same type of ‘budget’ constraint. More generally some propositions, which make sense at an intuitive, personal level, fail to hold an aggregate level. This is referred to as the **fallacy of composition**. A number of examples, both economic and non-economic, are provided. We then discuss what macroeconomics should be able to explain and outline two empirical examples relating to unemployment and the conduct of fiscal policy in which there are sharp theoretical differences between MMT and orthodoxy.

Finally in the Appendix, we provide a brief outline the **Buckaroos model**, which has been implemented at the University of Missouri at Kansas City (UMKC) in the United States. UMKC students are required to undertake a certain number of hours of voluntary labour for community service providers prior to graduation. The Buckaroos model is a means of operationalising the administration of this scheme and provides insights about the operation of a modern monetary economy.

## 2.2 Thinking in a Macroeconomic Way

Macroeconomics is a controversy-ridden area of study. In part, this is because the topic of study is seen as being of great significance to our nation and our daily lives even though the details that are discussed are mostly difficult for us to understand.

The popular press and media in general are flooded with macroeconomics – the nightly news bulletins invariably have commentators speaking about macroeconomic issues – such as the real GDP growth rate, the inflation rate or the unemployment rate. The population has been more exposed to macroeconomic terminology over the last two or so decades and the advent of social media has given voice to anyone who wants to be a macroeconomic commentator.

The so-called blogosphere is replete with self-styled macroeconomic experts who wax lyrical about all and sundry, often relying on intuitively logical arguments to make their cases. The problem is that common sense is a dangerous guide to reality and not all opinion should be given equal privilege in public discourse. Our propensity to generalise from personal experience, as if the experience constitutes general knowledge, dominates the public debate – and the area of macroeconomics is a major arena for this sort of problematic reasoning.

A typical statement that is made in the public arena is that the government might run out of money if it doesn’t curb spending. Conservative politicians who seek to limit the spending ambit



of government often attempt to give this statement authority by appealing to our intuition and experience.

They draw an analogy between the household and the sovereign government to assert that the microeconomic constraints that are imposed on individual or household choices apply equally without qualification to the government.

So we are told that governments, like households, have to live within their means. This analogy resonates strongly with voters because it relates the more amorphous finances of a government to our daily household finances.

As we noted in Chapter 1, we know that we cannot run up our household debt forever and that we have to moderate our spending when we reach the borrowing limits on our credit cards. We can borrow to enhance current spending, but eventually we have to sacrifice spending to pay the debts back. We cannot indefinitely live beyond our financial means.

Neoliberals draw an analogy between the two – household and government – because they know we will judge government deficits as being reckless, more so if fiscal deficits rise. But the government is not a big household. It can consistently spend more than its revenue if it creates the currency.

Whereas households have to save (spend less than they earn) or borrow to spend more in the future, governments can purchase whatever they like whenever there are goods and services for sale in the currency they issue. Governments always have the capacity to spend in their own currencies. Governments like Britain, the United States, Japan and Australia can never run out of money. We make brief reference to the Japanese economy later in this Chapter.

In addition, fiscal surpluses (taxation revenue greater than government spending) today do not provide greater capacity to governments to meet future spending needs, nor do fiscal deficits (taxation revenue less than government spending) erode that capacity.

MMT teaches that our experience in managing our own household budgets provides no guidance about the management of the government fiscal position, yet on a daily basis, we are told it does. We are users of the currency that the government issues.

The government has to consider the real resources that are available to the economy and how best to deploy them. These are not financial considerations – there are no intrinsic ‘financial’ constraints that are relevant to a currency-issuing government.

A household always has to consider its financial means. Common sense tells us that if we have ‘too much debt’ then we can save and reduce that debt. But, putting aside whether public debt is problematic (see Chapter 14), if the government tries to ‘save’ (another inapplicable conceptual transfer from the individual level) then public debt will probably rise.

Indeed, in the 1930s macroeconomics started life as a separate discipline of study from microeconomics because the dominant way of thinking at the time was riddled with errors of logic that led to spurious analytical reasoning and poor policy advice.

Microeconomics develops theories about individual behavioural units in the economy – the person, household, or firm. For example, it might seek to explain the employment decisions of a firm or the saving decisions of an individual income recipient. However, microeconomic theory ignores knock-on effects on others when examining these firm level or household level

decisions. That is clearly inappropriate if we look at the macroeconomy, where we must consider the impacts on others.

We have learned that macroeconomics studies the aggregate outcomes of the behaviour of all firms and households. The question is how do we go from the individual unit (microeconomic) level to the economy-wide (macroeconomic) level? This is a question that the so-called aggregation problem seeks to address.

Prior to the 1930s, there was no separate study called macroeconomics. The dominant neoclassical school of thought in economics at the time considered macroeconomics to be a simple aggregation of the reasoning conducted at the individual unit or atomistic level.

To make statements about industry or markets or the economy as a whole, they sought to aggregate their atomistic analysis. For reasons that will become clearer, simple aggregation proved to be flawed.

The solution was to fudge the task and introduce the notion of a 'representative household' to be the demand side of a goods and services (product) market and the 'representative firm' to be the supply side of that market. Together they bought and sold a 'composite good'. These aggregates were fictions and assumed away many of the interesting aspects of market interaction.

For example, if we simply sum all the individual demand relationships between price and spending intention we could form a representative household demand function.

But what if the spending intentions of each household or a segment of them were interdependent rather than independent? What if one household changed their demand once they found out what the spending intentions of the next-door neighbour were (for example, the notion of keeping up with the Joneses!)? What if the actions of one household impinge on the feasible choices of another? Then a simple summation of demands is inappropriate.

But these issues were abstracted from and the representative firm and household were just bigger versions of the atomistic unit and the underlying principles that sought to explain the behaviour of the representative firm or household were simply those that were used to explain behaviour at the individual level which ignored any impacts on others. Accordingly, changes in behaviour or circumstances that might benefit the individual or the firm are automatically claimed to be of benefit to the economy as a whole.

In the Great Depression, this erroneous logic guided policy in the early 1930s and the crisis deepened. At that time, John Maynard Keynes and others sought to expose the logical error that the dominant orthodoxy had made in their approach to aggregation. In that debate this mode of thinking was considered to incorporate a compositional fallacy. This led to the development of macroeconomics as a separate discipline from microeconomics. Karl Marx had appreciated this fallacy in the mid-1800s but his contributions were largely ignored in the popular economic theory of the early 20th century.

**Compositional fallacies are errors in logic that arise when we infer that something, which is true at the individual level, is also true at the aggregate level.** The fallacy of composition arises when actions that are logical, correct and/or rational at the individual or micro level have no logic (and may be wrong and/or irrational) at the aggregate or macro level.

Keynes led the attack on the mainstream thinking at the time – mid-1930s – by exposing several fallacies of composition, including the paradox of thrift and the wage cutting solution to unemployment.

A contemporary example of the flawed reasoning that follows a fallacy of composition is the proposal to engage in fiscal austerity in response to higher fiscal deficits. Prior to considering the paradox of thrift and fiscal austerity, let us first consider two simpler examples, the first of which is non-economic and the second has economic relevance.

Consider a large crowd attending a sporting event. The stadium provides seating for all attendees. A spectator would get a better view of an incident occurring near the sideline by standing up. Would all members of the crowd get a better view by standing up? Clearly the answer to this question is no.

Consider an employee who loses their job on Thursday evening. On Friday morning they consult the vacancies advertised in the local newspaper and online and apply for suitable jobs. They also knock on the doors of local employers and present a C.V. and request a job. Within a week they have secured a new job, following their thorough job search. Would it be correct to argue that if all the unemployed searched as conscientiously for jobs, then the unemployment problem would be solved? The answer is no. To make the discussion simple, assume all the unemployed are qualified to fill the available job vacancies, but 100 workers are competing for 50 jobs. At best, 50 of these job seekers will remain unemployed, irrespective of how thoroughly they search for jobs. This topic is further discussed in The Parable *One hundred dogs and 94 bones* (CofFEE).

The paradox of thrift tells us that what applies at a micro level that is the ability to increase saving if one is disciplined enough, does not apply at the macro level. Thus if everyone attempts to increase saving, overall incomes would fall and total saving in the economy would not increase.

There is an old saying – look after the pennies and the pounds will look after themselves.

So by reducing their individual consumption spending a person can increase the proportion they save and enjoy higher future consumption possibilities as a consequence. The loss of spending to the overall economy of this individual's adjustment would be small and so there would be no detrimental impacts on overall economic activity, which is crucially driven by aggregate spending.

But imagine if all individuals (consumers) sought the same goal and started to withdraw their spending *en masse*? Then total spending would fall significantly and, as you will learn from Chapter 7 national income falls (as production levels react to the lower spending) and unemployment rises. The impact of lost consumption on aggregate demand (spending) would be such that the economy would plunge into a recession and everyone would suffer.

Moreover, as a result of the lower national income, it is possible that total saving would actually fall along with consumption spending so the economy as a whole would be saving less. As we will see later, if poor sales due to an increased desire to save, negatively impact investment, aggregate saving would certainly fall.

The paradox of thrift tells us that what applies at a micro level (that is, the ability to increase saving if one is disciplined enough) does not apply at the macro level (if everyone attempts to increase saving, overall incomes fall and individuals would be thwarted in their attempts to increase their savings in total).

Why does the paradox of thrift arise? In other words, what is the source of this compositional fallacy?

The explanation lies in the fact that a basic rule of macroeconomics, which you will learn once you start thinking in a macroeconomic way, is that spending creates income and output. This economic activity, in turn, explains how employment is generated. Adjustments in spending drive adjustments in total production (output) in the economy as firms react to higher (lower) sales by increasing (reducing) employment and output.

So if all individuals reduce their spending (by attempting to save) the level of income falls rather than stays constant. By contrast if just one person reduced their spending it is safe to assume that their income would not be affected and that the impacts on all others would be so small they could be ignored.

But we know that if all consumers act *en masse* then not only does their spending fall, but national income also falls and the logic that applied at the individual level will be spurious or fallacious at the aggregate level.

As total saving (the sum of all household saving) is a residual after all households have made their consumption spending choices from the available disposable income then national income changes, in turn, feedback on total saving. When national income falls, consumption falls and total saving may decline in absolute terms.

Certainly total saving will be less than individuals planned due to the fall in equilibrium national income.

By assuming that we could simply add up the microeconomic relations to get the representative firm or household, the mainstream at the time were assuming that the aggregate unit faced the same constraints as the individual sub-units. So the individual saver might reasonably assume that changing their consumption choices would not impact their income.

During the Global Financial Crisis (GFC), the conservative reaction to the increasing government deficits has been to enact fiscal austerity measures, by cutting government expenditure and/or increasing taxes, and to encourage nations to cut domestic costs in order to stimulate their export sectors via increased competitiveness.

In isolation, that is, where one nation does this while all other nations are maintaining strong economic growth, this strategy might have a chance of working. But if all nations engage in austerity and cut their growth rates, then overall spending declines, and imports will fall across the board, as will exports. This is another example of a Fallacy of Composition.

It is the interdependence between all countries via trade, as well as a fall in net government spending that undermines the policy prescription in this case. Further, it is obvious that not all countries can rely on export-led growth (to more than offset a decline in net Government spending) since for every exporter there must be an importer.

MMT contains a coherent logic that will teach you to resist falling into intuitive traps and compositional fallacies. MMT teaches you to think in a macroeconomic way.

Keynes and others considered that fallacies of composition, such as the paradox of thrift, provided a *prima facie* case for considering the study of macroeconomics as a separate discipline. These examples show that we must be very careful when drawing general conclusions on the basis of our own experience (that is, specific-to-general reasoning).

## 2.3 What Should a Macroeconomic Theory be Able to Explain?

Any macroeconomic theory should help us understand the real world and provide explanations of historical events and reasonable forward-looking forecasts as to what might happen as a consequence of known events – for example, changes in policy settings. A theory doesn't stand or fall on its absolute predictive accuracy because it is recognised that forecasting errors are a typical outcome of trying to make predictions about the unknown future.

However, systematic forecast errors (that is, continually failing to predict the direction of the economy) and catastrophic oversights (for example, the failure to predict the 2008 Global Financial Crisis) are an indication that a macroeconomic theory is seriously deficient.

In this section we present some stylised facts about the way in which modern industrialised economies have performed over the last several decades. These facts will be referred to throughout the textbook as a reality check when we compare different approaches to the important macroeconomic issues such as unemployment, inflation, interest rates and government deficits.

The facts provide a benchmark against which any macroeconomic theory can be assessed. If a macroeconomic theory generates predictions which are consistently at odds with what we observe then we conclude that it doesn't advance our understanding of the real world and should be discarded.

### Real GDP growth

Real Gross Domestic Product is the measure of actual production of goods and services in the economy over the course of a particular period. We will learn how the national statistics offices measure it and how we interpret movements in real GDP in Chapter 4 when we study the National Income and Product Accounts (NIPA). For now, we consider economic growth to be measured by the percentage change in real GDP, and in that sense, it is one measure of the prosperity of a nation. We will learn that employment growth is also dependent on output growth and so a higher real GDP growth usually means higher employment and lower unemployment.

Table 2.1 shows the average annual real GDP growth rates by decade from 1960 for various countries. The sample of nations chosen include the three large industrialised European nations representative of the 'north' and 'south' (Germany, Italy and Spain) all of which are members of the Eurozone; Britain, a European nation outside the Eurozone; a small open economy predominantly exporting primary commodities and with a relatively underdeveloped industrial base (Australia), and two large, non-European industrialised nations (Japan and the USA).

**Table 2.1 Average annual real GDP growth by decades, per cent**

	Australia	Germany	Italy	Japan	Spain	UK	US
1960-70	5.0	4.5	5.7	10.2	8.6	3.1	4.7
1970-80	3.3	3.3	4.0	5.2	5.3	2.6	3.2
1980-90	3.4	2.0	2.6	4.4	3.0	2.7	3.1
1990-00	3.2	2.2	1.5	1.5	2.8	2.1	3.2
2000-10	3.2	0.8	0.5	0.6	2.7	1.9	1.8
2010-15	2.6	2.0	-0.5	1.4	-0.3	2.0	2.1

Source: National statistical agencies.

Several things are clear. First, real economic growth has been lower on average in the current period than in the 1960s for each country. Second, the European nations (Italy and Spain) have clearly performed poorly in the recent period. Third, the European nations within the Eurozone, including Germany, have performed relatively poorly since 2000. Fourth, Australia has generally performed better than the other nations in the table.

Among the questions that our macroeconomic approach needs to be able to answer in a consistent fashion are: Why has real GDP growth on average slowed? What explains Australia's superior growth rate between 2010 and 2015? Why have Italy and Spain endured negative growth in the period 2010 to 2015?

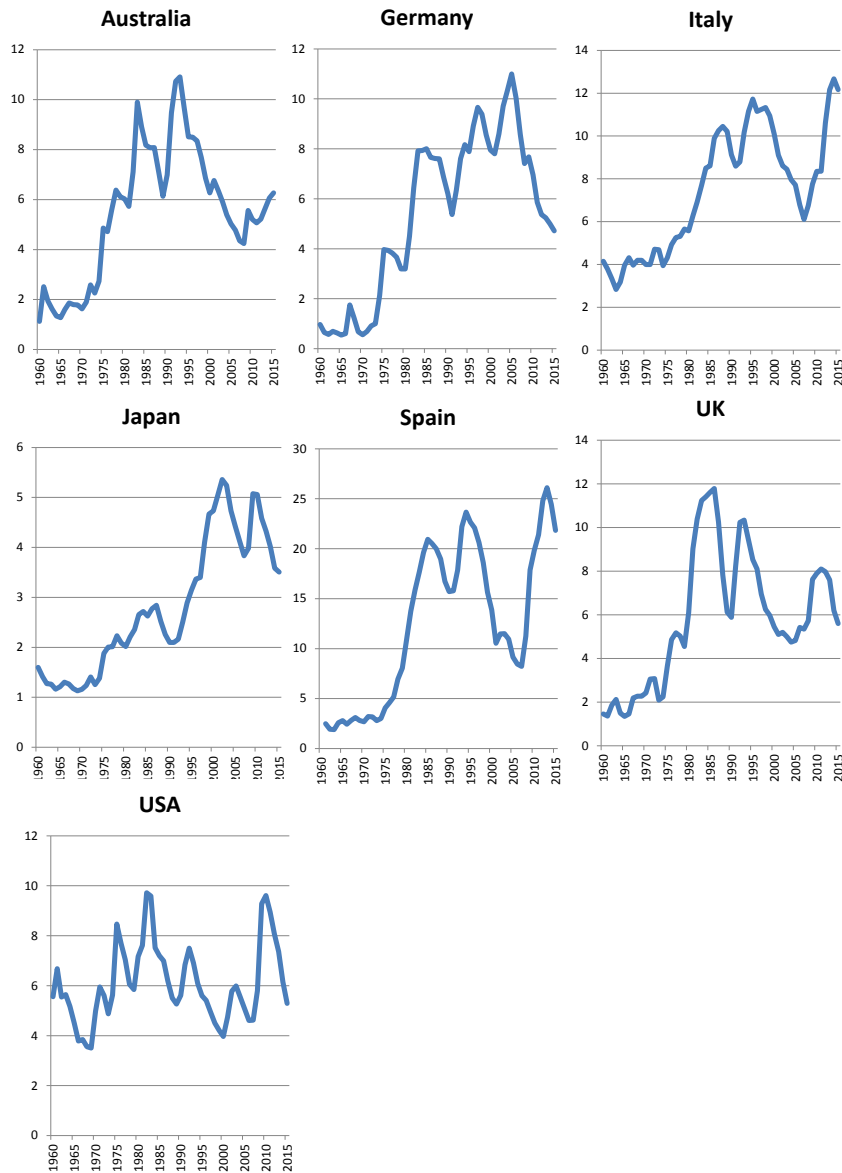
### Unemployment

One of the stark facts about modern economies has been the way in which unemployment has evolved over the last three or more decades. While different nations have recorded varying experiences, the common thread is that unemployment rates have risen overall and, in most cases, endured at higher levels for many years.

In Figure 2.1, the unemployment rates – the percentage of willing workers who are unable to find work – are shown for the seven nations depicted in Table 2.1 from 1960 to 2015. Please note that the vertical scales are different.

The accompanying data in Table 2.2 provides further information upon which to assess the historical behaviour of unemployment.

**Figure 2.1 Comparative unemployment rates, per cent, 1960 to 2015**



Source: Australian Bureau of Statistics; Federal Statistical Office, Germany; National Institute of Statistics, Italy; Ministry of Finance, Japan; National Statistics Institute, Spain; Office of National Statistics, Britain, Bureau of Economic Analysis, US.

The data show that unemployment rose in all nations shown during the 1970s and persisted at these high levels well into the first decade of the new century. Unemployment rates in Japan have been significantly below that of the other nations shown.

The data also show quite clear cyclical patterns. Australia is an example where cyclical patterns have been pronounced. Unemployment was below 2 per cent for most of the early post-World War II period and then rose sharply in the mid-1970s and continued rising as the economy went into a deep recession in the early 1980s.

**Table 2.2 Average unemployment rates by decade, per cent**

	Australia	Germany	Italy	Japan	Spain	UK	US
1960-70	1.7	0.8	3.8	1.3	2.6	1.8	4.8
1970-80	4.1	2.5	4.8	1.7	5.0	3.8	6.3
1980-90	7.5	6.7	8.5	2.5	17.3	9.2	7.1
1990-00	8.5	7.8	10.4	3.2	19.0	7.8	5.6
2000-10	5.4	8.7	8.0	4.7	12.1	5.7	5.9
2010-15	5.6	5.5	10.7	4.2	23.1	7.2	7.6

Source: See Figure 2.1.

Economic growth in the second-half of the 1980s brought the rate down from its 1982 peak but never to the level that had been enjoyed in the 1950s, 1960s and early 1970s.

The 1991 recession then saw the unemployment rate jump up again very quickly and reach a peak higher than the 1982 peak. The unemployment rate started to fall again as growth ensued after the recession was officially over but it took many years to get back to levels prior to the 1991 downturn. The US follows a similar pattern, although unemployment rates were higher in the early post war period but lower than Australia's in the 1990s. The GFC largely bypassed Australia but led to high unemployment in the USA, which has fallen somewhat since.

Unemployment rates tend to behave in an asymmetric pattern – they rise very sharply and quickly when the economy goes into a downturn in activity but then only gradually fall over a long period once growth returns.

Any credible macroeconomic model needs to provide convincing explanations for these movements. How was unemployment kept at low levels during the 1950s and 1960s? Why did unemployment rates rise in the 1970s and persist at the higher levels for several decades? What determines the cyclical pattern of the unemployment rates – that is, the asymmetry? Is there a behavioural relationship between the GDP growth data shown in Table 2.1 and the unemployment data in Table 2.2?

In answer to the first two questions, MMT would refer to the key proposition in macroeconomics that total spending determines output and employment, and indirectly unemployment. Then variations in unemployment must be attributable to variations in total spending.

On the other hand, some orthodox or mainstream economists claim that variations in output and employment occur due to decisions made on the supply side of the economy. Workers decide whether they wish to work under prevailing wages and conditions. Then high unemployment is caused by high quit rates, which is a supply side phenomenon.

If this were true, the onus would then be on these economists to explain why in 2008 the unemployment rate rose quite dramatically in four of the six countries depicted in Figure 2.1, but not in Australia and Japan, and to a certain extent, Germany. However, quit rates are procyclical, so workers tend to quit their current jobs, when plenty of alternative job opportunities are available. Thus high unemployment is not associated with high quit rates, which suggests that a demand side (spending) explanation of variations in unemployment is more plausible.



## Real wages and productivity

In 1957, the renowned British economist Nicholas Kaldor wrote an article in *The Economic Journal*, about the nature of long-term economic growth. He noted that there were six “remarkable historical constancies revealed by recent empirical investigations” (page 591), which he later considered to be the stylised facts regarding economic growth. He noted that these constancies were not necessarily immune to cyclical variation (as the economic cycle moves up and down), but were relatively constant over longer periods.

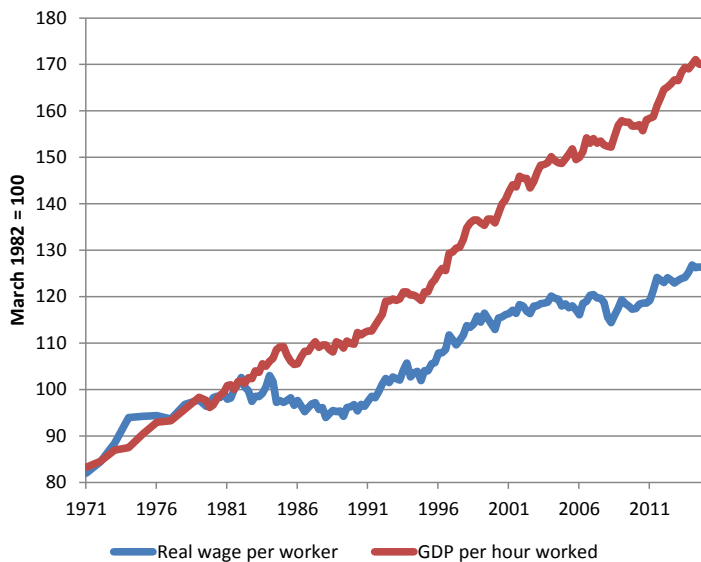
Among his stylised facts of economic growth was the observation that:

... the share of wages and the share of profits in the national income has shown a remarkable constancy in ‘developed’ capitalist economies of the United States and the United Kingdom since the second half of the nineteenth century (pages 592-93).

This observation was repeated by many economists for other nations in terms of the distribution of national income between labour (wages) and capital (profits).

We will learn in later chapters that for the share of wages and the share of profits in national income to remain constant over time, real wages must grow in proportion with labour productivity. Real wages are the purchasing power equivalent of the wage a worker receives in money terms. Labour productivity is the output that is produced per unit of labour hour. Kaldor’s stylised fact in relation to national income shares thus meant that real wages grew in proportion with labour productivity over a long time period.

**Figure 2.2 Real wage and productivity indexes, Australia, 1971 to 2015, March 1982=100**



Source: Australian Bureau of Statistics, National Accounts.

Figure 2.2 picks up the story in early 1971 for the Australian economy, which is representative of the trends that have been observed over this time period in a number of advanced economies of

the world. Up until the early 1980s, real wages continued to grow in proportion with labour productivity (GDP per hour worked).

After 1981, a gap opened up between these two time series and has widened ever since. Since March 1982, labour productivity has grown by around 70 per cent and real wages have only increased by around 28 per cent in Australia. Don't worry if you are having trouble interpreting the graph and its underlying data at this stage. During the course of this textbook, we will develop the techniques necessary to allow you to achieve competence when viewing empirical material.

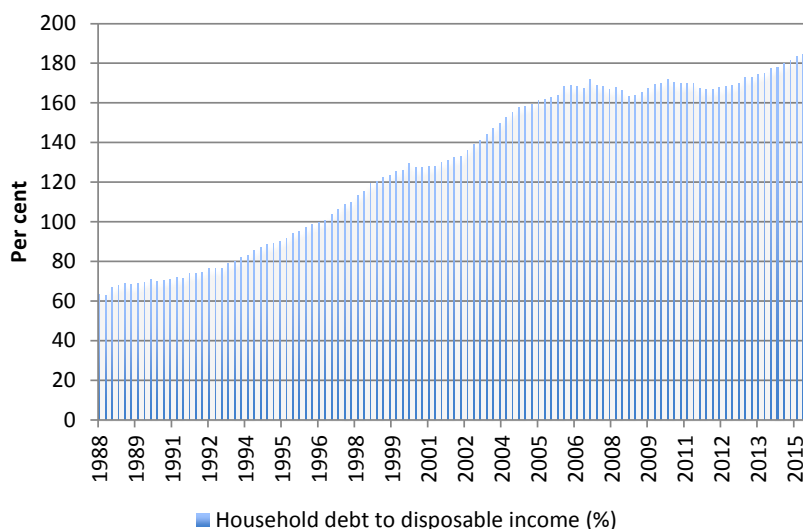
In terms of shares of national income, the growing gap between real wages and labour productivity has meant that there has been an on-going redistribution of real income away from workers (wages) towards capital (profits). In Australia, the wage share has dropped from around 60 per cent in the early 1980s to around 52 per cent in 2015.

How do we explain this shift in national income shares? Why did Kaldor's stylised constancy of national income shares end? What are the implications of such a substantial redistribution of national income away from real wages, which have until the last few decades been the primary driver of household consumption expenditure? What other factors now influence the growth in household consumption expenditure?

### Private sector indebtedness

Taking the example of Australia again (as representative of what has happened elsewhere in the advanced world), Figure 2.3 shows the rise in household debt as a share of disposable income since the early 1970s. Prior to 1988 (the beginning of our sample), the ratio was relatively steady at around 60 per cent. In the 1990s, the ratio began to rise and by the early 2000s had reached more than 150 per cent. The ratio fell marginally during the Global Financial Crisis but has since begun to increase again and by the end of 2015 was above 180 per cent and rising.

**Figure 2.3 Household debt to disposable income ratio, Australia, per cent, 1998 to 2015**



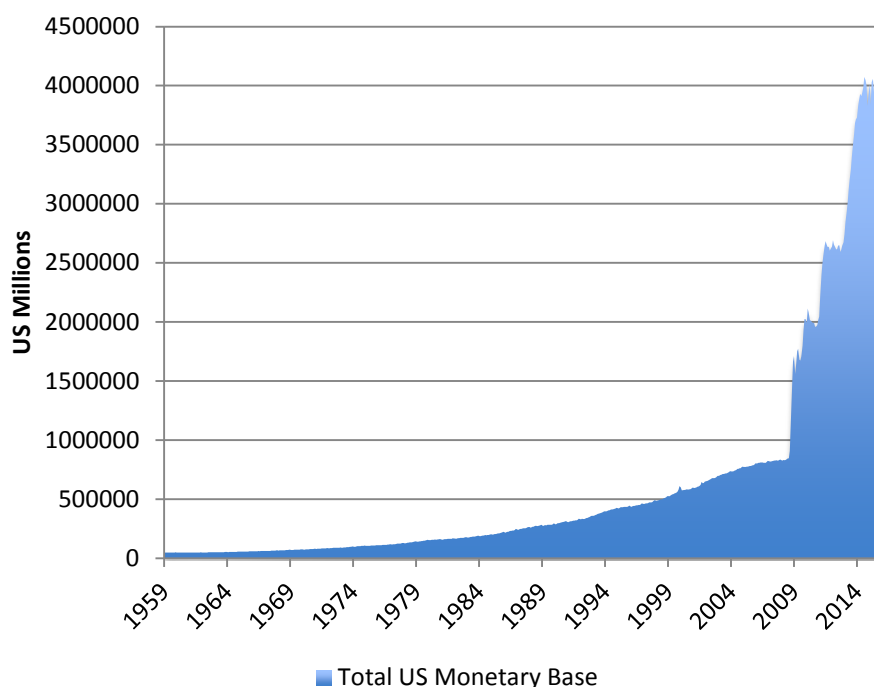
Source: Reserve Bank of Australia.

Is this large increase in the household debt ratio linked to the distributional shifts in national income implied by Figure 2.2? What other factors might explain this shift? What are the implications of the elevation in the household debt to disposable income ratio? Was the Global Financial Crisis linked to this movement?

### Central bank balance sheets

Figure 2.4 shows the so-called monetary base of the US economy administered by the Federal Reserve Bank. We will learn about the monetary base in later chapters but for now we can simply consider it to be the total reserves of the US banking system held at the central bank (the Federal Reserve Bank) plus currency (notes and coins) in circulation. The monetary base represents liabilities on the balance sheet of the US central bank. Up until 2008, the monetary base was predominantly comprised of currency on issue. In December 2015, bank reserves were around 65 per cent of the total monetary base and that proportion had increased in the period from 2008.

**Figure 2.4 US Federal Reserve Bank monetary base, 1959 to 2015, US dollar millions**



Source: Federal Reserve Bank, US.

In January 2008, the US monetary base equalled \$US830,632 million. It then accelerated upwards very quickly and by December 2015, stood at \$US3,835,800 million, a huge increase by any standard.

The rise in bank reserves at the US central bank is not an isolated event and similar balance sheet shifts have occurred in recent years in other nations (for example, Japan and the UK). Many

mainstream economists predicted that the substantial rise in central bank reserves would flood each economy with money and cause inflation. History tells us that inflation is low and in retreat.

How do we explain this massive shift in the balance sheet of the US Federal Reserve Bank? What are the implications of this shift? How does the monetary base relate to the money supply? Can the central bank carry liabilities of this size indefinitely?

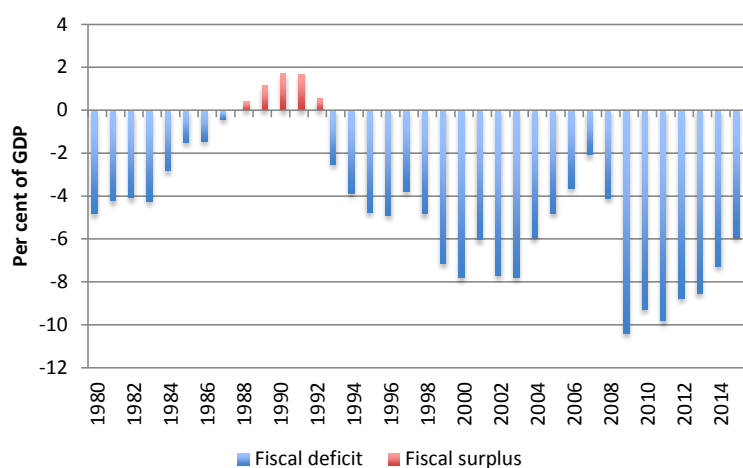
### Japan case study

Consult almost any macroeconomics textbook and you will find the following propositions stated in some form or another:

1. Persistent fiscal deficits push up short-term interest rates because the alleged need to finance higher deficits increases the demand for scarce savings relative to its supply.
2. The higher interest rates that result, undermine private investment spending (the so-called 'crowding out' hypothesis).
3. Persistent fiscal deficits lead to bond markets demanding increasing yields on government debt.
4. The rising public debt to GDP ratio associated with the persistent fiscal deficits will eventually lead bond markets to withdraw their lending to the government and the government will run out of money.
5. Persistent fiscal deficits lead to accelerating inflation and potentially hyperinflation, which is highly detrimental to the macro-economy.

Japan was the second largest economy after its reconstruction following the Second World War led to spectacular growth in the 1960s. It is now the third largest economy behind the United States and China. The period since 1990 provides a very interesting case study for macroeconomists because it has been marked by a number of macroeconomic outcomes, which are at odds with orthodox thinking.

**Figure 2.5 Government fiscal balance as % of GDP, Japan, 1980 to 2015**



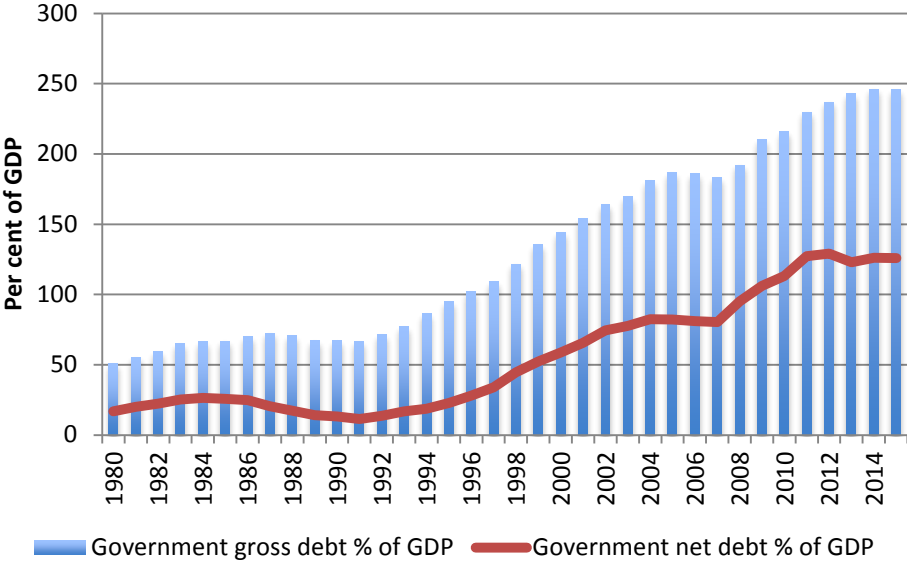
Source: IMF World Economic Outlook dataset (<http://www.imf.org/weo>).

As we can see in Figure 2.5, Japan has run a persistent deficit since 1992. A massive build-up of private indebtedness associated with a real estate boom, accompanied the five years of fiscal surpluses from 1987 to 1991. The boom crashed spectacularly in 1991 and began a period of lower growth and the need for higher deficits. The convention in Japan is that the national government matches its fiscal deficit with the issuance of bonds to the non-government sector, principally, the private domestic sector in Japan.

Figure 2.6 shows the evolution of the public debt levels as a per cent of GDP since 1980. Gross public debt is the total outstanding public debt issued by Japan’s national (general) government sector. But the Government also has investments itself, which deliver returns and when we subtract them from the Gross public debt we get the Net public debt.

Unsurprisingly, given the institutional practice of issuing debt to the private bond markets to match the fiscal deficits, the debt ratio has risen over time as a reflection of the on-going deficits that the Japanese government has been running to support growth in the economy and maintain relatively low unemployment rates (see Figure 2.1).

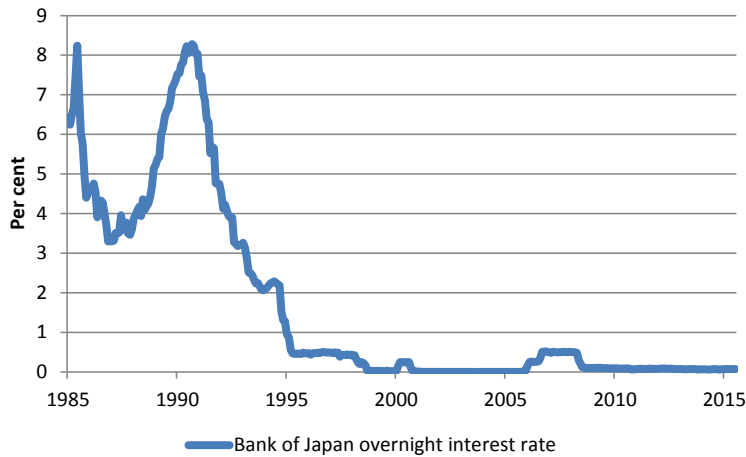
**Figure 2.6** Gross and net public debt as % of GDP, Japan, 1980 to 2015



Source: IMF World Economic Outlook dataset (<http://www.imf.org/weo>).

If the neoclassical propositions summarised above correctly captured the way the real world operates, then we should have expected to see rising interest rates, increasing bond yields, and accelerating inflation in Japan, given the persistent fiscal deficits.

**Figure 2.7 Japan overnight interest rate, per cent, July 1985 to December 2015**



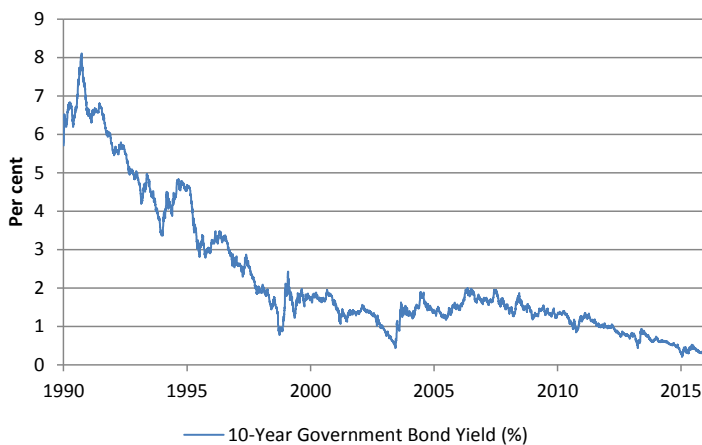
Source: Bank of Japan ([http://www.stat-search.boj.or.jp/index\\_en.html](http://www.stat-search.boj.or.jp/index_en.html)).

Did the persistent fiscal deficits in Japan drive up interest rates and government bond yields? The answer is clearly no! Figure 2.7 shows the overnight interest rate in Japan, which is administered by the central bank, the Bank of Japan. This is the interest rate that banks use to borrow. It has been exceedingly low and has not responded adversely to the persistent fiscal deficits.

Figure 2.8 shows that long-term (10-year) bond yields (interest rates) on government debt are also very low and have not responded adversely to the persistent fiscal deficits. There is no suggestion that bond market investors have become increasingly scared of buying the Japanese government bonds. If investors considered the government debt had become increasingly risky to purchase they would have demanded increasing yields to compensate for that risk.

The corollary is that the investors have also not signalled an unwillingness to purchase the debt and demand for the bonds remains high and yields remain low.

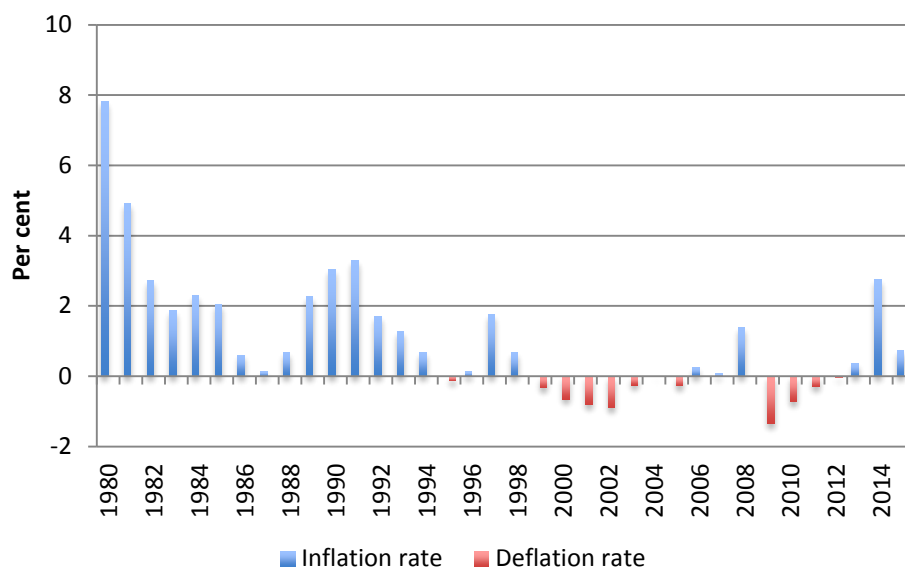
**Figure 2.8 Japan government 10-year government bond yield, per cent, 1990 to 2015**



Source: Ministry of Finance, Japan ([http://www.mof.go.jp/english/jgbs/reference/interest\\_rate/historical/jgbcme\\_all.csv](http://www.mof.go.jp/english/jgbs/reference/interest_rate/historical/jgbcme_all.csv)).

Figure 2.9 shows the inflation and deflation rates for Japan between 1980 and 2015. Inflation occurs when there is an ongoing increase in the general price level whereas deflation describes a situation when the general price level is continuously falling (negative inflation). You can see that in the period after the property boom crashed and the Japanese government began to run persistent and, at times, large fiscal deficits, the inflation rate has been low and often negative. There is clearly not an inflationary bias in the modern Japanese economy, as predicted by the mainstream economic theories.

**Figure 2.9 Inflation and deflation in Japan, per cent, 1980 to 2015**



Source: IMF World Economic Outlook dataset (<http://www.imf.org/weo>).

The above evidence shows that, despite persistent deficits and a rising public debt to GDP ratio, along with a downgrade of Japan’s credit rating by international ratings agencies, including Fitch in April 2015, international bond markets have not ‘punished’ the Japanese government with high 10-year interest rates on public debt nor has the central bank lost control of the overnight interest rate. Second, the persistent deficits have not lead to high rates of domestic inflation.

It is clear that the mainstream macroeconomic explanation of the relationships between fiscal deficits, interest rates, bond yields and inflation rates is unable to adequately capture the real world dynamics in Japan. Such a categorical failure to provide such an explanation suggests that the mainstream theory is seriously deficient. An MMT explanation of these empirical outcomes will be provided in Chapter 14, when students will have developed a thorough understanding of the workings of a modern monetary economy with a sovereign currency and the operation of fiscal and monetary policy.

### Summary

These examples demonstrate that macroeconomics is a highly contested discipline in terms of theory and policy prescription. When assessing the statements made by financial commentators and economists in the public debate, one has to continually refer back to the stylised facts.

It is important that students gain familiarity with the language of macroeconomics and understand the key concepts and theories, which will be developed in the following chapters.

## Appendix

### The Buckaroos model

A modern monetary economy is characterised by a currency regime, whereby transactions between economic agents (e.g. households, firms, financial institutions and government) can take place. This may involve, for example, the purchase of goods and services by households from firms; the purchase of assets (by households and firms); the payment of taxes to the Government or the receipt of transfers (e.g. unemployment benefit) from government.

The real world Buckaroos model demonstrates the roles of the currency, spending and taxes in a simplified economy.

At the University of Missouri at Kansas City (UMKC) in the USA, students are required to undertake a specified number of hours of Community Service (CS) during each year of their degree program. Failure to complete the required hours of Community Service over the duration of the student's degree program has negative implications for the final grade the student receives. The Economics Department ran the initial pilot program and designed a monetary system to administer the scheme. We briefly outline the scheme below.

Each student is assumed to be subject to a community service tax of 25 hours work per semester, payable to the University Treasury. Assume there are University- approved community service (CS) providers (for example, child care, aged care, environmental services, etc.) who submit bids for student hours to Treasury. Treasury awards paper notes (let's call these Bs as in 'Buckaroos') to the CS providers (assuming health, safety and environmental standards are met). In this economy assume one hour of 'average community work' is equal to B1. Paper notes are printed, with the inscription 'this note represents one hour of community service by a UMKC student'.



For example, Treasury may agree that students can do a total of 100 hours of work this semester at, say, the XYZ not-for-profit agency, which provides support for elderly people who are living alone. Treasury provides XYZ with B100, enabling 100 hours of student labour to be purchased.

CS providers then draw on their Bs to pay students for their hours of service. This can be considered 'spending' by the University Treasury, through the CS provider. If the student has



undertaken 25 hours of CS in the semester, then they can then pay their B25 tax, when they return these Bs to the University Treasury. This transfer of Bs by each student to the Treasury extinguishes their tax liability for the semester.

The University Treasury burns the Bs received from students, or stockpiles them to be used for future Treasury spending - whichever is more cost efficient. The number of Bs supplied to any CS provider is limited by its need for student labour but also its ability to attract student workers.

### **Implications of the Buckaroos model**

Treasury is the only source of Bs, which cannot be counterfeited. Treasury cannot collect B taxes until it has spent some Bs. Treasury can only be deemed to have spent when Bs are handed over to students for work done. Treasury cannot collect more Bs in payment of taxes than it has previously spent.

A possible Treasury outcome is a 'balanced budget', with tax 'revenues' equalling B spending. Thus Bs acquired by CS providers from Treasury are used to buy student labour which are then returned to Treasury as tax payments by the students. On the other hand, a surplus (deficit) arises in say Semester 1, if total Treasury spending is less (more) than the total taxes collected over that period.

### **References**

Centre of Full Employment and Equity (CofFEE) (n.d.) One hundred dogs and 94 bones, available at: [http://e1.newcastle.edu.au/coffee/pubs/briefs/dogs/dogs\\_and\\_bones.cfm](http://e1.newcastle.edu.au/coffee/pubs/briefs/dogs/dogs_and_bones.cfm).

Kaldor, N. (1957) 'A Model of Economic Growth', *The Economic Journal*, 67(268), 591–624.

# Chapter 3: A Brief Overview of Economic History and the Rise of Capitalism

## Chapter Outline

- 3.1 Introduction
- 3.2 An Introduction to Monetary Capitalism
- 3.3 Tribal Society
- 3.4 Slavery
- 3.5 Feudalism
- 3.6 Revolts and the Transition to Capitalism
- 3.7 Capitalism
- 3.8 Monetary Capitalism
- 3.9 Global Capitalism
- 3.10 Economic Systems of the Future?

## Learning Objectives

1. Recognise that throughout history the dominant mode of production evolves and can be overthrown.
2. Acknowledge that capitalism has evolved and may change fundamentally in the future, so there is nothing natural or everlasting about our current mode of production.

### 3.1 Introduction

In this chapter we will briefly examine what makes capitalism different. It is important to understand that humans have not always organised their economies around money. Throughout most of human history, the economy either operated entirely without money, or with money playing a relatively unimportant role in the provisioning process. However, with the rise of capitalism, it is not misleading to say that money came to play a dominant role.

We should not, however, link use of money only to the capitalist economy. Money has existed for at least 4000 years, while capitalism's origins can be traced back approximately half a millennium. To be sure, money's origins are not really known and might never be known, but there is no doubt that money was used for thousands of years before capitalism rose to replace feudalism. This chapter starts by noting that capitalism takes different forms. We briefly analyse different modes of production which have preceded capitalism, namely tribal society, slavery and feudalism and then examine the transition to capitalism. The repercussions of global capitalism are explored and then we speculate about economic systems of the future.

### 3.2 An Introduction to Monetary Capitalism

Today all of the major nations have economic systems that conform to the general structure that is called capitalism. Sometimes these are inaccurately called **market systems**—a term that is both too general (markets predate capitalism by thousands of years) and too narrow (while markets are certainly important to capitalist systems, they are only a part of the economy). They are also called **mixed economies** to indicate that the government sector as well as the private sector is important in the economic processes.

A somewhat more technical description used by some economists (including Marx and Keynes) is **monetary production economy** that captures the primary purpose of production for profit denominated in the money of account. While that does draw attention to the importance of money and the profit motive, it, again seems to neglect the role played by government—which is not operated for monetary profit. For that reason, the simple term **capitalism** seems more appropriate for our purposes.

It should not be thought that there is only one monolithic form of capitalism—a one size fits all version with carefully delineated institutions, rules of behaviour, and roles for government and other sectors. Capitalism takes a wide variety of forms, from a system comprised mostly of small-scale firms employing simple tools with a lot of the production farmed out to households, to a system utilising modern large-scale industrial production with literally thousands of highly skilled and unionised workers per factory.

Capitalist firms might operate under the constraints of **dog eat dog** cut-throat competition, or they might be organised into large cartels that carefully control competition for mutual advantage. Capitalism can be mean as described in the works of Charles Dickens, with most families eking out a miserable existence on low wages and long hours of work. Or it can be generous, with a strong unionised workforce demanding good working conditions, adequate pay, and a social safety net that takes care of aged persons, persons with disabilities, and children.

Capitalist systems can perform well, with rising living standards for most people, and they can collapse into great depressions as they did in the 1930s. They can grow fairly rapidly for long

periods (Italy after 1960, or Japan until 1990) or they can stagnate with slow growth (Japan after 1990).

Finally, capitalist systems can have big governments that actively manage the economy to the benefit of the majority of the population, or they can have downsized neoliberal governments that cater to the rich and powerful even as unemployment and poverty rates rise.

We might say that there are capitalisms, not a single kind of capitalism.

Finally, even if we recognise that there are many forms of capitalisms, it is equally important to realise that capitalism is by no means the only kind of economic system. As we discuss in the following subsections, humans have lived in other kinds of systems, and might choose to live in as yet unknown forms in the future.

### **3.3 Tribal Society**

This is not the place for a detailed analysis of the history of the development of capitalism, but we will provide a brief outline. We begin with brief examinations of other general forms of economic organisation. Historically, humans lived first (and for the longest period) in tribal societies. Both native Americans and native Australians lived in varieties of tribal forms of organisation at the time of the invasions by Europeans—as did all Europeans until the rise of Greece and Rome a few thousand years ago.

While recognising that there are many forms of tribal societies, we are able to generalise because they did share many characteristics—at least at earlier stages. First, tribes operated as egalitarian, communal, kinship-based social organisations. Members of a tribe were related by blood, with rules regarding marriage, initiation (and adoption of new members who were not related by blood), and expulsion.

Typically, tribal society was matrilineal (one's heritage was traced through the mother's side) and matrilocal (upon marriage, the male joined the female's family) although there were also examples of patrilineal (patrimony traced through father's side) and patrilocal (wife joined the husband's family) tribes. (It is possible that development of patrilineal and patrilocal practices came later.)

An egalitarian society is one in which members have equal rights and responsibilities, although there could be gender-specific and age-specific distinctions. It should also be noted that some tribal societies also practiced slavery (generally, captured enemies were enslaved, killed, or adopted) so that egalitarianism did not apply to some living within a tribe.

Finally, a communal society is one in which production and distribution is undertaken by all of the members according to well-defined rules of participation. In other words, an individual (or individual family) would not be responsible for deciding what to produce, how to produce it, and when and how to consume the production. Instead, the tribe would decide what to produce, how to produce it, and how to distribute the fruits of the tribe's labour among its members. These decisions would follow custom, although adaptations would be made over time.

The great anthropologist, Margaret Meade, observed the complex rules adopted by a particular tribe regarding distribution of the meat from a hunt and of crops from farming.

There is some dispute over the possibility that tribal society made use of practices that might approximate market activity. It is doubtful that markets would have been used within a tribe, as

production and distribution followed communal practice. An individual would not decide to produce something in order to sell it or to formally exchange it with another producer in the same tribe for some other item. Production and distribution decisions were made communally, so there would be no need to exchange essential items.

It is, however, well-known that members of tribal society had elaborate ceremonies of gift-exchange (something like Christmas gift-giving), but the items exchanged usually had little practical value. The most likely purpose of these ceremonies was to bring the members of the tribe together to enhance social relationships. Further, it was common to offer gifts to the family of the bride at marriage ceremonies—often called **bride-price**. However, to view this as a market in which one buys a bride certainly seems to be erroneous—as silly as viewing Christmas gift exchange around an evergreen tree as market activity.

Perhaps the activity that comes closest to something that we might be willing to view as market exchange was the practice of gift-giving between tribes. Some researchers have claimed that there are examples in which such exchanges involved trade of useful items that the receiving tribe would not otherwise be able to obtain. For example, one tribe that lives in a rain forest might offer products that can only be produced from rain forest resources, while another tribe that lives on grassland plain offers produce from its environment. In this case, the exchange of gifts is mutually beneficial in terms of improving living standards in each tribe, while also enhancing the social relations between the two tribes (reducing warfare).

Some might view this as akin to a moneyless market exchange, called barter. Even so, it is obvious that most provisioning is done within the tribe and through communal production and distribution that does not involve markets, and there is little evidence for activity within tribal society that came close to markets with sales and purchases utilising IOUs denominated in a money of account.

### 3.4 Slavery

We noted that slavery existed in some tribal societies. There are also entire economies that are based on slavery, in which a large portion of the production of the essentials of life is done by the slaves. A relatively recent and well-known slave society existed in the southern states of the US until the Civil War of the 1860s. Students are probably also familiar with the slave societies of Greece and Rome.

Production decisions are mostly made by the owners of slaves, who also own the output of the slaves. Slaves can be bought and sold in markets, although there can be wide variations of the laws governing treatment of slaves and their families. Typically, production by slaves is mostly used for consumption by the owner of the slaves and for subsistence of the slaves, however, slave production can also be used to provide goods and even services sold on the market.

Like tribal society and capitalist society, there are different forms of slave society. Some are much harsher in their treatment of slaves; some allow greater freedom for slaves or at least for their children. Some allowed slaves to gain freedom with the human rights enjoyed by other members of society. The US version of slavery was particularly repulsive because it was combined with virulent racism that denied that blacks could even become human. By contrast, slavery in the ancient societies was not based on racism, so that freed slaves could gain rights of citizenship.

However, the most important point to recognise about slave society is that it is operated for the benefit of the slave owners that are relatively few in number, and that the (typically) larger number of slaves recognise that their lives would improve through revolution and emancipation. Even in the most enlightened form of slave society, force is required to preserve slavery.

Further, because most of the benefits go to the owners of slaves, there is little incentive for slaves to increase productivity and expand output. Overseers are required to ensure at least a minimum work effort. Technological advance tends to be slow—both because slaves have little incentive or opportunity to innovate, and because more complex means of production are typically easier to break and more costly to repair.

Slave societies are inherently weak and subject to revolt (many students will recall the name of the most famous slave to lead a revolt—Spartacus). When faced with a military invasion, slave societies cannot arm slaves out of fear that the weapons would be turned against slave owners. Further, slaves are likely to use the opportunity of the invasion to initiate their own revolt. For these reasons, slave societies tend to be unstable.

### 3.5 Feudalism

Most students are at least passingly familiar with another important economic system, called feudalism. (In Western Europe, the feudal period more or less coincides with the time period called Middle Ages or Medieval period, but it is more accurate to use the term feudalism.) Knights and castles, lords and peasants, sword battles and jousting matches are all part of the western lore passed down for generations. Western and Eastern Europe as well as China and Japan each had a long experience with their own versions of feudal society. In Western Europe, feudal society emerged out of the fall of Rome in the fifth century BC, and lasted for a thousand years, although its institutions were beginning to break down by the 11th century, with a nascent capitalism beginning to replace feudalism from the 13th to 17th century (depending on the region). As is always the case, one cannot put an exact date either on the beginning of an economic system, or on its death. And, as always, there are different versions of the system we call feudalism.

For Western Europe, the most characteristic form reached its peak from the 7th to the 11th centuries. The two primary classes were the peasants and the feudal lords. The peasants had a right based on custom to agricultural land, with periodic redistribution of the land among the peasants (to account for changes to family size, fertility of the land, and so on). Claims to the land typically went back as far as memories permitted - perhaps to pre-Roman, tribal society times. Planting and harvesting were still typically done communally (as in tribal society), however, each family would receive the output of the land to which they were assigned.

The right to farm the land should not be confused with modern conceptions of ownership of land—for a family was not free to **alienate** (sell) land. Indeed, a family could leave the region, and generations later an heir could return to claim a right to the land to farm based on the family's ancient customary right - even after a market in land had developed. The institution of private property in land was not consistent with the feudal system, and once it developed, private property in land would help to bring an end to feudalism.

The relatively small upper class consisted of the feudal lords. In some cases, these also had a customary right to land in the region. Because lords were lords, and lords didn't work, peasants

were expected to work the lord's allocated land, and to turn the produce over to the lord. The labour required to work the lord's allocated land was called rent—paid in the form of labour.

In other cases, the lords did not have a right to land, but rather, exercised a right—based on custom—to a portion of the output of each peasant's land. This portion was also called **rent**—paid in the form of agricultural output. The knights and other armed men who had sworn allegiance to the lord enforced these customs through the threat of force. They were at the service of the sheriff, charged by the lord to collect rent—whether in the form of labour or agricultural products.

Many students know the story of Robin Hood, his band of merry men (or thieves), and the evil Sheriff of Nottingham. The original version dates from feudal times, with Robin stealing from lords, battling their sheriffs, and siding with the peasants—robbing from the rich to give to the poor. The later version was adapted to take account of the rising power of kings, with a good King Richard added.

Over time, rents were gradually converted to money rents. Roman and Greek societies had money, and indeed had used coins. While these were scarce in the early days of feudalism—and were not necessary for maintenance of feudal relations—as coins and other forms of money became more common, lords would agree to accept them. Like the institution of private property in land, the growing use of money also helped to break down feudal society.

The payment of rent is often presented as an exchange, in which the peasants are buying protection from the feudal lord. In some respects this was true. The feudal lord would use his armed force to prevent other feudal lords from collecting rent from his peasants. However, the main protection peasants received was from their own feudal lord and his knights, much as a storekeeper buys protection from the Mafia today. If the storekeeper doesn't pay, the Mafioso breaks his leg, threatens his family, and sets the store afire. However, payment of protection money ensures that the Mafioso will also prevent other gangs from trying to obtain payment for protection. Similarly, if the peasant refused to pay rent, the knights would attack. If the peasant did pay, the knights would protect the peasant from attacks by other knights. The peasants would have been happier without the need for protection from feudal lords and knights, just as the storekeeper would be better off without rival Mafia gangs.

### 3.6 Revolts and the Transition to Capitalism

Again, just as in the case of slavery, it was not difficult for the peasants to conclude that they would be better off without the feudal lords. Hence, just as there were continual revolts by slaves in slave society, peasants revolted periodically against lords. The most famous revolts in the West took place in 1381, and were moderately successful at forcing concessions from the lords. (The movie *Braveheart* depicts a peasant revolt that was also nationalistic as peasants sought to drive out foreign lords and their armies.)

By that time, many changes had already occurred in the nature of European feudalism. One important change was the increasingly common practice of paying rent in the form of money-denominated IOUs. Another change that was advanced by the 1381 revolts was the development of the recognition of property rights in land—moving toward alienability of land. This, combined with the enclosure movement, helped to bring an end to the feudal system in Europe.

The enclosure movement is fairly well known to students. Originally, a portion of the land in each region was preserved as a **commons**, containing forests, pastures, wetlands, and other land that was not farmed. The commons was important to peasant families, as a source of wood for building and fires, for game, and for grazing cattle. However, feudal lords over time gradually exerted a claim to the commons, using threat of force to keep the peasants out. The lords also claimed all of the game and other resources of the commons. (Robin Hood, of course, lived illegally in the lord's forest, taking the lord's game and attacking carriages of the rich when they tried to pass through the forest.)

The **enclosure** (that is seizing) of the commons made an already difficult life truly unbearable for the peasants—they could no longer supplement their meals with game, they could not collect wood, and they could not graze their cattle. As a result, some would look for paid work to supplement their meagre output from farming; some would even sell their land and abandon farming altogether. Wage receipts and receipts from the sale of land would often end up in the hands of the feudal lords, as peasants paid overdue rent.

As we can see, portions of the economy became increasingly monetised as feudal relations broke down. Working for wages became more common. Payment of rents in money form rather than in terms of labour or agricultural produce, became more common. Land was bought and sold, displacing the customary rights to land.

At the same time, cities were becoming more important, acting as magnets for peasants who were leaving the land. In cities, one could perhaps find a position as an apprentice in a handicraft shop, learning skills producing furniture, silverware, or shoes. With luck and hard work, one might advance to a position as master craftsman. Markets became increasingly important through specialisation—the craftsman would produce shoes for market, and use proceeds from sales to purchase food and other necessities and perhaps even some luxuries. Peasants, too, could sell a portion of their output, paying money rent and perhaps purchasing some consumer items they had previously made themselves (or perhaps had done without). Markets and money became increasingly important.

In addition to the enclosure movement, other tactics were used to force peasants from the land. Some were run off through violent attacks by armies of lords and kings. Others were displaced by seizure of Catholic church lands. The Catholic church was by far the largest feudal lord exerting control over and collecting rent from vast areas. After a dispute with the Vatican, King Henry VIII confiscated the church's lands in England. Some left voluntarily because they could no longer support their families. Others lost their land to creditors through excessive debt burdens.

The vacated land could then be consolidated for pastures, particularly for sheep in the case of Scotland. The wool was shipped to the growing textile manufacturing industries. At the same time, the displaced peasants had to find alternative means of livelihood, so many were converted to wage labourers working in those same manufacturing industries. Thus, taking land out of the feudal arrangement simultaneously created a displaced workforce for the rising capitalist sector as it cleared the land for agricultural products bound for capitalist production and hence to markets.

We gradually see a transformation of the economy to something that looks a lot more familiar: employment, marketed output, cities, and even factories. Rather than an economy based on lord and peasant, we see workers and their employers, the capitalists. Even in the agricultural areas—



formerly thoroughly feudal—we see owners of land employing wage labour to work the fields and tend the herds. A rising portion of output goes to market.

Marketed output seeks profits, measured in terms of the money of account. While markets and money denominated sales, and money denominated liabilities, are thousands of years old, they had never dominated the economy previously. Most people's livelihoods over the previous centuries and millennia had not depended on producing marketable output; most consumption had been satisfied by direct production of the consumer (or by their extended family).

With the breakdown of feudalism, all of that began to change rapidly.

### 3.7 Capitalism

The capitalist mode of production was altogether different from all previous economic systems. With the development of capitalism, most of the producers (workers) had no right to the things they produced, and they worked with tools and machinery owned by others—the capitalists. Indeed, most of the workers could not have produced much at all unless they worked for capitalists, because they had no other access to the necessary tools and machines. This is very different from feudal society, in which peasants had a customary right to the land, which was necessary for agricultural production.

The worker has no *right* to the means of production, and hence no means of securing a livelihood *unless* they can convince an owner of the means of production (a capitalist) to employ them for wages. This gives access to means of production, and the wage provides access to the means of livelihood. There is no guarantee that the worker will obtain employment, and no guarantee that even if they do, the wage will be sufficient to purchase the necessary means of livelihood to produce an agreeable living standard.

While workers are sometimes called **wage slaves**, capitalism deviates in important ways from slave society. It is true that slaves in a slave society also had to work for somebody, generally did not own the tools they used, and did not own the output they produced. However, slaves were never unemployed. If an owner did not need a slave, the slave would be sold to someone who did. Further, a slave could not normally quit and search for a different owner. Workers are usually free to quit their jobs and to seek alternative employment, however they can become unemployed because they cannot force capitalists to hire them.

In the mean capitalist systems, unemployed workers and their families would starve. In a slave society, by contrast, the rational owner of the slave would provide at least a subsistence level of necessities to protect the investment in the human property. Still, wage slavery (working for wages) is surely better than true slavery in which humans are reduced to the status of property of others.

All societies experienced change through time: customs and beliefs evolved; technology including know-how changed and improved; animal power replaced human power and then was replaced by the power of machines. The types of things—and how they were consumed—changed; populations migrated; civilizations rose and then fell. However, the pace of change accelerated almost unimaginably under capitalism.

In previous forms of economic organisation, children could expect to live a life not noticeably different from the life led by their parents, their grandparents, and even their great-grandparents. Economic growth (in the sense of output of the production of goods and services) generally

occurred over time, but it was so slow that it was barely noticed. Capitalism changed all of that. While it is certainly not true to state that capitalism always and everywhere improved living standards, there was a nearly inexorable longer-term trend toward increased output that could not be overlooked.

In a very real sense, the whole discipline of economics was created with the rise of capitalism, to make sense of this new form of society that was continually changing. Often, the standard of living of one's children would be very different from that of their parents, and certainly the children would face a new array of products and styles and jobs that could not have been imagined for them by their parents just 20 years earlier. The changes were obvious, and required the development of a field of social science that would explain the forces that drove them.

Capitalism represented a social revolution and indeed, the institutionalisation of mechanisms that ensure continuing change at a relatively rapid pace. It is very difficult for us today to comprehend how rapid change is today when compared to the experience of all generations of humans prior to the rise of capitalism. There are still people alive today in Australia and America who mostly travelled by horse-drawn cart and who lived in sod houses and who did without radio (and television!) when young. However, we should not be misled into thinking that rapid change is limited to the past half century; a similar pace of change was present even in the early days of capitalism.

The application of the term **industrial revolutions** (plural because there were more than one) reflects the perceived pace of change—so fast that it resembled revolution. Even those great critics of capitalism, Karl Marx and Frederic Engels, conceded that capitalism had **unleashed the forces of production** in a novel manner:

The bourgeoisie, during its rule of scarce one hundred years, has created more massive and more colossal productive forces than have all preceding generations together. Subjection of nature's forces to man, machinery, application of chemistry to industry and agriculture, steam navigation, railways, electric telegraphs, clearing of whole continents for cultivation, canalization of rivers, whole populations conjured out of the ground -- what earlier century had even a presentiment that such productive forces slumbered in the lap of social labor? (Marx and Engels, 1848)

These words come from the Communist Manifesto (1848), which went on to proclaim:

The weapons with which the bourgeoisie felled feudalism to the ground are now turned against the bourgeoisie itself. But not only has the bourgeoisie forged the weapons that bring death to itself; it has also called into existence the men who are to wield those weapons -- the modern working class -- the proletarians. (Marx and Engels, 1848).

Note how Engels and Marx celebrate both the rise of the monumental productive forces of **social labour** and also declare that social labour—the proletariat - will prove to be capitalism's undoing.

Before we turn to the **death of capitalism**, let us look in some detail at the capitalist system in which most of the people of the world live today.

### 3.8 Monetary Capitalism

Capitalism is also very different from all previous economic systems because of its thoroughly monetary nature. We have hinted that money and markets long pre-date capitalism. However, what is important about capitalism is that the *purpose* of production is different. In all previous economic systems, the immediate purpose of production is to generate real goods and services to be consumed—by the producers (members of the tribe, slaves, peasants) and by others (feudal lords and their knights, kings and queens, slave owners, ecclesiastical officials, and so on).

It is true that a portion of production always took the form of investment—tools, machines, infrastructure—that would be used in the future to produce goods and services for consumption. These were not available for immediate consumption by their producers. And, depending on the arrangements, some goods were produced for sale in markets even before capitalism. However, the majority of production never entered markets but rather was for the direct (even if not immediate) satisfaction of the tribe, the peasants and lord on the feudal manor, or the slaves and slave owner. And most of the goods and services consumed by the producers were not purchased in markets.

The capitalist form of production, however, is undertaken with a view to making monetary profits. It is true that the produced goods and services *will* be consumed (or form part of the productive capacity to be used for future consumption), however, this takes place only once they are sold.

Further, the production will not occur in the first place unless the capitalist believes the goods and services can be sold at a *profitable* price. No matter how badly the population *needs* to consume or *wants* to work, the production will not occur unless it is deemed profitable. Earlier we explained that because capitalists own most of the means of production, even if the vast majority of the population is eager to produce the necessities of life, production cannot take place unless the small minority that owns the means of production is willing to produce. That willingness, in turn, is mostly determined by expected profitability.

Not only is money the goal of production, it is also necessary to allow production to proceed in the first place. Marx famously described the capitalist production sequence as  $M \rightarrow C \rightarrow M'$ . Production begins with the use of monetary liabilities, used to purchase all of the inputs to the production process, including equipment, raw materials, and labour. This is represented by the initial **M**. The production process, in turn, results in **commodities**, or **C**. Commodities are goods and services that will be sold on the market; these include output that is sold directly to consumers as well as output sold to other firms to be used in other production processes. If all goes according to plan, the commodities will be sold at a sufficiently high price to reap profits. This requires that the total value of the money liabilities received in sales ( $M'$ ) is greater than the total value of the money liabilities used to engage in the production process ( $M$ ). That is, profits require that  $M' > M$ .

We can see that there are two main barriers to the production process that must be overcome: first, the capitalist must be able to obtain money liabilities ( $M$ ) to begin production, and second, the capitalist must believe that sales of production will generate monetary profits ( $M' > M$ ). Production can be prevented by either barrier.

Hence, in a *monetary production economy*, production begins with money on the expectation that it will end up with more money. In an important sense then, money can be blamed for

unemployment of labour and other resources. These labour and other resources sit idle when capitalists either cannot obtain money to start the production process, or if they believe that production will not be sufficiently profitable (in terms of money.)

### 3.9 Global Capitalism

As capitalism evolved, more of the production process was brought into the market. In the early days of capitalism, the family of the worker might still produce a large proportion of the food consumed by the family. Milk, cream and butter would come from the family's cow; the garden produced vegetables; and eggs came from the chickens. Much of the clothing and bedding would be made at home. Few services were purchased in the market.

However, there is a tendency for the capitalist form of production to continually expand into new areas. Today in modern developed capitalist countries, food mostly comes from global agribusiness, clothing is produced by large conglomerates employing cheap labour in Asia, and many of the services that households formerly performed for themselves are now bought in the market. For example, today in most US suburbs, even working class families hire gardening firms to mow the lawns—a service that would have been purchased only by the rich a few generations ago.

Not only does capitalism become more *intensive* in the sense that it continually expands its reach into new markets within a nation, but it also becomes more *extensive* as it spreads over the globe and brings all peoples into the capitalist form of production. Beginning less than five centuries ago in northern Europe, the capitalist mode of production now dominates production almost everywhere in the world.

However, we should not overstate the importance of capitalist production. Even in the most developed nations such as the US and Australia, much of the production that is absolutely essential for social survival takes place outside capitalist enterprises.

First, households still produce many of the goods and services required to support the family: rearing children, cooking meals, routine maintenance of housing, gardening, financial services (balancing the cheque book), entertainment (playing word games), and so on. Even if much of this *could* be purchased in markets, families perceive quality differences, and also enjoy working together. It is probable that reproducing healthy families requires that a large portion of family life be preserved from the reach of the market system.

Second, as we emphasise throughout this text, much of the production is better-suited to public organisation and provision rather than to for-profit production. In recent years, there has been a strong push by neoliberal politicians and think-tanks to down-size government while either abandoning its responsibilities or contracting-out services to private firms. This is justified by claims that private firms are more efficient and that the market produces the right incentives.

In some cases this is probably true, but in many others it opens the way for abuse, cronyism, and corruption. Further, since private firms are profit-seeking, they rationally prefer to provide goods and services to those who are willing to pay and can afford to pay. For these reasons, there will always be room for production outside the market, by families, by government and by not-for-profit organisations to meet needs that are not fulfilled by for-profit production.

The push for globalisation has been very strong in recent decades, as evidenced by various free trade deals (in the case of the US and its neighbours, the most important recent development was

NAFTA). There was a similar period of globalisation at the end of the 19th century. In both cases, imports and exports became relatively more important, and huge international corporations took substantial control of international trade. In both the late 1900s and in recent decades, finance was also internationalised to a great extent.

That earlier period of globalisation and its accompanying rise of international finance collapsed in the Great Depression of the 1930s. The US and other countries reformed finance, downsizing it and exerting more control over it. For a number of reasons, international trade became somewhat less important—and trade barriers were restored. However, over time, production, sales, and finance gradually became global—even more so than they had been in the early 20th century.

The global financial crisis of 2008 temporarily slowed the advance of global capitalism. However, the rescue of global financial institutions as well as of some of the huge global nonfinancial corporations (such as General Motors) by the US Federal Government seems to have renewed its advance.

### **3.10 Economic Systems of the Future?**

All economic systems evolve, but it is impossible to predict the direction of change. We can be sure that the economy will look different a hundred years from now, but do not know how different. From the vantage point of the early 21st century, the form taken by capitalism in the major developed countries appears to be environmentally and socially unsustainable. In a later version of this book we will examine some of the issues related to environmental sustainability.

In later chapters we will explore some of the social problems—particularly unemployment, inequality, and poverty—that result from the way that many modern economies function today. Undoubtedly, capitalism will continue to change, and informed policy can help to resolve these sorts of problems.

However, many critics of capitalism—currently and historically—foresee a day when capitalism will be replaced by alternative economic (and socio-political) systems. We will briefly outline two such systems: socialism and communism. We will distinguish these from capitalist economic systems, and from each other.

Our definitions follow those usually used by the major advocates of such systems. However, it must be noted that the following is necessarily conjectural because we are describing possible future economic systems and we have no way of knowing how things will actually turn out.

Finally, we realise that much controversy—and confusion—surrounds these terms. This is in part because several real world societies have variously claimed to be socialist or communist—or were accused by others of being socialist or communist. Here we set out clear definitions that are not meant to describe any of those real world economies.

A *socialist* economic system is one in which the means of production are collectively owned. In such a system, there is no functioning capitalist because private ownership of the means of production is prohibited. (There still might be those who would like to be capitalists, but their desires are frustrated by the prohibition.) To be sure, there is still private ownership—of clothing, of automobiles, of housing, and perhaps even of small family farms. However, a significant share of the means of production is not privately owned.

Without private ownership of means of production, there is no significant private employment of other humans. Employment of family members or perhaps others within the household or on a family farm might be permitted. Most workers would be employed in organisations with collective ownership of the means of production.

Alternative arrangements would be possible. At one extreme, the workers of the collective would share communal ownership of the factory and all of its associated tools, buildings, financial assets, and so on. The collective would make all production decisions—what to produce, how to produce it, and how to price it.

At the other extreme, all means of production would be owned and managed by the workers of the nation as a whole. Decisions concerning what to produce, how to produce it, and how to price it would be made at the level of the nation as a whole by representatives of the workers. In this case, the purpose of production would be to achieve national goals. Unlike capitalism, production would not be undertaken with a view to obtaining monetary profits.

As a simple example of the difference, in a socialist economy, necessities (food, clothing, shelter, medical care, education) would be assigned prices sufficiently low that all members of society could afford them. Prices of luxury goods or of harmful products and practices (tobacco, gambling) would be set high enough to discourage their use.

By contrast, in a capitalist economy, prices are set to ensure that capitalist owners of firms achieve desired profits.

A *communist* economic system shares some of the characteristics of the socialist system: there are no capitalists and no private ownership of the means of production. Further, no one wants to be a capitalist—all aspire to become workers and the practice of privately employing other humans is abhorrent to all members of society.

All production decisions are made democratically. Unlike socialism, there is no need for wages or prices. All production is freely and universally available to all. The forces of production are so great that all material needs and desires are easily satisfied, hence, there is no reason to ration output.

Further, many of the social problems that spur conspicuous consumption and invidious distinction will have been removed from society. Hence, the sort of profligate consumption (shop ‘till you drop) that is common in the developed capitalist economies will have disappeared. Rather than shopping malls and glossy advertisements that try to lure families to consume more than they need and want, there will be communal warehouses at which families can obtain whatever they need.

Further, the threat of deprivation will not be needed in order to induce people to work. All will want to contribute to society, and hence, will voluntarily participate in the social production process to the best of their ability.

There is a very simple way to distinguish between socialist and communist societies. The motto of socialism is: from each according to ability, to each according to her contribution to production (Marx, 1875). In other words, the distribution of social output is largely determined by the contribution to the production process. This means that inequality of the distribution of output will continue under socialism: those who produce more will receive more. While the distribution will be less unequal than the distribution under capitalism, some inequality will remain. Of course, just as in other economic systems, there will be some who cannot produce

very much. For example, people with disabilities, or people too young or too old to work, or parents with young children, might not be able to contribute very much to the production process. Thus, there will be some deviation from the socialist motto to ensure that all receive necessities.

The motto of communism is: from each according to ability, to each according to need. (Marx, 1875: 13-30). In this case, there is no attempt to, and no need to ration output on the basis of the contribution to production. This is because the communist economic system can easily satisfy all reasonable needs, and the members of such a system will not have unreasonable desires. Each will take only what they need. Compulsion is not needed because each will contribute as much as they can.

If we compare either of these systems to capitalism, it is obvious that there are big differences. In a capitalist system, one's income includes earnings that are due to one's ownership of the means of production, which allows one to employ others and to thereby receive income generated by the production of others. The capitalist owner receives profit income not because they work, but rather because they own the factories and other establishments in which production takes place.

The capitalist system concentrates ownership of the means of production in the hands of a few, and then all others must work for the capitalist owners to generate profit income for them. While it is commonly claimed that capitalists also contribute to the production process by providing entrepreneurial skills, in practice these skills can be hired—there will be a hired management team, a hired research and development team, and so on. Even after paying all of these teams, there still must remain profit income or the capitalist owners will not allow their means of production to be used. What the capitalist owners actually provide is the means of production that they have effectively monopolised.

Socialism and communism eliminate capitalist income by eliminating private employment and private ownership of the means of production.

Will the economic system of the future look anything like these alternative systems? We cannot know. There have been some real-world experiments to implement socialist economic systems (the Paris Commune in 1871, the USSR in 1917, China in 1949). So far, it appears that none of these have been able to build a viable alternative to capitalism (the Paris Commune was crushed, the USSR collapsed, and China appears to be moving toward capitalism, albeit in a form that is rather different from that in the western developed nations).

However, it must be remembered that the transition to capitalism required many false starts and several hundred years before it replaced feudalism throughout Europe. We can be sure that the economic system will continue to evolve and it is unlikely that capitalism is the final form of economic organisation. Capitalism, itself will likely undergo many transformations in the coming decades.

It also must be remembered that tribal society endured for perhaps tens of thousands of years and feudalism persisted in Europe for about a thousand years. By comparison, capitalism is still a young upstart.

## References

Marx, K. (1875) 'Critique of the Gotha Program', in Marx/Engels Selected Works (1970), Vol. 3, Progress Publishers, Moscow.

Marx, K. and Engels, F. (1848) Communist Manifesto, English translation by Samuel Moore, 1888. Available online: <http://www.anu.edu.au/polsci/marx/classics/manifesto.html>.



# Chapter 4: The System of National Income and Product Accounts

## Chapter Outline

- 4.1 Measuring National Output
- 4.2 Components of GDP
  - Consumption ( $C$ )
  - Investment ( $I$ )
  - Government spending ( $G$ )
  - Exports ( $X$ ) minus imports ( $M$ ) or Net Exports ( $NX$ )
- 4.3 Equivalence of Three Measures of GDP
  - Expenditure approach
  - Production approach
  - Income approach
- 4.4 GDP versus GNP
- 4.5 Measuring Gross National Income
- 4.6 GDP Growth and The Price Deflator
- 4.7 Measuring CPI Inflation
  - The CPI index
  - Rate of growth of the CPI index
- 4.8 Measuring National Inequality
- Appendix: Advanced Material
  - Chain weighted real GDP
  - Measuring net national income
  - Difficulties in using the CPI to accurately measure inflation

## **Learning Objectives**

1. Understand how GDP is measured and why different measures of GDP are equivalent.
2. Recognise the deficiencies of GDP as a measure of welfare.
3. Derive (CPI) indexes and their rate of change.
4. Interpret measures of income inequality

## 4.1 Measuring National Output

The System of National Income and Product Accounts (NIPA) is the framework assembled by national statisticians for **measuring** economic activity.

In this chapter we look at national income accounting - that is how we measure total national spending and its components as well as national income and its components. The most important measure of economic production is Gross Domestic Product or GDP. Let us first provide a formal definition.

**GDP is the measure of all currently produced final goods and services evaluated at market prices.**

Note that GDP is a flow measure, hence it must have a time dimension: month, quarter, and year are the most common periods over which the flow of production is measured.

Let us emphasize the most important parts of the definition:

**Currently produced:** This includes only goods and services produced over the time period, and would exclude goods sold this period that had been produced previously. Hence this measure excludes sales of 'used' goods.

**Final goods and services:** This includes only goods and services sold to final users - whether these are consumers, firms or government. Households buy final consumer goods and services; firms buy investment goods to increase capacity, and government buys goods and hires services. Intermediate goods and services are excluded. For example, an auto manufacturer buys tyres to put on new cars for sale. These are intermediate goods and if we were to count those tyres as part of GDP, and then count the value of automobiles produced, we would double-count the value of tyres (since the value of the automobiles would include all the intermediate goods and services that go into producing the automobiles). For that reason, we count only the value of final goods and services.

**Evaluated at market prices:** We calculate the value of final goods and services at market prices. This means that GDP is calculated at nominal values. We use another measure of GDP to take account of the impact of price changes, called real GDP. Note that unless specifically designated as 'real GDP' when we say GDP we mean nominal GDP, calculated at current market prices. We will discuss real GDP below.

A system used for calculating the value of a nation's output and income has been developed, called the National Income and Product Accounts, or NIPA. Note that the statisticians who compile these accounts must make many decisions about what to include and what to exclude. While the decisions are not arbitrary, it is important to recognise that they are conventions - in other words, there is nothing sacrosanct about them, and the conventions could be changed by agreement.

For example, washing your own dishes at home is not included in GDP. However, if you hire your neighbour to wash your dishes, that should be counted in GDP as dishwashing services. (Note that we said 'should' because if you pay your neighbour 'under the table' and neither of you report it, the transaction might not get captured in the official numbers.) This makes some sense because in the first case there was no monetary exchange and no market price at which the service took place, while in the second there is the market price that you paid for the service. However, by excluding all the unpaid household services performed including cleaning, repairs

and upkeep, and child and elder care, the NIPA numbers exclude a huge proportion of the nation's production.

More importantly, it undercounts the contribution made especially by women to production, since they perform a disproportionate amount of unpaid work. Many economists have called for reform of the accounting conventions to include more unpaid work in order to give greater recognition to the social value of 'women's work'.

GDP also excludes black market, grey market, and much of the production in the informal sector. This has largely to do with the difficulty of collecting the data. Black market transactions are illegal, even though the good or service, *per se*, may be legal. For example, the sale of cigarettes on which duty has not been paid is illegal. On the other hand, the drug and sex trade involves illegal transactions in illegal goods and services.

In the grey market legal, non-counterfeit goods are sold outside normal distribution channels. For example, if a brand of cameras is very expensive in a particular country, an enterprising local trader may import them from a country where the price is low and sell them in competition with the official supplier(s) of the camera. Many nations do attempt to estimate such activity and even include at least some of it in official measures of GDP. Much of the informal activity is similar to household production discussed above. For example, in many developing nations, much of the food production does not reach formal markets—it is consumed by farmers and shared or sold in local markets without being subject to proper recording. Other activity is 'under the table', and unrecorded to escape taxes. While the size of the black market is sometimes estimated in countries, typically it is not included in their official measures of GDP. However in late 2014 the office of national statistics in Italy announced that the estimation of its GDP would in future include illegal activities, notably "drug trafficking, prostitution and smuggling services (cigarettes and alcohol)." (The Economist, 2014).

Another problem is that GDP is not necessarily a good measure of production as a contribution to economic well-being. For example, a factory might pollute the air and water supply while it is producing 'widgets' (widget is a generic term for a produced commodity). The social, health, and environmental costs are not deducted from the value of the widgets produced for the purposes of measuring GDP. However, if society had to hire workers and produce machinery in order to clean up the pollution coming from the widget factory, that would be counted toward GDP. Ironically, production of widgets would then count twice toward GDP, once for the value of the widgets produced and secondly for the value of cleaning up the environmental mess.

Furthermore, if neighbours of the widget factory get sick from the pollution, then the healthcare spending required to treat them also gets counted in GDP. For that reason, GDP can be a poor measure of economic well-being, as the polluting industries might actually make a negative contribution to our general earning standard even as they increase GDP.

Still another problem is inequality. It does not make any difference to the calculation of GDP whether almost all production goes to the top 10 per cent of individuals or households, so that the bottom 90 per cent gets next to nothing. The GDP measure simply adds up production without taking account of the distribution of the output. This can make GDP a bad measure for comparing earning standards across countries.

It is common to divide a nation's GDP by its population, to derive per capita GDP. We can then rank nations according to per capita GDP, classifying some as rich, some as middle income, and

some as poor. However, per capita GDP simply provides a measure of the average and that can be highly misleading as a guide to the standard of living of the typical resident of a nation.

For example, the average could be \$35,000 per capita in two very different nations. In Country A, the share of GDP of the top 1 per cent might be 90 per cent, leaving the remaining 99 per cent to share only 10 per cent of the nation's output, while in Country B the distribution could be nearly equal, with 99 per cent of the population living within a few thousand dollars of the \$35,000 average. Clearly, economic well-being would be more widely shared in Country B, with very few poor people but also few people living much above the average. There is a measure of inequality that measures distribution called the **Gini Coefficient**, which we discuss later in this Chapter.

There are alternative measures of economic well-being that attempt to get around these problems. Some try to measure household production. Others take account of inequality, poverty, and access to education and healthcare. Some measures deduct social, health, and environmental costs. For example, our hypothetical widget factory just discussed might actually make a net negative contribution to economic well-being - it would be beneficial to close the factory and thereby increase social welfare even while foregoing consumption of widgets.

As a real world example, tobacco smoking increases GDP due to sales of tobacco, spending to capture smoke to make indoor air cleaner, and tremendous amounts of spending on healthcare for tobacco users and all those who suffer from the effects of second-hand smoke. Eliminating tobacco use would undoubtedly enhance well-being but might reduce GDP. For these reasons, when addressing economic, social, and environmental well-being, we need alternatives to GDP.

Still, GDP is the most commonly used measure and it does have one big advantage: it focuses largely on monetary value of output. As we have discussed, the profit motive drives capitalistic production. It can be characterised as M-C-M', that is it begins with money (M) to produce commodities for sale (C) for more money (M'). For that reason, GDP is an appropriate measure for the capitalist sphere of production as it focuses on production for sale in exchange for money.

Still, it is not perfect even for that narrow purpose. For example, GDP does include imputed monetary values for some production that is not actually sold. The most important example is the 'services' of owner-occupied housing. The idea is that the homeowner 'consumes' housing services over the course of the period. If the home is not owned, we can use instead the rent paid as a value of the housing services consumed by the renter. The problem is that many families live in homes they purchased so there are no market transactions that takes place over the period.

Note that when a new home is purchased, that is counted as residential investment (included in the investment category, not the consumption category, see the next section). It would not make sense to count the **entire** market value of a home as consumption over the period. Further, most homeowners have purchased a 'used' home, so that purchase will not show up in either the investment category or the consumption category. For that reason, the imputed monetary value of the housing services over the period is counted as consumption – whether or not the home is new or used. Still, by including imputed values, our measure of GDP deviates from the ideal of capturing the total value of production that is sold at market prices over the period.

## 4.2 Components of GDP

The National Income and Product Accounts (NIPA) divide the nation's output into four main categories, and add a fifth to account for foreign production that is available to the nation's residents. These are: Consumption, Investment, Government Expenditure, Exports and Imports. Each of these can be further subdivided.

### Consumption (*C*)

This includes domestic consumption by **households** of goods and services. Keep in mind from our definition of GDP that only currently produced final goods and services are included. Intermediate goods and services are excluded, as are sales of used goods.

Generally speaking, all current period spending on new goods and services by households is included as consumption. The only major exceptions are the purchase of a newly built house, which is included as investment spending (see below), and the inclusion of 'imputed' housing services of owner-occupied homes, which is counted as consumption.

What is most confusing for students is that household 'investment' in shares and bonds is not included in GDP at all. This is because shares and bonds are not currently produced goods and services. Indeed, purchase of financial assets of any type is treated by the NIPA System as saving, not as spending.

### Investment (*I*)

This includes three main categories: capital investment by firms, inventory investment by firms, and real estate investment by households. Investment expenditure increases the productive capacity of the economy and expands what we think of as **potential GDP**. So it adds to current spending but increases the capacity of the economy to absorb increases in future spending without inflation.

Capital investment includes spending on plant and equipment - factories and machines, for example. Increasingly, investment includes purchases of software and other non-physical but long-lasting inputs to production.

As discussed, we do not want to include intermediate goods in GDP, so purchases by firms of inputs that are 'used up' in the production process are not included as investment - inputs such as electricity, oil and other natural resources, marketing services, and so on. Note that the precise division between an intermediate input and an investment is necessarily somewhat arbitrary, and so will rely on accounting conventions and will be related to the input's useful life.

Again, purchases of financial assets are not included as investment. For example, if one firm takes over another, that is not an investment for purposes of measuring GDP. Also note that if a household buys a car it is counted as consumption, but if a business buys a car it is counted as investment - even if the firm operates out of a home office of the same household!

The value of unsold goods is defined as inventory investment. An increase of inventories is also treated as an investment, even if the firm did not plan to increase its inventories. For example, a firm might have produced output that it was not able to sell by the end of the accounting period. If a firm sells more output than planned, its inventories are reduced. This is treated as negative

investment. Swings of inventory investment can be quite wide as it is difficult for firms to sell precisely the amount that they planned.

Finally, real estate investment includes new construction of residential and non-residential buildings. Sales of existing homes as well as existing commercial buildings are not included as investment. Sales of land also would not be counted as investment.

When in doubt whether the sale of an asset would be counted as investment or simply a purchase of an asset, a useful rule of thumb is to consider whether labour was used during the period to produce the asset. If it was, then this is investment; if not, then it is simply an asset purchase - which is treated as a portfolio adjustment, but not an investment. Newly produced machines, factories, houses, and apartment buildings all required current labour services to produce them, and hence, count as investment. Sales of stocks, bonds, existing houses, or existing factories do not use labour - at least in the current period - to produce them, so they are not investment.

### **Government spending ( $G$ )**

This includes government purchases of final goods and services.

Note that it does not include government transfer payments, such as spending on welfare and social security. This is because if we were to include transfers we would double count since most transfer payments will be spent on consumption goods and services, hence, included in 'C' as described above. Government transfer payments are not purchases of currently produced goods and services, so are not part of GDP.

Government purchases can be further divided between 'consumption' and 'investment' or capital expenditures. The division between these two subcategories is somewhat arbitrary. Government consumption expenditures are for goods and services that are used relatively quickly (fire-fighting services, postal delivery, and air traffic control), while government investment purchases are for long-lasting improvements (fire trucks, roads, and airports). Typically, any spending whose impacts are exhausted within a 12-month period are considered to be consumption, otherwise, they are classified as investment. Do not get confused by the use of the terms 'consumption' and 'investment' when applied to the division of government spending by type - these are under the 'G' category and not under the 'C' or 'I' categories discussed above.

### **Exports ( $X$ ) minus imports ( $M$ ) or Net Exports ( $NX$ )**

Exports are goods and services sold abroad; imports are goods and services produced abroad for domestic use. If imports are greater than exports, then net exports are negative; alternatively, if imports are less than exports, then net exports are positive. Again, these can be consumption-type goods or investment-type goods but if they are sold abroad or bought from abroad they are counted in the  $NX$  category, but not in the  $C$  or  $I$  category.

Exports add to domestic spending to stimulate production, whereas imports represent a drain on domestic spending.

### 4.3 Equivalence of Three Measures of GDP

GDP can be measured in three ways, namely the expenditure approach, the production approach and the income approach, and which, subject to the statistical discrepancy, should be equal.

The expenditure approach is conceptually the simplest because it works on the principle that total expenditures denote the value of the product that been bought, and given the inclusion of inventory investment in the definition of investment, it measures the value of total production. The production (or value added) approach is based on summing the gross outputs of every class of enterprise and then netting out intermediate consumption. The income approach works on the principle that the incomes of the productive factors (producers) must be equal to the value of their product, and determines GDP by finding the sum of all producers' incomes.

#### Expenditure approach

The first way to estimate GDP is to calculate the sum of final expenditures on goods and services measured in current market prices. As we discussed above, GDP ( $Y$ ) is the sum of consumption ( $C$ ), investment ( $I$ ), government spending ( $G$ ) and net exports ( $X - M$ ).

$$(4.1) \quad Y = C + I + G + (X - M)$$

#### Production approach

This approach measures gross value added. First it is necessary to measure the gross value of domestic output over say a year. This will include the value of output at all stages of production (see example below<sup>4</sup>). This will include intermediate consumption that are the costs of (raw) materials, supplies and services which were used up in the production of gross output. We then subtract the intermediate consumption from the gross value of domestic output to obtain the gross value added. If we do not subtract the intermediate consumption, then we are double counting.

Consider a three stage production process which culminates in the final sale of woollen coats to consumers. Initially sheep farmers incur costs of feed etc in rearing the sheep and pay wages to the shepherds and to the sheep shearers and then sell the wool to a woollen mill, which processes it by the employment of labour and other producers. The woollen mills then sell the processed wool to the manufacturer of the coats, which employs labour and other producers in the production of the woollen coats. For simplicity, we assume that the manufacturer sells these final goods to consumers. At each stage of the production process the value added by the producers must be calculated, so for example value added by the woollen mill is the value of sales of the processed wool minus the costs of buying the unprocessed wool and raw materials used to process the wool and the electricity costs incurred in the production process. Then we can write: value added in the production of woollen coats = gross value of output – value of intermediate consumption, which has been summed over all stages of production.

The sum of the value added across every class of enterprise is known as GDP at factor cost. GDP at factor cost plus indirect taxes less subsidies on products is GDP at producer price.

---

<sup>4</sup> If the production of all final output is vertically integrated, so a single firm is responsible for all stages of the production for each good or service, then there is no intermediate consumption.



## Income approach

The third way of measuring GDP is to calculate the sum of primary incomes distributed by resident producers of goods and services. Thus this method adds together the producers' incomes that firms pay in exchange for their services, namely wages for labour, interest for capital, rent for land and profit for capitalists. This defines GDP at factor cost. It is then necessary to add indirect taxes minus subsidies to get a measure at market prices, and in turn depreciation (or capital consumption allowance) to obtain GDP.

Under the production approach, the value added at each stage of production is the additional income, which is generated, so the equivalence of the production and income approaches to the measurement of GDP is clear.

### 4.4 GDP versus GNP

GDP is the total value of goods and services produced within a nation regardless of the ownership of the firm producing them; GNP is the total value of goods and services produced by residents of the nation regardless of the location of the production.

**GDP** includes earnings from production in the domestic economy that goes to foreigners.

**GNP** does not, but includes foreign earnings of domestic firms and residents operating abroad. Thus the financial flows between the domestic and external sectors are not confined to net exports.

Until the early 1990s the USA tended to use GNP while many nations used GDP. However, since then the USA has conformed and adopted GDP although it still reports GNP. For the USA, there is no major difference between GDP and GNP because earning from production in the USA that go to foreigners is nearly balanced against foreign earnings of US residents. For many other nations, however, there is a large difference between GDP and GNP because, for example, their residents have large investments in factories operating abroad.

In the Appendix we show how net national income is measured (Advanced Material – Chain weighted real GDP).

### 4.5 Measuring Gross National Income

We initially examine **Gross National Income** (GNI) from the perspective of what can be done with income: an individual can consume it, pay taxes, or save it. As a simplification we ignore the difference between GNP and GDP, so we can write:

$$(4.2) \quad Y = C + S + T = GDP = C + I + G + NX$$

We use  $Y$  to stand for income;  $C$  is consumption,  $S$  is gross saving and  $T$  is total taxes paid. We can think of  $S$  as a residual: it is all after-tax, (disposable) income that is not spent on consumption.

Another way of looking at GNI is from the point of view of which sector is receiving income: wages ( $W$ , which go to workers), profits ( $P$ , which goes to capitalists; note this is a gross concept which includes interest income) or taxes ( $T$ , revenue to government).

In that case we can write:

$$(4.3) \quad Y = W + P + T = GDP = C + I + G + NX$$

Of course that also means that:

$$(4.4) \quad Y = C + S + T = W + P + T = GDP = C + I + G + NX$$

We can easily manipulate the identity to obtain a useful identity:

$$\text{Keynesian Saving Equation: } S = I + (G - T) + NX$$

What is G-T? It is government deficit spending?

We'll make more use of identity (4.4) later.

## 4.6 GDP Growth and The Price Deflator

We have defined nominal GDP as a measure of the value of output at current market prices. We often want to measure economic growth, as measured by the growth of GDP over time. The problem is that prices, as well as output, change over time. If we find that GDP (nominal) today is 100 times greater than it was a hundred years ago, does that mean that we enjoy 100 times more physical output? Clearly not, if prices have risen. To take account of this, we often want to 'deflate' GDP, that is, to correct our measure for the change in prices to get an idea of 'real' economic growth.

The idea is simple, but in practice this is a very difficult thing to do. Let us start with the conceptual problem.

Suppose we want to compare GDP of 2015 to GDP of 2002 to see how much 'real' output grew over the thirteen-year period. To find nominal GDP in each year we take the 'current' market price of that year and multiply by the quantity produced that year. For exposition purposes, we are simplifying here by taking the quantity and price of a single aggregate good we call GDP:

$$(4.5a) \quad GDP_{2002} = P_{2002} * Q_{2002}$$

$$(4.5b) \quad GDP_{2015} = P_{2015} * Q_{2015}$$

where  $GDP_t$  measures GDP at current prices in year  $t$ , based on production level ( $Q_t$ ) and market price ( $P_t$ ).

However we are interested in a comparison of levels of 'real' GDP over time that corrects our measure for the change in prices. In that case, we have to decide which year's prices to use as a 'base'. We always calculate 'real GDP' over time in terms of a base year. We could choose 2002 or 2015 or any other year as the base. Let us say we choose to use the prices of 1985 (this makes it clear that we do not have to use prices of 2002 or of 2015).

Then we do the following calculation:

$$(4.6a) \quad RGDP_{2002} = P_{1985} * Q_{2002}$$

$$(4.6b) \quad RGDP_{2015} = P_{1985} * Q_{2015}$$

where  $RGDP_t$  denotes real GDP in year  $t$  based on 1985 prices.

So long as we have used the same base year to calculate real GDP for both years, we can determine real GDP growth over the thirteen-year period, but the measure will reflect to some

degree the choice of the base year prices, when we consider many goods rather than a single good.

In practice, statisticians update the base year through time so that they will always use a fairly recent base year. Thus you would be unlikely to use 1900 as the base year to calculate real GDP for 2015! The older the base year used for calculations, the greater the problems encountered in calculating real GDP. We will return to these problems shortly. Before we do, there are two other useful concepts related to calculation of real GDP.

First there is the GDP deflator, which is an indicator of price changes. It is defined in year  $t$  as follows:

$$(4.7) \quad GDPD_t = GDP_t / RGDP_t$$

where  $GDPD_t$  denotes the GDP deflator for year  $t$ .

Changes in the magnitude of the GDP deflator over time give us a measure of prices changes for output as a whole. Note that it is possible for prices in general to go down as well as up. However, over the past century deflations have been relatively rare.

Our goal has been to develop a method for adjusting GDP for price changes. In practice it is much more difficult than suggested by the earlier discussion. As noted, we were using a simplification to calculate nominal GDP as ‘Price times Quantity’ of a single good.

However, GDP is defined as the value of total output measured at current prices. Conceptually we have a set (vector) of prices (one for each good or service sold) and a set (vector) of quantities (an entry for every item sold), and then we sum each individual sale ( $P^i * Q^i$  for the  $i$ th item) to obtain GDP. That does not seem too difficult—we simply recognise that output is heterogeneous and so it can only be aggregated in nominal terms, not in ‘quantity’ terms.

In practice, major problems are created if we try to measure the value of real GDP in terms of another year’s prices. Let us say we again use 1985 as our base year, and apply 1985 prices to the goods and services sold in 2015. How do we put a 1985 price on an iPad sold in 2015? There were no iPads sold in 1985 and indeed nothing comparable existed.

To reverse the problem, how can we find a 1985 price for manual typewriters sold in 1900 to value real GDP that year (in terms of 1985 prices)? Clearly, the composition of output changes both in terms of what is sold and the quality of items sold (the typical personal computer sold today is very much faster than one sold in 1990 even though the nominal price has hardly changed). It should be obvious that the older the base year chosen, the more difficult the problem. That is why statisticians have favoured the use of a chain-weighted measure of GDP since it involves a lag of only a year. In the Appendix we discuss this measure in more detail (Advanced Material – Measuring net national income).

## 4.7 Measuring CPI inflation

### The CPI index

In this section we look at the measurement of the prices of consumer goods (bought by households) and make brief reference to producer goods (bought by firms, including raw materials and intermediate goods to be used in production). These prices could go down, but the usual trend is for rising prices.

The index most commonly used to calculate inflation of consumer goods prices is the Consumer Price Index, or CPI. It is defined as follows:

**CPI:** An index based on the cost of a fixed basket of consumer goods and services.

In the construction of the CPI index, the statistician needs to decide what consumer goods and services to include, their respective quantities (weights) and how to calculate the corresponding prices. It is assumed that the chosen basket of goods and services is representative of the purchases made by a typical household. The statistician chooses a base year (much like the choice of the base year to be used in calculating real GDP). The CPI then represents the cost of a market basket of consumer goods and services.

The measure is usually expressed for a specific spatial area such as a capital city or a weighted-average of all capital cities in a nation.

The items included in the Australian CPI published by the Australian Bureau of Statistics are shown in Table 4.1. Within each major group there are many items included.

**Table 4.1 Items in Australian CPI, September 2013**

---

#### All groups CPI

Food and non-alcoholic beverages

Alcohol and tobacco

Clothing and footwear

Housing

Furnishings, household equipment and services

Health

Transport

Communication

Recreation and culture

Education

Insurance and financial services

---

Source: Australian Bureau of Statistics, Consumer Price Index, <http://www.abs.gov.au/AUSSTATS/abs@.nsf/mf/6401.0>

If the prices of all the items in the basket changed at the same rate from one period to the next, then the change in the cost of the basket would be easy to calculate, period by period. But, in reality, the individual prices generally change at different rates, so that relative prices are also changing. The statistician thus needs a single summary measure to determine whether the basket overall is rising in cost or not. That is the role that the price index plays. It is a weighted-average of the price movements in the given basket relative to some base period.

In compiling a summary measure such as the CPI, the statistician has various options. Two broad options are whether to use base-weighting or current-weighting to compile the index.

A base-weighted index examines the shifts in prices of the basket of goods and services using the base-period quantities purchased and is referred to as a *Laspeyres* index after the German economist who first compiled such measures. The base-weighted index allows us to see how much a basket that consumers bought in the base period would cost in the current period.

A current-weighted index uses the current quantity purchased of each good and service in the basket as the weight to compile the average measure. This is commonly called a *Paasche* index after the German statistician who developed this measure.

The current-weighted index allows us to see how much a basket that consumers buy in the current period would have cost in the base period.

Both measures provide different ways of estimating the change in the cost of the basket of goods and services over time. However, statisticians tend to favour the use of the *Laspeyres* index to calculate the CPI because it requires less information to be available. The only new information that is required is the current prices of the items in the basket. The quantities making up the basket and the corresponding base year prices are already known.

This allows for a more timely publication of the CPI, which is a central policy variable used by central bankers and treasuries in formulating monetary and fiscal policy, not to mention, its use in labour and other contracts and indexing the values of different transfers, such as some pensions and other benefits.

To simplify our analysis, imagine a basket of goods and services comprises two items (Bread and Cheese). Yes, the obvious question is don't these people wear clothes?!

Table 4.2 shows the hypothetical data we will be working with to illustrate the construction of the price index.

In Year 1, the price per unit of Cheese is \$4 and 3 units are consumed overall. So total expenditure on Cheese in Year 1 is \$12. The price of a loaf of Bread is \$2 and 9 units are consumed in Year 1, making total expenditure on Bread \$18. Overall, the basket of goods costs \$30 in Year 1 (Column 3).

In Year 2, Cheese rises to \$5 per unit and 4 units are consumed whereas Bread rises to \$3 per loaf and 10 units are consumed. Overall, the basket of goods in Year 2 now costs \$50 (Column 7).

Note that if we wanted to know what the quantities purchased in Year 1 would cost in Year 2, Column (8) provides that answer, \$42.

We calculated that using the following data:

Cheese  $\$5 \times 3 = \$15$

Bread  $\$3 \times 9 = \$27$

Total = \$42

The 3 and the 9 are the quantities of each good purchased in Year 1. Conversely, Column (7) shows the expenditure in Year 2 based on Year 2 prices and Year 2 purchases.

Similarly, if we wanted to know what the basket would cost in Year 1 based on Year 1 prices and Year 2 purchases we would look to Column (4).

**Table 4.2 Hypothetical data for basket of goods and services**

	Price Per Unit	Quantity	Expenditure	Expenditure based on Year 2 Quantities
	(1)	(2)	(3)	(4)
	\$	Units	\$	\$
<b>Year 1</b>				
Cheese	4	3	12	16
Bread	2	9	18	20
Total			<b>30</b>	<b>36</b>
<b>Year 2</b>				
Cheese	5	4	20	15
Bread	3	10	30	27
Totals			<b>50</b>	<b>42</b>

What would be the price index values in this example?

*Base-weighted CPI*

Using base-weights (Year 1 quantities), we will set the index in Year 1 to 100, so we express the index as follows:

$CPI_{Year 1} = 100 \times \text{Total Expenditure in Year 1 (Column 3)} \div \text{Total Expenditure in Year 1 (Column 3)}$

$$CPI_{Year 1} = (100 \times \$30) / \$30 = 100$$

In Year 2, the index would be (using Year 1 weights):

$CPI_{Year 2} = 100 \times \text{Total Expenditure in Year 2 (Column 8)} \div \text{Total Expenditure in Year 1 (Column 3)}$

$$CPI_{Year 2} = (100 \times \$42) / \$30 = 140$$

### *Current-weighted CPI*

Using current-weights (Year 2 quantities), the index in Year 1, which we will again set to 100 would be expressed as:

$CPI_{Year\ 1} = 100 \times \text{Total Expenditure in Year 1 (Column 4)} \div \text{Total Expenditure in Year 1 (Column 4)}$

$$CPI_{Year\ 1} = (100 \times \$36) / \$36 = 100$$

In Year 2, the index would be (using Year 1 weights):

$CPI_{Year\ 2} = 100 \times \text{Total Expenditure in Year 2 (Column 7)} \div \text{Total Expenditure in Year 1 (Column 4)}$

$$CPI_{Year\ 2} = (100 \times \$50) / \$36 = 138.9$$

### **Rate of growth of the CPI index**

We have generated two CPI indexes (one base-weighted and one current-weighted) over two years, so we can calculate a measure of the overall movement in prices, and provide a measure of the change in the cost-of-living. The growth rate of the CPI measures the rate of inflation (if positive) or deflation (if negative), acknowledging that strictly inflation (deflation) is an on-going, rather than one off, increase (decrease) in the price level.

We can write the percentage rate of inflation (deflation) as:

$$(4.8) \quad CPI_{Gt} = 100 * [(CPI_t - CPI_{t-1}) / CPI_{t-1}]$$

where  $CPI_t$  denotes the index magnitude in period (say, year)  $t$  and  $CPI_{Gt}$  denotes the growth of the CPI from period  $t-1$  to period  $t$ . So a rate of change (growth) can be expressed as one hundred multiplied by the change in the index, divided by the initial value of the index.

It can be readily shown that the respective rates of change for the Base and Current Weighted Price Indexes between Year 1 and Year 2 are 40 per cent and 38.9 per cent.

You will appreciate that the current-weighted index takes into account changes in prices and the quantities purchased following these price changes, whereas the base-weighted approach considers price changes only and ignores the fact that people will change their expenditure patterns over time as relative prices change.

In practice, household expenditure patterns change and new goods and services are sold, so statisticians periodically revise the weights in the basket of goods and services in line with other information that they collect. They have complex methods to splice the new and the old indexes together. In the Appendix, we explore the biases associated with using the CPI to accurately measure inflation (Advanced Material – Difficulties in using the CPI to accurately measure inflation).

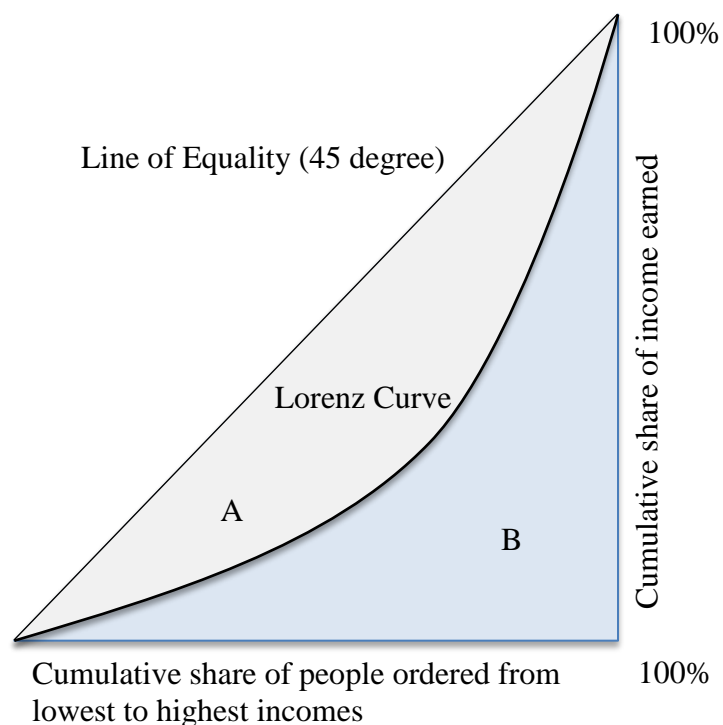
Finally it should be recognised that there are other published price indexes, including those based on wholesale and retail prices. For example, the US Producer's Price Index is based on the wholesale prices of approximately 3000 items, including raw materials and semi-finished goods.

## 4.8 Measuring National Inequality

As discussed above, our measures of national output (GDP) and income (GNI) do not directly take account of the distribution of output and income. Economists typically use the **Gini Coefficient** derived from a **Lorenz Curve** as an index of income distribution.

The Lorenz Curve plots the share of total income received (vertical axis) by the lowest X per cent of income earners (horizontal axis) (see Figure 4.1). It is easy to see that in our example the distribution is not equal because as we move from the origin at the left end of the horizontal axis, the share of income going to those with the lowest income initially increases slowly. As we move to the higher income people, the cumulative share of income increases more rapidly. The 45-degree line shows the case of perfect equality, so that 30 per cent of people have 30 per cent of total income; 60 per cent of people have 60 per cent of total income and so on.

**Figure 4.1** The Lorenz curve



We can calculate the **Gini Coefficient** as a ratio using the two areas, A and B in Figure 4.1:

$$(4.9) \quad \text{Gini Coefficient} = A/(A+B)$$

Different shaped Lorenz curves can generate the same value of the Gini Coefficient. In addition, there are different ways to measure income, for example before or after taxes, and before or after income transfers. Statisticians have developed a number of algebraic formulations of the Gini, but each yields the same value of the Gini Coefficient for a given dataset. There are also alternative indexes to the Gini Coefficient. It is important to realise that different indexes exhibit different properties and the choice of which index to use should be made in light of the objectives associated with measuring inequality.



A Gini Coefficient of zero means that income is perfectly equally distributed as the economy is lying on the Line of Equality. Alternatively, a Gini coefficient of one means that income is perfectly unequally distributed (that is, one person has all the income).

Table 4.3 shows the Gini Coefficients for all the nations that belong to the Organisation for Economic Co-operation and Development (OECD), for which comparable data is available for the years 2004 and 2012. The Gini Coefficients mostly range between the values of 0.25 to 0.50. There is considerable diversity among these nations with respect to income inequality. Sweden had the least inequality in 2004, while Mexico had a large degree of inequality.

Note also that inequality has increased in many nations between 2004 and 2012, while it declined in other nations (see the + and – signs).

**Table 4.3 Gini coefficients for several OECD nations, 2004 and 2012**

<b>Country</b>	<b>2004</b>	<b>2012</b>	<b>Change</b>
Australia	0.315	0.324	+
Austria	0.269	0.276	+
Belgium	0.287	0.262	-
Czech Republic	0.269	0.252	-
Estonia	0.346	0.326	-
Finland	0.267	0.261	-
France	0.283	0.306	+
Germany	0.285	0.289	+
Greece	0.336	0.340	+
Iceland	0.262	0.252	-
Ireland	0.323	0.302	-
Italy	0.331	0.326	-
Luxembourg	0.263	0.299	+
Mexico	0.474	0.482	+
Norway	0.276	0.253	-
Poland	0.381	0.300	-
Portugal	0.383	0.341	-
Slovak Republic	0.266	0.249	-
Slovenia	0.247	0.251	+
Spain	0.332	0.335	+
Sweden	0.234	0.274	+
United Kingdom	0.331	0.351	+
United States	0.360	0.389	+

Source: OECD Statistics.

## Appendix: Advanced Material

### Chain weighted real GDP

In recent years, there has been growing use of another measure to take account of price changes, called the Chain Weighted real GDP, defined as follows:

$$RGDP_{2015} = \{(P_{2014} + P_{2015})/2\} * Q_{2015}$$

This measure averages the prices of two years and, as we discuss below, this is particularly useful for measuring real GDP growth.

In practice, economists are more interested in real GDP growth rather than in levels of real GDP. This favours the chain weighted measure even more over the calculation of real GDP with a base year that is periodically changed. Every time the base year is changed, real GDP needs to be recalculated for every year. That, in turn, will change the calculations of real GDP growth rates over time. In an important sense economic history is 'rewritten' every time the base year is changed.

With the chain weighted approach, however, the calculation of real GDP growth is invariant to changes of the base year. Changing the base year will change the calculated levels of real GDP but not the growth rate for the historical series of real GDP that will instead use the chain weighted measure.

Changes in this measure are calculated using the weights of adjacent years. These annual changes are 'chained' (multiplied) together to form a time series that allows for the effects of changes in relative prices and in the composition of output over time. Thus, BEA is able to calculate an index that uses weights appropriate for each period and thereby avoids the rewriting of economic history that results from updating the base period of a fixed-weighted index as well as the substitution bias that is inherent in fixed-weighted indexes (Landerfeld and Parker, 1997: pp 59-60).

In other words, once the BEA has calculated real GDP growth for any set of years using the chain weighted approach, it will not need to do any recalculations because the base year prices used for that set of years will not change. This is still more difficult than it sounds, but, we will not go into further details here.

### Measuring net national income

At the aggregate level, national income equals national output because as discussed previously, production of output generates equivalent income. We will define Net National Income as NI, and then will derive a number of subcategories of income.

It is more convenient to begin with GNP (so it includes foreign earnings of domestic residents). GNP equals Gross National Income, or GNI. To calculate *NI* we need to subtract two items.

Over the course of a production period (month, quarter, or year) some of the production facilities (plant and equipment) 'wears out' or 'depreciates'. We subtract depreciation from our Gross National Product to obtain Net National Product (NNP).

We then subtract indirect business taxes (sales and excise taxes) to obtain Net National Income (*NI*). The reason for deducting depreciation and these taxes is to obtain a measure of national income that is actually available to purchase national output. We subtract the depreciation

because producers must set aside a portion of gross income to replace the capital that is wearing out; we subtract indirect business taxes because these reduce the amount of income that can be paid out of production.

To summarise, begin with GNP

Subtract Depreciation = NNP

and subtract Indirect Taxes (sales and excise taxes) = NI

Next we want to obtain a measure of Personal Income flowing to households. We subtract corporate taxes, payroll taxes, and undistributed profits since the taxes go to government and undistributed profits are retained by producers, leaving us with the income to be paid out to households.

However we need to add transfer payments made by government to households as well as personal interest income received by households to obtain Personal Income (*PI*). To summarise these operations:

To obtain Personal Income, PI we:

Subtract corporate taxes and undistributed profits and payroll taxes

Add transfer payments and personal interest income

Equals PI

We need to get a measure of PI after taxes paid by individuals, so we subtract personal taxes to obtain Personal Disposable Income: PDI. This is the after-tax income available to individuals to spend, PDI.

Subtracting Personal Consumption, Interest Paid to Business and Transfer Payments made to Foreigners from PDI gives us Personal Saving (*PS*).

We start from Personal Disposable Income

less Personal Consumption

less interest paid to business

less personal transfer payments to foreigners

Equals: Personal Saving.

Note that gross saving *S* (defined above) is not the same as Personal Saving, as it is based on total income (not PDI) and we have not deducted interest paid to business and transfers to foreigners (see the definition for PS above).

### **Difficulties in using the CPI to accurately measure inflation**

#### *Measurement biases*

There are many difficulties in using the CPI to get an accurate measure of inflation. For example, if consumers increase the percent of purchases at 'discount' outlets, the CPI will overstate the actual rate of inflation experienced by the typical consumer - this is the '*outlet substitution bias*'- because the index does not adequately take into account such shifts.

In addition, consumers will change the composition of the basket of consumer goods purchased over time; since the composition of the basket used to calculate the price index is completely revised on an irregular basis, and therefore will bias results. Economists identify three different kinds of bias associated with changing baskets: substitution bias, quality change bias, and new product bias. In addition there is growing recognition of a fourth kind of bias called the formula bias.

The first refers to the impact that changing relative prices would have on the composition of the basket. If, for example, the price of tea rises relatively to that of coffee, economic theory suggests that consumers will substitute coffee for tea. However, as the CPI basket might be changed only once per decade to reflect the switch to coffee, the index will be calculated as if no substitution had occurred, leading to overstatement of inflation due to a *substitution bias*.

Often when prices rise, this reflects increases of the quality of products (products might last longer or provide a higher level of services). In most cases, it is very difficult to calculate what portion of a price increase should be attributed to quality changes, and the BLS does not even attempt to calculate this for many products. Thus, inaccurate measures of quality change introduce a *quality change bias*.

Finally, new products are introduced all the time; the BLS includes these in the basket only with long and variable lags, which introduces a *new product bias* into the CPI. In the case of some goods, a considerable bias results. For example, many high-technology consumer goods follow a price cycle that begins with very high prices for goods sold to high income classes, then prices fall rapidly as the goods are introduced to lower income classes, and then prices gradually rise again as the market matures. If the BLS introduces the goods into the basket only after prices have reached their minimum, the CPI will not capture the period during which prices fell rapidly, but will include the mature period in which prices rise. In recent years, this bias should be expected to be quite important.

There is yet another source of bias called the '*formula bias*'. This bias results because price data are collected on a disaggregated basis and then aggregated in a very complex manner that can introduce anomalies. For example, the calculation method used in recent years gives too much weight to items on sale; somewhat paradoxically, this generates formula-induced inflation as the items go off-sale. The degree of this bias can increase with the frequency of rotation (of outlets included in the sample) because the bias results from short-run price variability and a method that gives greater weight to lower-than average prices.

Researchers had noticed that surveys of average prices actually paid by consumers showed rates of inflation well below the rates of inflation reported by the CPI for relatively disaggregated components of the consumer basket; while part of this could be attributed to the outlet substitution bias, most of it could not. Estimates of the formula bias run as high as six-tenths of a percentage point for owner-occupied housing and one percentage point for apparel, an item often on sale.

#### *The housing component*

The housing component of the CPI is very large—in the US it is above 40% of the index and during high inflation periods it contributes up to half of the measured inflation. There are two alternative ways to calculate the contributions of the housing sector to a price index: the flow of services approach versus the homeowner or 'user' cost approach. The method currently used in the US —imputed rental cost— is based on the flow of services approach and has been in place

in the US since 1983. Previously, the BLS tried to calculate user cost of housing, but it was believed that this method mixed investment and consumption features of home ownership.

The largest portion of the housing component is ‘shelter’ services, which account for more than two-thirds of the housing sector’s relative importance; of this, nearly three-quarters is owner-occupied costs (the rest is renters’ costs—since most Americans own their own homes, this is not surprising). The majority of homeowners’ costs are the owners’ equivalent rent—which is an estimate of the amount that homeowners would pay to rent an equivalent home.

The BLS uses a survey of rental units to obtain data regarding changes of rental price. The results are adjusted through a weighted averaging process, and quality adjustments are made to deal with aging and improvements. The imputation of renter’s costs to be included in the CPI is therefore fairly straightforward.

However, the method used for owners’ equivalent rent (OER) is more complicated. Field agents ask owners for the rental price the homeowner believes the house could rent. Agents may enter their own estimate if they believe the owner’s estimate is unreasonable. These survey data are used to establish the base year imputed rent. Subsequent values of implicit rent for a given unit are derived by using changes in market rents that occur in a specific subsample of the residential rent units used for the residential rent unit. In other words, the rate of increase of OER is obtained by applying the rate of increase of prices of rental units that are thought to be similar in certain respects (location, structure type, and quality).

There are situations in which this method of calculating ‘inflation’ of the housing component could lead to erroneous results (for example, where the statisticians impute high inflation when actual housing prices are falling). We need not go into that here. What we want instead to make clear is that construction of an index is difficult, subject to controversial decisions, and to error.

Further, it is important to understand that the CPI is comprised of components that have ‘imputed prices’—prices formulated by statisticians rather than obtained from markets. This is because US statisticians want to obtain a measure of the cost of a relatively complete consumer basket that includes items that are not bought annually (such as housing ‘services’ enjoyed by those who own their own homes).

There is a trade-off made between calculating a CPI that takes a ‘hedonic’ approach (tries to put a price on the ‘enjoyment’ one receives from the entire consumption basket) and one that attempt to focus on what is actually happening to market prices of things that are bought. The problem with the first is that statisticians must make a lot of guesstimates. The problem with the second is that it does not deal adequately with quality adjustments (such as improvements to computers so that \$1000 today buys a much better computer than \$1000 bought a decade ago).

What all this means to the student is that you should take CPI measures of the inflation rate with the proverbial grain of salt! Especially at low measured rates of inflation, we cannot be sure if the prices of things people actually buy are rising, steady, or falling.

## References

The Economist (2014) 'Sex, drugs and GDP: Italy's inclusion of illicit activities in its figures excites much interest', 31 May.

Landefeld, J.S. and Parker, R.P. (1997) "BEA's Chain Indexes, Time Series, and Measures of Long-Term Economic Growth" *Surveys of Current Business*, June, 58-68.

# Chapter 5: Sectoral Accounting and the Flow of Funds

## Chapter Outline

- 5.1 Introduction
  - 5.2 The Sectoral Balances View of the National Accounts
    - How can we use the sectoral balances framework?
  - 5.3 Revisiting Stocks and Flows
    - Flows
    - Stocks
    - Inside wealth versus outside wealth
    - Non-financial wealth (real assets)
  - 5.4 Integrating NIPA, Stocks, Flows and the Flow of Funds Accounts
    - Causal relationships
    - Deficits create financial wealth
  - 5.5 Balance Sheets
  - 5.6 The Flow of Funds Matrix
    - Flow of funds accounts and the national accounts
- Appendix
- A graphical framework for understanding the sectoral balances

## Learning Objectives

1. Understand the relationship between sectoral balances and changes to net financial assets.
2. Recognise the distinction between vertical and horizontal transactions in their impact on net financial assets.
3. Interpret a balance sheet (stocks) and period to period changes of its items (flows).

## 5.1 Introduction

In Chapter 4, we saw that the national accounts divided the national economy into different expenditure categories – consumption by persons/households (note housing is investment); investment by private business firms; spending by the government; exports to and imports from the foreign sector.

The most basic macroeconomics rule is that one person's spending is another person's income. Another way of stating this rule is that the use of income by one person (i.e. spending) will become the source of income for another person or persons.

In this Chapter, we extend our understanding of the national accounts, which record these different flows of expenditure and income. The **sectoral balances perspective** of the national accounts brings the uses and sources of national income together. We show that when appropriately defined, the sectoral balances must sum to zero. We expand our discussion of stocks and flows and then introduce the flow of funds by reference to the sectoral balances.

The sectoral balances approach helps us to understand the relations among the spending and income balances of the households, firms, government, and foreign sectors of the economy. For example, they allow us to conclude that it is impossible for all sectors to run surpluses (that is, to 'save overall' – spend less than their income) simultaneously. For one sector to run a surplus, we need at least another to run a deficit (spend more than their income). You will learn that for those nations, which run external deficits against the rest of the world, then, in order for households and firms together (that is, the private domestic sector) to run surpluses (spending less than income in order to save overall) it is necessary for the government to run fiscal deficits (spend more than taxes). There are many useful insights that can be gained from an understanding of a nation's sectoral balances.

## 5.2 The Sectoral Balances View of the National Accounts

The Australian Bureau of Statistics publication – *Australian System of National Accounts: Concepts, Sources and Methods, 2014* – provides an excellent source for understanding the background concepts that are used to derive the sectoral balances framework. The discussion is generally applicable to all countries.

From this framework, economists derived what is called the basic **income-expenditure model** in macroeconomics to explain the theory of income determination that forms the core of the so-called **Keynesian** approach (see Chapter 7).

The income-expenditure model is a combination of accounting identities drawn from the national accounting framework and behavioural theories about how flows of expenditure by households, firms, governments, and foreigners combine to generate sales, which in turn, motivate output and income generation.

Remember, that an expenditure **flow** is measured as a certain quantity of dollars that is spent per unit of time. So for example, in the June-quarter 2015, the Australian Bureau of Statistics estimated that household consumption in Australia was \$A220,913 million in real, seasonally-adjusted terms.

Conversely, a **stock** is measured at a point in time and is the product of prior, relevant flows. For example, the Australian Bureau of Statistics estimated that total employment in Australia in



October 2015 was 11,838.2 thousand. The flows that generated this stock of employment were all the movements of workers between the different labour force categories: employment, unemployment, and not in the labour force. Of course, most workers remained in the same labour force category as they were in September 2015.

The accounting aspects that underpin the income-expenditure model draw on different ways of thinking about the national accounts.

We can view the national accounts in several ways. First, from the perspective of the **sources** of national income, we can write out the sources of spending that flow into the economy over a given period, using the following shorthand.

$$(5.1) \quad GDP \equiv C + I + G + (X - M)$$

that is, total national income (GDP) is the sum of total final consumption expenditure ( $C$ ), total private investment ( $I$ ), total government expenditure ( $G$ ) and net exports ( $X - M$ ). Note the use of the mathematical symbol  $\equiv$ , which denotes an **Identity** that is true by definition.

At this stage we simply take these flows of expenditure as given and understand them to be parts of the national accounts of a nation.

When these components of spending are summed, they equal **aggregate demand for goods and services** in a particular period. Aggregate demand, in turn, generates a response by producers (private and public) in the form of production, which, in turn, generates flows of income to suppliers of inputs into production (wages, profits). The sum of those flows equals national income.

As we noted in Chapter 4, the trade account is only one aspect of the financial flows between the domestic economy and the external sector. We must include net external income flows (FNI), which arise from the dividend and income flows that accrue to investments that residents make abroad minus the dividend and interest flows that are paid to foreign investors who have interests within the nation.

Adding in the net external income flows (FNI) to Equation (5.1) for GDP we get the familiar definition of gross national product or gross national income measure (GNP):

$$(5.2) \quad GNP \equiv C + I + G + (X - M) + FNI$$

At this stage, we could make the analysis quite complicated by considering retained earnings in corporations and the like, but here we assume that all income generated by firms and corporations ultimately is received by households.

To obtain the sectoral balances form of the identity, we subtract total taxes net of transfers ( $T$ ) from both sides of Equation (5.2), using the rules that govern the manipulation of equations, as outlined in the *Methods, Tools and Techniques Appendix*.

We thus obtain:

$$(5.3) \quad GNP - T \equiv C + I + G + (X - M) + FNI - T$$

Now we can collect the terms by arranging them according to the three sectoral balances:

$$(5.4) \quad (GNP - C - T) - I \equiv (G - T) + (X - M + FNI)$$

The terms in Equation (5.4) are relatively easy to understand now. The term  $(GNP - C - T)$  represents total income less the amount consumed by households less the amount paid by households to government in taxes net of transfers. Thus, it represents household saving.

The left-hand side of Equation (5.4),  $(GNP - C - T) - I$ , thus is the **overall saving** of the **private domestic sector**, which is distinct from total household saving denoted by the term  $(GNP - C - T)$ .

In other words, the left-hand side of Equation (5.4) is the **private domestic financial balance**. If it is positive then the sector is spending less than its total income (so the sector is adding to its stock of net financial assets) and if it is negative the sector is spending more than its total income.

The term  $(G - T)$  is the **government financial balance** or primary fiscal balance and is in deficit if government spending ( $G$ ) is greater than government tax revenue ( $T$ ), and in surplus if the balance is negative.

Finally, the other right-hand side term  $(X - M + FNI)$  is the **external financial balance**, commonly known as the Current Account Balance (CAB). It is in surplus if positive and deficit if negative. It is the balance between the spending/income flows of foreigners in the nation and the spending/income flows by residents that go to foreign nations.

We can say that:

**The private domestic financial balance equals the sum of the government financial balance plus the current account balance.**

This is an accounting statement.

Note that by re-arranging Equation (5.4) we get another version of the sectoral balances equation:

$$(5.5) \quad (S - I) + (T - G) - CAB \equiv 0$$

which shows that, when suitably defined, the balances sum to zero.

For example, let us assume that the external or foreign balance equals zero. Let us further assume that the private domestic sector's income is \$100 billion while its spending is equal to \$90 billion, which delivers an overall surplus of \$10 billion over the year. Then, from the identity, Equation (5.5), the government sector's fiscal deficit for the year is equal to \$10 billion. We know that the private domestic sector will accumulate \$10 billion of net financial wealth during the year, consisting of \$10 billion of domestic government sector liabilities (given that the external balance is zero).

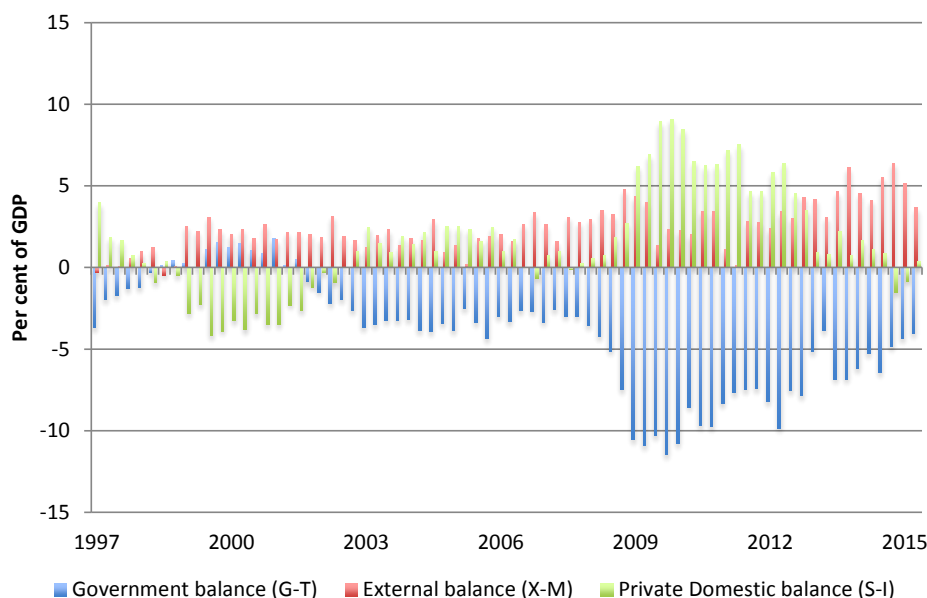
As another example, assume that the foreign sector spends less in the nation in question relative to the income it receives from that nation, which generates a current account deficit of \$20 billion. At the same time, the government sector also spends less than its income, running a fiscal surplus of \$10 billion. From our accounting identity, we know that over the same period the private domestic sector must have run an overall deficit equal to \$30 billion (\$20 billion plus \$10 billion). At the same time, its net financial wealth will have fallen by \$30 billion as it sold assets and/or issued debt. Meanwhile, the government sector will have increased its net financial wealth by \$10 billion (reducing its outstanding debt or increasing its claims on the other sectors), and

the foreign sector will have reduced its net financial position by \$20 billion (also raising its outstanding debt or reducing its claims on the other sectors).

It is apparent that if one sector is going to run a surplus, at least one other sector must run a deficit. In terms of stock variables, in order for one sector to accumulate net financial wealth, at least one other sector must be in deficit. **It is impossible for all sectors to accumulate net financial wealth by running surpluses.**

How can we use the sectoral balances framework?

**Figure 5.1 UK sectoral balances, 1960 to 2014**



Source: OECD (2015) (see also Watts and Sharpe, 2016). Note: Imports (M) include net income flows in this graph.

The UK sectoral balances shown above (Figure 5.1) replicate Equation (5.5), except that the balances which sum to zero, are expressed as percentage shares of GDP.

At this stage 3 observations are appropriate:

1. Despite the contemporary rhetoric, the UK has rarely run an annual fiscal surplus. Indeed seven surpluses have been achieved since 1960.
2. Like a number of other developed economies, including the USA and Australia, current account surpluses have also been relatively rare.
3. Private sector balances have typically been in surplus. The limited occurrence of private sector deficits have been often accompanied by fiscal surpluses. The three annual fiscal surpluses between 1998 and 2000 were accompanied by current account deficits and relatively large private sector deficits (7.3 percent of GDP in 2000). The 2001 economic slowdown followed (Watts and Sharpe, 2016). Wray (1999) notes that fiscal surpluses usually have been followed by recessions in the USA. A similar pattern is evident in most advanced economies.

In Chapter 7 we will develop an understanding of how expenditure drives income generation via the principle of aggregate (effective) demand. The principle tells us that total income in the

economy per period will be exactly equal to total spending from all sources but also details the behavioural processes involved that bring that equality into line.

We will outline theories of the components of expenditure. For example, there are various theories of household consumption expenditure but all of them suggest that consumption is determined positively by changes in disposable income. The response of consumption to a change in income is called the Marginal Propensity to Consume (MPC). It is normally hypothesised that the MPC will be less than one, so that the residual of disposable income not consumed will be positive. That constitutes saving.

So the private domestic sector financial balance ( $S - I$ ) will increase, other things equal, when national income rises.

Similarly, taxation revenue (net of transfers) is considered to be a positive function of national income. So, other things equal, the government financial balance ( $G - T$ ) falls when national income rises, and vice versa. Similarly, government spending automatically increases when national income falls as a result of welfare payments rising. In this way, the government fiscal deficit (surplus) is said to operate as an automatic stabiliser, with net expenditure being higher when national income is lower and vice versa.

Imports are also considered to be a positive function of national income – so when national income increases, we simultaneously buy more locally-produced goods and more imported goods. So the external balance falls when national income rises, and vice versa, other things equal.

In turn, changes in financial balances by sector are driven by joint impact of changes in expenditure and national income flows, as outlined above.

The accounting structures that underpin the sectoral balances framework also allow us to check logic. For example, if a politician says that the government and non-government sectors should simultaneously reduce their net indebtedness (increase their net wealth) (assuming neo-liberal public debt issuance strategies) then we know that is not possible. We don't have to resort to theory to make those sorts of conclusions.

But the accounting structures do not allow us to determine the validity of a political statement that austerity measures will stimulate growth. At that point we need theory but we should still use the sectoral balances framework to draw inferences about the overall macroeconomic outcome when sectoral balances respond to the imposition of austerity.

### 5.3 Revisiting Stocks and Flows

#### Flows

In this section we re-examine the concepts of stock and flow variables, which were briefly outlined in Chapter 1, and delineate their differences, as well as the relationship between the two. This will enable us to clearly set out the necessary relationships between deficit spending and saving, and between financial deficits and debts. This Chapter will clarify these fundamental accounting relationships.

**Flow variables** are measured over time. The simplest example is personal income, which can be stated as \$10 per hour, or \$400 per week, or \$20,000 per year. The important point is that without a clear statement of the time component, any statement about a flow is incomplete and

somewhat meaningless: if one says one's income is \$100, we need to know whether that is per hour, per day, per week, or per year to make sense of it. It is also useful to work with growth of flow variables, often calculated as annual growth rates. For example, your employer might offer a labour contract that provides for annual cost of living increases equal to 4% per year. In the first year you would receive \$20,000, while in the second you would receive a wage income of \$20,800 (\$20,000 plus 4% of \$20,000, which is equal to \$800).

What flows? When we speak of the flow of a river, it is obvious that it is water, which is flowing, measured in terms of thousands of cubic metres per second. However, it is not so clear what is flowing when we refer to flows of income and expenditure. For example, what flows to provide a wage income equal to \$20,000 per year? The simple answer is 'dollars'. You work for your employer 8 hours a day, 5 days a week, and after two weeks you receive a cheque drawn on a bank or an electronic transfer for the sum of \$800 (ignoring possible deductions for taxes and benefits). Even on payday, it is difficult to conceive of the pay cheque as the 'dollars' that were flowing while you were working. Actually, as we will see in Chapter 6, the cheque is really just an IOU issued by your employer's bank that is denominated in your nation's money of account - the dollar in our example.

In fact, we can conceive of your work for hourly wages as an implicit accumulation of the IOUs of your employer. Over the course of the two weeks during which you worked, you earned a flow of wages equal to \$10 for each hour worked, received in the form of an implicit promise from your employer to pay you in dollars at the end of the two week period. Indeed, in the event of a dispute, the court system would recognize the legal obligation of your employer to pay dollars to you for hours worked. In this sense, we can conceive of each hour worked leading to your accumulation of IOUs of your employer denominated in dollars. On payday, your employer extinguishes their IOUs by delivering to you a cheque or a transfer for the total obligations accumulated over the two-week period. Two important conclusions follow from this example.

Flows are measured in terms of money. The money of account is the means by which we measure flows of income or spending. The associated flow of currency can take a physical form of notes and coins, but equally can be an electronic entry, say in a private bank account. Thus, in contrast to a flow of water, the flows of spending or income do not always take a physical form.

As we will explore later, metal coins and paper currency are really nothing more than government IOUs denominated in the money of account. While government currency is in some respects different from the cheques issued by banks and from the implicit IOUs you accumulate against your employer, all share a common characteristic because all are IOUs denominated in dollars.

We also need to differentiate between flows of income and spending denominated in the money of account from the associated flows of (labour) services and goods and services. In principle, consumer goods and services are used up to satisfy the needs and desires of households, however, consumption purchases made this week could include goods that will be used for many months or even years. Economists typically record consumption at the time the purchase is made and at the dollar value of the purchase even while recognising that goods and services purchased might provide a stream of 'satisfaction' over a long period of time.

## Stocks

**Flows accumulate as stocks.** The flow of water in a stream can be accumulated in a reservoir behind a dam, or in the cup we dip into the stream. The stock of water is then the number of cubic metres in the reservoir, or the half litre in the cup. Unlike a flow, a stock can be measured without reference to a time period as it exists at a point of time. We can measure the stock of water in a lake at noon on the last day of the summer as 1.5 billion cubic metres, and at noon on the last day of the following winter as 2.0 billion cubic metres. Because the stock has increased, we can surmise that the inflow of water during the passing of six months has been greater than the outflow of water over that period, by an amount equal to 0.5 billion cubic metres.

Let us continue to assume that you receive a biweekly pay cheque equal to \$800, twenty-five times a year for a total annual income of \$20,000. On payday, you deposit your employer's cheque in your bank account, increasing your deposit by \$800. Your bank deposit represents a portion of your wealth, held in the form of a financial asset, which is a claim on your bank. Because wealth is measured at a point in time, it is a stock variable. In addition to your bank account, you might also hold other forms of financial wealth (stocks and bonds, currency in your pocket, other types of bank deposits) as well as real wealth (a car, real estate, a business firm, art and jewels). Again, all of these are stock variables whose value is measured in terms of the money of account at a point in time.

Once you have deposited your \$800 pay cheque, you begin to draw down your bank account to finance your purchases. Let us continue to assume that your annual consumption will be \$18,000 for the year, comprised of purchases of consumer goods (food, fuel for your automobile, clothing) and consumer services (entertainment, medical care, legal services). Hence, between pay cheques, you spend a total of \$720 for consumption, drawing down your bank account by that amount to finance these purchases.

Over the year, your flow of wage income has been equal to \$20,000 and you have spent \$18,000 of that on consumption. Then you have accumulated a stock equal to \$2000 - which is equal to the inflow of income less the outflow of spending. Recalling our definition from above, your flow of saving over the year is also equal to \$2000, because saving is defined as the residual dollar value of income that has not been spent over the period.

This will accumulate as an addition to your stock of wealth. If you allow the funds to accumulate in your cheque account - which we will initially assume does not earn interest - the annual addition to your financial wealth will be \$2000. Alternatively, you could instead purchase interest-earning bonds, another form of financial wealth. In this case, however, you will also have a flow of interest earnings, in addition to your labour income. The flow of interest income - let us say it amounts to \$200 over the course of the year - will also add to your stock of financial wealth (so that the total addition to your stock of financial wealth is \$2200).

However, there are many other possible uses of your saving flow. You might decide to buy stocks or other kinds of financial assets. Or, you might purchase real assets - a collectable car, real estate, or equipment for your family's business firm. The saving decision can be analysed as a two-step process: first as a decision to withhold a portion of one's income flow from spending, and second a decision as to the form in which wealth will be accumulated. An income flow is first realised as an accumulation of IOUs - normally, claims on a bank in the form of a deposit - that in the second step is used to purchase an asset.

**One's financial asset is another's financial liability.** It is a fundamental principle of accounting that for every financial asset there is an equal and offsetting financial liability. The cheque deposit (also called a demand deposit or a sight deposit) is a household's financial asset, offset by the bank's liability (or IOU). A government or corporate bond is a household asset, but represents a liability of the issuer (either the government or the corporation). The household has some liabilities, too, including student loans, a home mortgage, or a car loan. These are held as assets by the creditor, which could be a bank or any of a number of types of financial institutions including pension funds, hedge funds, or insurance companies. A household's net financial wealth is equal to the sum of all its financial assets (equal to its financial wealth) less the sum of its financial liabilities (all of the money-denominated IOUs it issued). If that is positive, it has positive net financial wealth.

Examples of stocks include: stock of capital; inventories; financial wealth; and net worth.

### Inside wealth versus outside wealth

It is often useful to distinguish among types of sectors in the economy. The most basic distinction is between the public sector (including all levels of government) and the domestic private sector (including households and firms). Note here we are simplifying by excluding the foreign sector as if the economy was completely closed to trade and capital flows.

If we were to take all of the privately-issued financial assets and liabilities, it is a matter of logic that the sum of financial assets must equal the sum of financial liabilities. In other words, net financial wealth would have to be zero if we consider only private sector IOUs. This is sometimes called 'inside wealth' because it is 'inside' the private sector. In order for the private sector as a whole to accumulate net financial wealth, it must be in the form of 'outside wealth', that is, financial claims on another sector. Given our basic division between the public sector and the domestic private sector, the outside financial wealth takes the form of government IOUs. The private sector holds government currency (including coins and paper currency) as well as the full range of government bonds (short term bills, longer maturity bonds) as net financial assets, which is a portion of its positive net wealth.

Net private financial wealth equals public debt. Recall from our discussion above that accumulation of stocks requires flows. The private sector accumulation of net financial assets over the course of a year is made possible only because its spending is less than its income over that same period. In other words, it has been saving, enabling it to accumulate a stock of wealth in the form of financial assets. In our simple example with only a public sector and a domestic private sector, these net financial assets are government liabilities—government currency and government bonds. These government IOUs, in turn, can be accumulated only when the government spends more than it receives in the form of tax revenue. This is called a 'government deficit', which is the flow of government spending less the flow of government tax revenue measured in the money of account over a given period (usually, a year). This deficit accumulates to a stock of government debt—equal to the private sector's accumulation of financial wealth over the same period.

A complete explanation of the process of government spending and taxing will be provided in Chapter 13. What is necessary to understand at this point is that the net financial assets held by the private sector are exactly equal to the net financial liabilities issued by the government in our

two-sector example. If the government spending always equals its tax revenue, the private sector's net financial wealth would be zero.

**Rest of world debts are domestic financial assets.** We can broaden our analysis by considering the financial assets and liabilities of the rest of the world. So we now form three sectors in this open economy: a domestic private sector, a domestic public sector, and a 'rest of the world' sector that consists of foreign governments, firms, and households. In this case, it is possible for the domestic private sector to accumulate net financial claims on the rest of the world, even if the domestic public sector runs a balanced budget, with its spending over the period exactly equal to its tax revenue. The domestic sector's accumulation of net financial assets is equal to the rest of the world's issue of net financial liabilities. Finally, and more realistically, the domestic private sector can accumulate net financial wealth consisting of both domestic government liabilities as well as rest of world liabilities. It is also possible for the domestic private sector to accumulate government debt (adding to its net financial wealth) while also issuing debt to the rest of the world (reducing its net financial wealth). In the next section we turn to a detailed discussion of sectoral balances.

### **Non-financial wealth (real assets)**

One's financial asset is necessarily offset by another's financial liability. However, real assets represent one's wealth that is not offset by another's liability, hence, at the aggregate level net wealth equals the value of real (non-financial) assets. To be clear, you might have purchased an automobile by going into debt. Your financial liability (your car loan) is offset by the financial asset held by the auto loan company. Since those net to zero, what remains is the value of the real asset - the car. In most of the discussion that follows we will be concerned with financial assets and liabilities, but will keep in the back of our minds that the value of real assets provides net wealth at both the individual level and at the aggregate level. Once we subtract all financial liabilities from total assets (real and financial) we are left with non-financial (real) assets, or aggregate net worth.

## **5.4 Integrating NIPA, Stocks, Flows and the Flow of Funds Accounts**

The sectoral balances framework, which is derived from the national accounts framework, was explored in Section 5.2. It is intrinsically linked to the flow of funds analysis. They are different, but related, ways of considering national economic activity.

An early exponent of the flow-of-funds approach, Lawrence Ritter (1963:220) wrote that:

The flow of funds is a system of social accounting in which (a) the economy is divided into a number of sectors and (b) a 'sources- and-uses-of-funds statement' is constructed for each sector. When all these sector sources-and-uses-of-funds statements are placed side by side, we obtain (c) the flow-of-funds matrix for the economy as a whole.

Thus, the flow-of-funds accounts allow us to link a sector's balance sheet (statements about stocks of financial and real net wealth) to income statements (statements about flows) in a consistent fashion. In a monetary economy, flows of expenditures measured in terms of dollars spent over a period involve transactions between sectors in the economy, which also have logical stock counterparts, that is flows feed stocks. The flow-of-funds accounts ensure that all of these transactions are correctly accounted for.



This approach underpinned the work of the so-called New Cambridge approach that was part of the Cambridge Economic Policy Group at the University of Cambridge in the early 1970s. Key members of this group were Martin Fetherston, Wynne Godley and Francis Cripps, all of who were of a Keynesian persuasion.

While the sectoral balances approach had been understood much earlier (for example, by Nicholas Kaldor and others), it was popularised by the New Cambridge macroeconomic analysis which introduced the concept of the **Net Acquisition of Financial Assets (NAFA)** into the forefront of its Keynesian income-expenditure model (see below).

Like Lawrence Ritter, the Cambridge economists were interested in tracing the flow of funds between the different sectors of the economy, which they divided into the government sector; the private domestic sector and the external sector, as outlined above. These transactions have occurred in a given period, and these sectors could record a financial deficit or surplus.

We can re-write Equation (5.5) as follows:

$$(5.5) \quad (S - I) = NAFA = (G - T) + CAB$$

$(S - I)$  is the private domestic financial balance or NAFA of the private domestic sector. The private domestic sector is in financial surplus (deficit) when its disposable income ( $GNP - T$ ) exceeds (is less than) its spending on consumption goods and investment goods.

From a stock perspective, NAFA can also be measured by the difference between the private domestic sector's stock of net financial assets at time  $t$  and the stock at time  $t-1$ , where  $t$  could be 2016, so that  $t-1$  would be 2015.

Noting the stock/flow distinction, Equation (5.5) can be interpreted as meaning that if its right hand side is positive, government sector deficits ( $G - T > 0$ ) and current account surpluses ( $CAB > 0$ ) generate national income and additional net financial assets for the private domestic sector. Then  $NAFA > 0$ , which means that the private sector is running a surplus, and acquiring new assets and/or reducing its existing debt obligations, whereas the government financial balance is negative.

Conversely, fiscal surpluses ( $G - T < 0$ ) and current account deficits ( $CAB < 0$ ) reduce national income and undermine the capacity of the private domestic sector to net save and add to its stock of net financial assets. In this case  $NAFA < 0$ , so that the private domestic sector is running down its net financial position by borrowing from the other sectors and/or by liquidating some of its stock of accumulated wealth.

If  $G - T < 0$ , then the government sector is spending less than it is taking out of the economy in taxation and undermining the capacity of the other two sectors to accumulate net financial assets by running surpluses and vice versa.

CAB is the external sector financial balance (the Current Account Balance) and comprises the **trade balance** (that is, the difference between export and import revenue on goods and services) and the **net income flows** that accrue to residents as a consequence of interest and dividends received on overseas ownership (offset by similar payments to foreigners).

If the overall external sector balance is in deficit then the national economy is borrowing from abroad or running down its net financial position in other ways and foreigners are accumulating financial asset claims and vice versa.

Equation (5.5) can also be written as:

$$(5.6) \quad [(S - I) - CAB] = (G - T)$$

where the term on the left-hand side  $[(S - I) - CAB]$  is the non-government sector financial balance and is of equal and opposite sign to the government financial balance,  $T - G$ .

**This is the familiar Modern Monetary Theory (MMT) conclusion that a government sector deficit (surplus) is equal dollar-for-dollar to the non-government sector surplus (deficit).**

MMT adopts the same interpretation of these balances as the New Cambridge approach, but when applied to the government sector, any conclusion is somewhat meaningless other than in a purely accounting sense.

Importantly, transactions within the private domestic sector do not alter the net financial position of the sector overall. For example, if a bank creates a loan for one of its customers then its assets rise but on the other side, the liabilities of the customer increases by an equal amount – leaving no change in the **net** position of the sector.

The only way the private domestic sector can increase its net financial assets is through transactions with the government or external sector – for example, by acquiring a government bond or buying a foreign government bond (or a foreign corporate bond). These two points are key MMT insights.

Once we understand the interlinked nature of the three sectors then it is a simple step to realise that if one sector has improved its position by the net acquisition of financial assets, following a financial surplus, at least one other sector must have reduced its net financial assets or run a financial deficit.

The flow-of-funds framework allows us to understand that the funds a particular sector receives during a period from current receipts, borrowing, selling financial assets, and running down cash balances have to be equal to the total of its current expenditures, capital expenditures, debt repayments, lending, and accumulation of cash balances. The approach clearly allows us to trace the uses and sources of funds for each sector.

It should be emphasised that the flow-of-funds approach is based on accounting principles rather than being a behavioural (theoretical) framework for understanding the factors, which explain the magnitudes of these flows. Relatedly, there are no insights into the adjustment processes that govern the change in net financial assets in each sector.

That is not to be taken as a criticism of the approach – it is merely an observation. It also doesn't reduce the utility and insights that the approach provides. Often economists like to denigrate analyses that manipulate accounting identities as if they are too low brow. But any approach is valuable if it provides useful ways of thinking.

### **Causal relationships**

From the discussion above, it is clear that a non-government surplus is the same thing as a saving flow and leads to the net accumulation of financial assets. By the same token, a deficit reduces net financial wealth. If the private domestic or external sector runs a deficit, it must either use its financial assets that have been accumulated in previous years (when surpluses were run), and reduce its bank deposits, or it must issue new IOUs to obtain bank deposits to offset its deficits. Once it runs out of accumulated assets, it has no choice but to increase its indebtedness every

year that it runs a deficit. On the other hand, if the external or private domestic sector runs a surplus then it will be accumulating net financial assets. This will take the form of financial claims on at least one of the other sectors.

As we will discuss later, it is misleading to apply terminology such as ‘dis-saving’ or ‘borrowing’ to the sovereign government, which issues the currency.

While we have identified an accounting relationship between the sectoral balances, we can say something about causal relationships between the flows of income and expenditure and the impact on stocks.

Individual spending is mostly determined by income. For the individual, it is plausible to argue that income determines spending because one with no income is certainly going to be severely constrained when deciding to purchase goods and services. However, on reflection it is apparent that even at the individual level, the link between income and spending is loose - one can spend less than one’s income, accumulating net financial assets, or one can spend more than one’s income by issuing financial liabilities and thereby becoming indebted. Still, at the level of the individual household or firm, the direction of causation runs from income to spending even if the correspondence between the two flows is not perfect.

### **Deficits create financial wealth**

We can also say something about the direction of causation regarding accumulation of financial wealth at the level of the individual. If a household or firm decides to spend more than its income by running a deficit, it can issue liabilities to finance purchases. Another household or firm will accumulate these liabilities as net financial wealth. Alternatively, they might allow the government to run a fiscal surplus. Of course, for this net financial wealth accumulation to take place, we must have one household or firm willing to deficit spend, and another household, firm, or government willing to accumulate wealth in the form of the liabilities of that deficit spender. So ‘it takes two to tango’. However, the decision to deficit spend is the initiating cause of the creation of net financial wealth. No matter how much others might want to accumulate financial wealth, they will not be able to do so unless someone is willing to deficit spend. Still, it is true that the household or firm will not be able to deficit spend unless it can sell accumulated assets or find someone willing to hold its liabilities, such as a bank through the creation of a loan.

In the case of a sovereign government, there is a special power - the ability to tax, that guarantees that households and firms will want to accumulate the government’s debt. We conclude that while causation is complex, it tends to run from individual deficit spending to accumulation of financial wealth by another economic entity, and from debt to financial wealth. Since the accumulation of a stock of financial wealth results from a surplus, that is, from a flow of saving, we can also conclude that causation tends to run from deficit spending to saving. At the sectoral, rather than individual, level the same principles apply. Thus, one sector cannot run a deficit if no other sector will run a surplus. Equivalently, we can say that one sector cannot issue debt if no other sector is willing to accumulate the debt instruments.

Aggregate spending creates aggregate income. At the aggregate level, taking the economy as a whole, causation is more clear-cut. A society cannot decide to have more income, but it can decide to spend more. Further, all spending must be received by someone, somewhere, as income. Finally, as discussed above, spending is not necessarily constrained by income because it is possible for households, firms, or government to spend more than income. Indeed, as we

discussed, any of the three main sectors can run a deficit with at least one of the others running a surplus. However, it is not possible for spending at the aggregate level to be different from aggregate income since the sum of the sectoral balances must be zero. For all of these reasons, we must reverse causation between spending and income when we turn to the aggregate: while at the individual level, income causes spending, at the aggregate level, spending causes income.

In MMT, we differentiate between horizontal and vertical transactions within the economy. Horizontal transactions occur between people and firms within the non-government sector (for example, purchases of goods and services, borrowing from banks). Vertical transactions occur between the government sector and the non-government sector (for example, government spending and taxation).

**Horizontal transactions do not add to the stock of net financial assets held by the non-government sector.** Much of the debt issued within a sector will be held by others in the same sector. For example, if we look at the finances of the private domestic sector we will find that most business debt is held by domestic firms and households. In the terminology we introduced above, this is ‘inside debt’ of those firms and households that run budget deficits, held as ‘inside wealth’ by those households and firms that run budget surpluses. Likewise if households choose to deficit spend, that is, spend more than their flow of annual income, then they may secure bank loans. In this case the net asset position of the private sector is unchanged. These are horizontal transactions.

However, if the domestic private sector taken as a whole spends more than its income, it must issue ‘outside debt’ held as ‘outside wealth’, which would be held by the foreign sector, but the stock of net financial assets held by the non-government sector (private domestic plus foreign) is again unchanged, since these are horizontal transactions.

The initiating cause of the private sector deficit is assumed to be a desire to spend more than income, so the causation mostly goes from deficits to surpluses and from debt to net financial wealth. While we recognise that no sector can run a deficit unless another wants to run a surplus, this is not usually a problem because there is a propensity to net save and acquire financial assets.

**Vertical transactions do add to the stock of net financial assets held by the non-government sector.** On the other hand, assume that a fiscal deficit occurs (perhaps as a result of increased government spending), and for simplicity the CAB is zero, then the private sector achieves a net increase in its stock of financial assets. This transaction between the government and private sector is referred as a vertical transaction and, in this instance, leads to an increase in net financial assets held by the non-government sector. On the other hand, if the government runs a fiscal surplus (by taking net spending out of the economy), with the CAB zero, the non-government sector (specifically the private sector) suffers a loss in its net holdings of financial assets.

In this section, we demonstrate how a flow-of-funds approach to the analysis of monetary transactions highlights both the importance of the distinction between and vertical and horizontal transactions and the fundamental accounting nature of the so-called government ‘budget’ constraint (GBC) identity, which we will refer to as the government fiscal constraint.

## 5.5 Balance Sheets

Following Ritter, we can present a very simple ‘generalised balance sheet’, which would apply to any sector, as being depicted in the following T-account, Figure 5.2

Several points are worth noting. Real assets are treated differently to financial assets because they only appear on the balance sheet of the owner. Financial liabilities are different because their existence as debt (to some other sector) means they will be matched by a financial asset on at least one other sector’s balance sheet.

Financial assets denote monetary amounts owned by that sector, which by the same logic as before means that there will be a matching liability on at least one other balance sheet within the system.

When we consider the monetary system as a whole, we conclude that financial assets and financial liabilities net to zero – that is, the total value of the financial assets equals the total value of outstanding liabilities.

The accounting also tells us that for the overall economy, net worth equals to monetary value of the real assets in the economy

**Figure 5.2 A stylised sectoral balance sheet**

Assets	Liabilities and Net Worth
Financial assets:	Liabilities
1. Money	
2. Other	
Real assets	Net worth
$\Sigma$	$\Sigma$

The balance sheet depicts stocks but we can easily see how they might provide us with information about flows, in the way the national accounts does. A stock is measured at point in time (say, the end of the year) whereas flows measure monetary transactions over a period (say, a year).

If we examine the difference between a balance sheet compiled at say December 31, 2015, and a balance sheet compiled at December 31, 2016, we will be able to represent the information in the balance sheet about assets, liabilities and net worth as flow data.

Consider Figure 5.3 (where the  $\Delta$  symbol refers to changes over the period concerned). Now the entries in the T-account denote uses and sources of funds (that is, flows) over the period of interest. There are two components, one relates to financial assets and the other real assets and net worth.

A given sector (for example, household, firm, government) can obtain funds by increasing their liabilities by borrowing and incurring debt ( $\Delta L$ ). They can apply those funds to accumulating more financial assets ( $\Delta FA$ ) or building cash balances ( $\Delta M$ ).

**Figure 5.3 A uses-and-sources-of-funds statement**

Uses	Sources
$\Delta$ Financial assets (lending)	$\Delta$ Liabilities (borrowing)
$\Delta$ Money (cash balances)	
$\Delta$ Real assets (investment)	$\Delta$ Net worth (saving)
$\Sigma$	$\Sigma$

If we wanted to complicate matters we could decompose  $\Delta FA$ ,  $\Delta M$  and  $\Delta L$  further, by recognising that a given sector can also sell existing financial assets or run down cash balances to obtain new funds. Similarly, it might use funds to reduce liabilities (pay down debts). So the entries in Figure 5.3 are to be considered **net transactions**.

The second source and use of funds for a sector relates to changes in Real assets ( $\Delta RA$ ) and the change in net worth ( $\Delta NW$ ) over a given period.

In the national accounts framework (see Chapter 4), we considered the division between the capital account and the current account, where the former related to investment in productive capacity and the latter referred to recurrent spending and income. The capital account measured transactions, which change the real assets held and the net worth of the economy.

What do we mean by a change in real assets? In the national accounts, we considered gross capital formation or investment, which is defined as expenditure on productive capital goods (for example, plant and equipment, factories). This is a use of funds by firms in the current period. Depreciation represents the difference between gross and net investment. For now though we abstract from that real world complexity.

Finally, we consider the change in net worth for a sector in a given period is the residual after all the uses and sources of funds have been accounted for. From an accounting perspective, net worth is equal to the difference between total assets and total liabilities.

It follows that a change in net worth over the period of interest is equal to the difference between the change in total assets and the change in total liabilities. If total assets increase by more (decrease by less) than total liabilities increase (decrease) then the net worth of the sector has risen.

Another way of thinking about the change in net worth, which is a flow of funds, is to link it to the national accounts concept of saving.

In the national accounts framework, we consider household saving, for example, to be the difference between consumption (a use) and disposable income (a source). This concept generalises (with caution) to the statement that the surplus of a sector is the difference between its current revenue and its current expenditure.

What happens to the flow of surplus funds? If the current flow of income is greater than the current expenditure, then at the end of the period, the sector would have accumulated an

increased stock of total net assets – either by increasing the actual assets held and/or reducing liabilities owed.

The surplus between current income and current expenditure has to be matched \$-for-\$ by an increase in the stock of total net assets. We have already discussed total net assets above but in different terms.

We defined the change in net worth over a period as the difference between the change in total assets and the change in total liabilities. That difference is exactly equal to the surplus of current income over current expenditure.

Thus, from an accounting perspective, we can consider saving to be the change in net worth over a period.

Figure 5.3, however, only implicitly includes the current account transactions – the flow of current income and expenditure – inasmuch as we have defined the change in net worth ( $\Delta NW$ ) to be the difference between the two current flows.

The simplicity of Figure 5.3, however, makes clear an essential insight – if a sector is running a deficit (that is, it is spending more than it is earning or in the parlance used above, it is investing more than it is saving) then it must obtain the deficit funds from its available sources:

- Increased borrowing
- Running down cash balances
- Selling existing financial assets

Conversely, a sector that is running a surplus (that is, it is spending less than it is earning or in the parlance used above, it is investing less than it is saving) must be using the surplus funds to:

- Repay debt
- Build up cash balances
- Increase its financial assets (increasing lending)

We also have to be cautious in our terminology when considering the different sectors. If we are considering the household sector, then it is clear that if they spend less than their income and thus save, they are deferring current consumption in the hope that they will be able to command greater consumption in a future period. The increase in their net worth provides for increased future consumption for the household.

Similarly, for a business firm, if they are spending less than they are earning, we consider them to be retaining earnings, which is a source of funds to the firm in the future.

We consider the private domestic sector as a whole (the sum of the households and firms) to be saving overall, if total investment by firms is less than total saving by households. From the national accounts, we consider that households save and firms invest.

However, in the case of the government sector such terminology would be misleading. If the government spends less than they take out of the non-government sector in the form of taxation we say they are running a fiscal surplus. A fiscal deficit occurs when their spending is greater than their taxation revenue.

But a fiscal surplus does not increase the capacity of the sovereign government to spend in the future, in the same way that a surplus (saving) increases the capacity of a household to spend in the future.

As we saw in Chapter 1, a sovereign, currency-issuing government faces no intrinsic financial constraints, and can, at any time, purchase whatever is for sale in the currency that it issues. Its capacity to do so is not influenced by its past spending and revenue patterns.

Figure 5.4 provides the most comprehensive framework for analysing the flow-of-funds because it brings together the current transactions (income and expenditure), the financial transactions, and the capital transactions that we have dealt with earlier. The capital and financial transactions are captured in changes to the balance sheet (Figure 5.2).

Note when we talk about the sovereign government we are excluding the levels of government that do not issue the currency. State and local governments are more like households or firms in that respect, although they do have the capacity to tax and issue fines.

**Figure 5.4 A complete sector uses-and-sources-of-funds statement**

Uses	Sources
Current expenditure	Current receipts
$\Delta$ Net worth (saving)	
-----	-----
$\Delta$ Financial assets (lending)	$\Delta$ Liabilities (borrowing)
$\Delta$ Money (cash balances)	
$\Delta$ Real assets (investment)	$\Delta$ Net worth (saving)
$\Sigma$	$\Sigma$

The transactions above the dotted line comprise the income statement and record current expenditure (uses). The balancing item above the dotted line constitutes the change in net worth ( $\Delta$ NW) or ‘saving’.

The changes in the balance sheet are shown below the dotted line and the balancing item is once again, the change in net worth ( $\Delta$ NW).

You can see that we could cancel out the change in net worth ( $\Delta$ NW), which is the balancing item in both the income statement and the change in the balance sheet. This would leave us with the accounting statement that that sources of funds to a sector through current income and borrowing must, as a matter of accounting, be used – for current expenditures, investment, lending, and/or building up cash balances.



## 5.6 The Flow of Funds Matrix

The T-accounts tracing the sectoral sources and uses of funds can be summarised for all sectors in the economy by the Flow-of-Funds Transactions Matrix, a stylised version of which is shown in Figure 5.5.

The overriding accounting rule that governs the presentation of the flow-of-funds accounts is that for the economy as a whole and for each sector in the economy the total sources of funds must be equal to the total uses of funds. Remember that sources of funds provided by the various sectors in the economy are used by those sectors.

Figure 5.5 (taken from Ritter, 1963) shows three sectors and the total economy. At the most aggregate level, the three sectors could be the private domestic sector, the government sector and the external sector.

**Figure 5.5 A stylised three sector flow-of-funds matrix**

	Sector A		Sector B		Sector C		Total Economy	
<b>Flow</b>	<b>U</b>	<b>S</b>	<b>U</b>	<b>S</b>	<b>U</b>	<b>S</b>	<b>U</b>	<b>S</b>
Saving ( $\Delta NW$ )								
Investment ( $\Delta RA$ )								
Lending ( $\Delta FA$ )								
Cash balances ( $\Delta M$ )								
Borrowing ( $\Delta L$ )								

For each period being accounted for, the statistician would record the flows of funds that related to each of the row categories in the matrix. Most importantly, we have learned that for every deficit sector, which saves less than it invests, there has to be offsetting surpluses in at least one other sector.

Lawrence S. Ritter (1963:228-229) called the economy-wide flow-of-funds matrix:

... an interlocking self-contained system ... [which] ... shows, for a specified time period, the balanced sources-and-uses-of-funds statements for each sector, the interrelations among the sectors, and the aggregate totals of saving, investment, lending, hoarding, and borrowing for the economy as a whole. Any one sector may invest more or less than it saves, or borrow more or less than it lends. However, for the economy as a whole, saving must necessarily equal investment, and borrowing must equal lending plus hoarding.

Thus a deficit sector, which saves less than it invests, must be offset by at least one other surplus sector to net the flows to zero.

What are the practical uses of presenting economic data in this way?

Various uses can be made of the information provided in the flow-of-funds accounts.

The flow-of-funds accounts provide information of all financial flows within the economy on a sector-by-sector basis. They allow researchers and policy makers to understand how funds flow from one sector (say the household sector) through the banking system and onto final users by, for example, firms engaged in productive investment.

They also allow researchers and policy makers to monitor major economic trends such as the changing indebtedness of the sectors included and the sources of funding for the respective sectors. For example, an understanding of the flow-of-funds accounts would have provided insights into the growing indebtedness of the private sector prior to the Global Financial Crisis in 2008 and perhaps, alerted policy makers to the likely financial instability arising from these trends.

Economic researchers also use the flow-of-funds accounts to study saving patterns in the economy. The accounts can tell us where the savings of a sector are being deployed. The accounts can tell us which sector(s) are accumulating surpluses or deficits and the division between financial and real assets. They also allow us to understand patterns of gross capital formation.

Economic researchers also use the flow-of-funds accounts to examine the dynamics of such concepts as household wealth. We can learn how household balances sheets change over time and how that wealth is composed. For example, one of the hallmarks of the period leading up to the Global Financial Crisis in many countries was the shift in household wealth to riskier categories, such as share holding sourced from margin loans. The shift in importance in overall wealth from the more secure home mortgages to more risky sources of wealth was significant because it exposed the economies to an increased risk of financial instability.

Finally, central banks use the flow-of-funds accounts to help them estimate the sensitivity of the economy to changes in the availability of credit.

### **Flow of funds accounts and the national accounts**

The flow-of-funds accounts complement the national accounts and the balance of payments accounts, which are produced by national statistical agencies on a regular basis, as a way of measuring economic activity in total and across the broad economic sectors.

We will consider the balance of payments accounts in Chapter 16.

There are important differences between the flow-of-funds accounts and the national accounts, which can be summarised as:

- The national accounts contain no data pertaining to financial transactions – borrowing, lending or changes in cash balances. Only non-financial transactions are measured. The flow-of-funds accounts fill that void.
- The national accounts focus on the current flows of final expenditure, output and income. As we saw in Chapter 4, transactions that involve so-called double counting or intermediate transactions are excluded from the calculations of final expenditures. The flow-of-funds accounts allow us to trace transactions involving assets that have been created in past periods.

- The structure of the national accounts is such that consumer durable expenditure is included under current expenditure when conceptually it should be considered investment activity. In the flow-of-funds accounts all sectors can invest and save.

## Appendix

### A graphical framework for understanding the sectoral balances

From Equation (5.4) we learned that the sum of the sectoral balances is zero as a matter of accounting – so  $(I - S) + (G - T) + (X - M) = 0$ . We can construct an axis defining four quadrants. Figure 5.6 depicts the government fiscal balance on the vertical axis and the external balance on the horizontal axis.

So all points above zero on the vertical axis represent a government fiscal surplus ( $G < T$ ) and all points below zero on the vertical axis denote government fiscal deficits ( $G > T$ ).

Similarly, all points to the right of the zero line on the horizontal axis denote external surpluses ( $X > M$ ) and all points to the left of zero on the horizontal axis represent external deficits ( $X < M$ ).

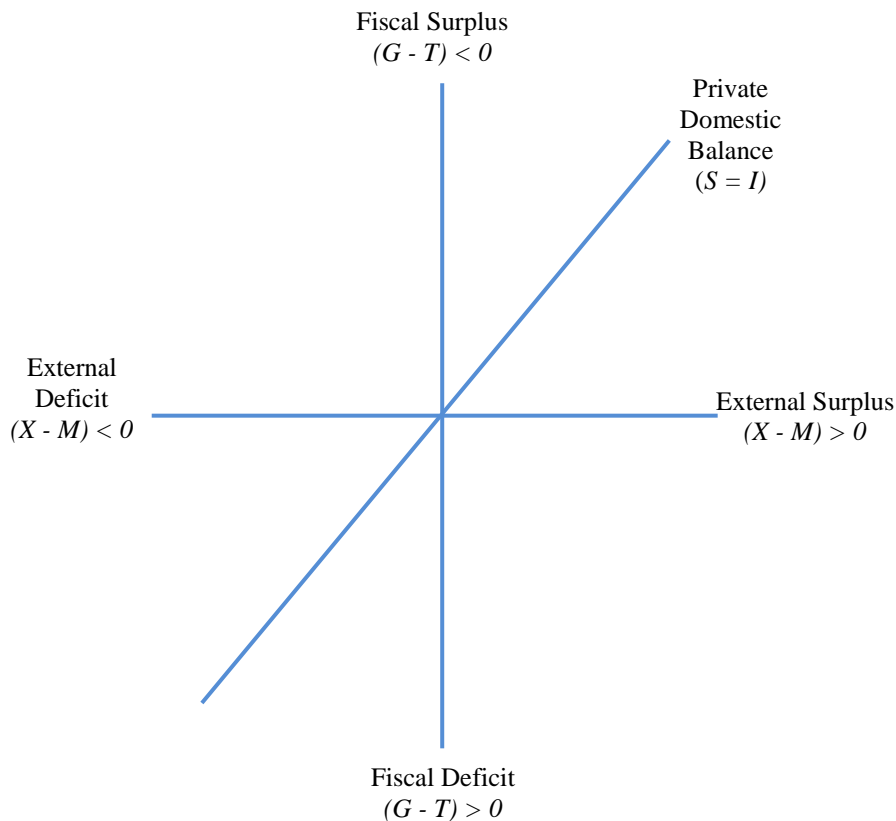
Clearly, the origin of the axis denotes a position where all balances are equal to zero. From the insight gained from Equation (5.4), we also know that when the private domestic balance is zero ( $S = I$ ), then the government fiscal deficit (surplus) has to equal the external deficit (surplus). From Figure 5.6, the diagonal 45-degree line thus shows all combinations of government fiscal balances and external balances where the private domestic balance is zero ( $S = I$ ). We will refer to this as the **SI line**.

We can use that knowledge to determine the segments of the diagram where the private domestic balance is in surplus ( $S > I$ ) and in deficit ( $S < I$ ). To make it easier, we can express the sectoral balances equation (5.4) in a different way:

$$(5.7) \quad (S - I) = (G - T) + (X - M)$$

Equation (5.7) is just another way of expressing the accounting rule but in this case isolates the private domestic balance on the left-hand side.

## 5.6 A graphical sectoral balances framework



Consider the orange dotted line  $A_0B_0$  in Figure 5.6a, which cuts the horizontal axis where the external deficit is 2 per cent of GDP. All points on that line segment correspond to a government fiscal deficit of less than 2 per cent of GDP or a government fiscal surplus depending on which side of the horizontal axis we are considering.

The private sector is in balance only at Point  $B_0$  where the budget deficit equals the external deficit. So all points along  $A_0B_0$ , except  $B_0$  itself, correspond to private domestic deficits ( $S < I$ ).

Now consider the green dotted line  $A_1B_1$ , which cuts the horizontal axis where the external surplus is 2 per cent of GDP. All points on that line segment correspond to a government fiscal surplus of less than 2 per cent of GDP or a government fiscal deficit depending on which side of the horizontal axis we are considering.

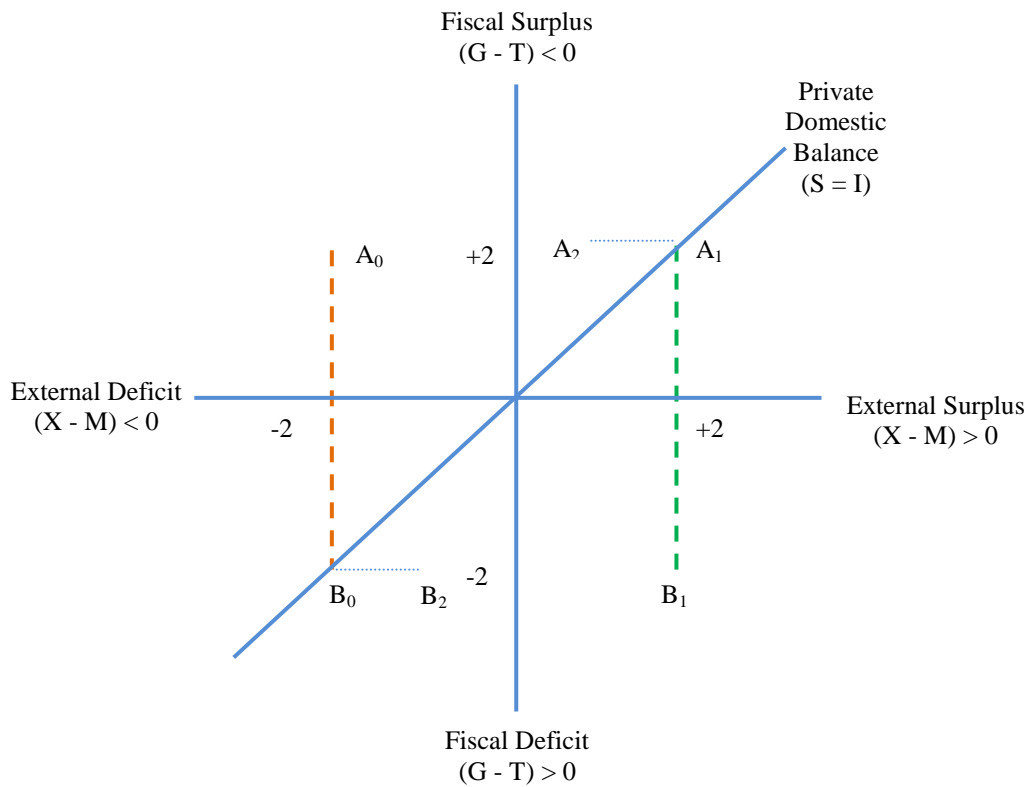
The private sector is in balance only at Point  $A_1$  where the budget surplus equals the external surplus. So all points along  $A_1B_1$  correspond to private domestic surpluses ( $S > I$ ).

Now consider the blue dotted line  $A_1A_2$ . At Point  $A_1$ , the private sector is in balance because the budget surplus equals the external surplus.

Moving along  $A_1A_2$  we encounter points where the external balance is less than the budget surplus and thus have to correspond to a private domestic deficit ( $S < I$ ).

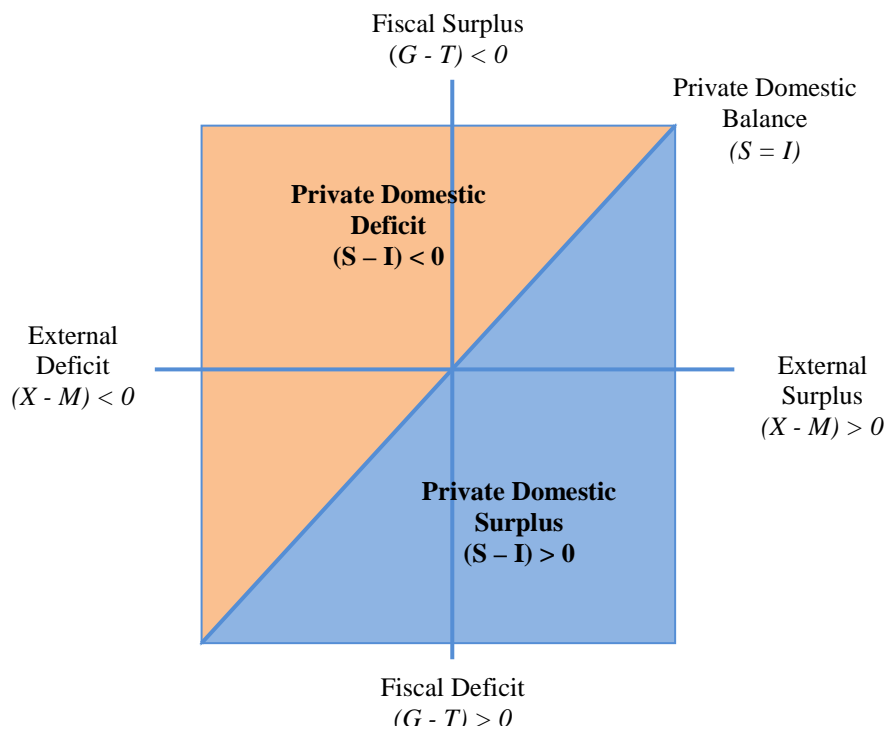
Similarly, along  $B_0B_2$ , we move from a private sector balance at  $B_0$ , to points where the external deficit is less than the budget deficit, which means the private domestic balance will be in surplus ( $S > I$ ).

**Figure 5.6a Deriving the private domestic sector balances**



We can thus generalise this knowledge and conclude that all points above the 45-degree line on either side of the vertical axis correspond to private domestic sector deficits and all points below the 45-degree line on either side of the vertical axis correspond to private domestic sector surpluses. Figure 5.6b renders this conclusion graphically.

**Figure 5.6b Private domestic surpluses and deficits**



This graphical framework thus allows us to examine the implications of different policy options.

For a sovereign, currency-issuing government any point in the four-quadrants is permissible. With private sector spending and saving decisions combining with the flows of income arising from trade with the external sector driving national income, the government sector can allow its balance to adjust to whatever magnitude is required to maintain full employment and price stability.

For example, if the external account was in deficit and the private sector was saving overall, then the drain on aggregate demand would require the government to run a deficit of sufficient size to ensure total spending was sufficient to absorb the real productive capacity available in the economy.

Alternatively, the external account might be in surplus which would add to aggregate demand while the private sector might be spending more than it is earning, that is, in deficit overall. In these situations the government would have to ensure it ran a surplus of sufficient size to ensure the economy did not overheat and exhaust its productive capacity.

The strong economy would be associated with robust tax revenue growth, which would help the government achieve its surplus. Discretionary adjustments in spending and taxation rates might also be required.

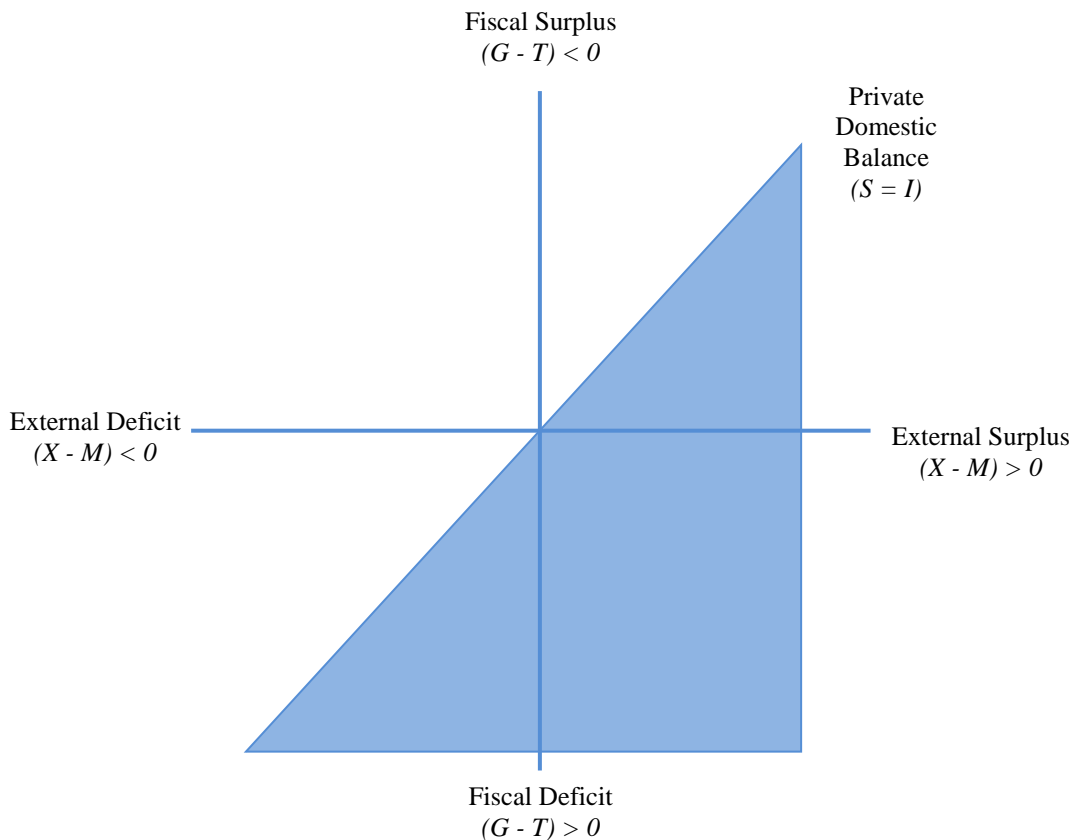
But while any point would be permissible, we know that the private sector cannot sustain deficits permanently. This is because the flows of spending which deliver deficits have to be funded. As we learned in the earlier section of this Chapter when we considered the flow of funds, private deficits ultimately manifest in an increasing stock of debt being held on the private sector's balance sheet.

This process of debt accumulation is limited because at some point the susceptibility of the balance sheet to cyclical movements (for example, rising unemployment) increases and the risk of default rises. In some historical instances, this process has collapsed after serious debt defaults occurred (for example, in the early months of the Global Financial Crisis in 2007-08). At other times, the private sector starts to reduce the precariousness of its balance sheet by reducing spending and increasing saving in order to bring the debt levels it is carrying down to more sustainable levels.

In the long-term, the only sustainable position is for the private domestic sector to be in surplus. An economy can absorb deviations around that position but only for short-periods.

Figure 5.7 shows what we might define as the sustainable space available to governments that issue their own currency. Note this excludes permanent private sector deficits, which are unsustainable.

**Figure 5.7 Sustainable space for sovereign governments**



Now imagine that the government is forced to operate under a fiscal rule that bans fiscal deficits greater than 3 per cent of GDP as shown by the red line in Figure 5.8. The formation of the European Monetary Union introduced just such a fiscal rule under its' Stability and Growth Pact. The aim was to restrict the capacity of each member state to run government budget deficits.

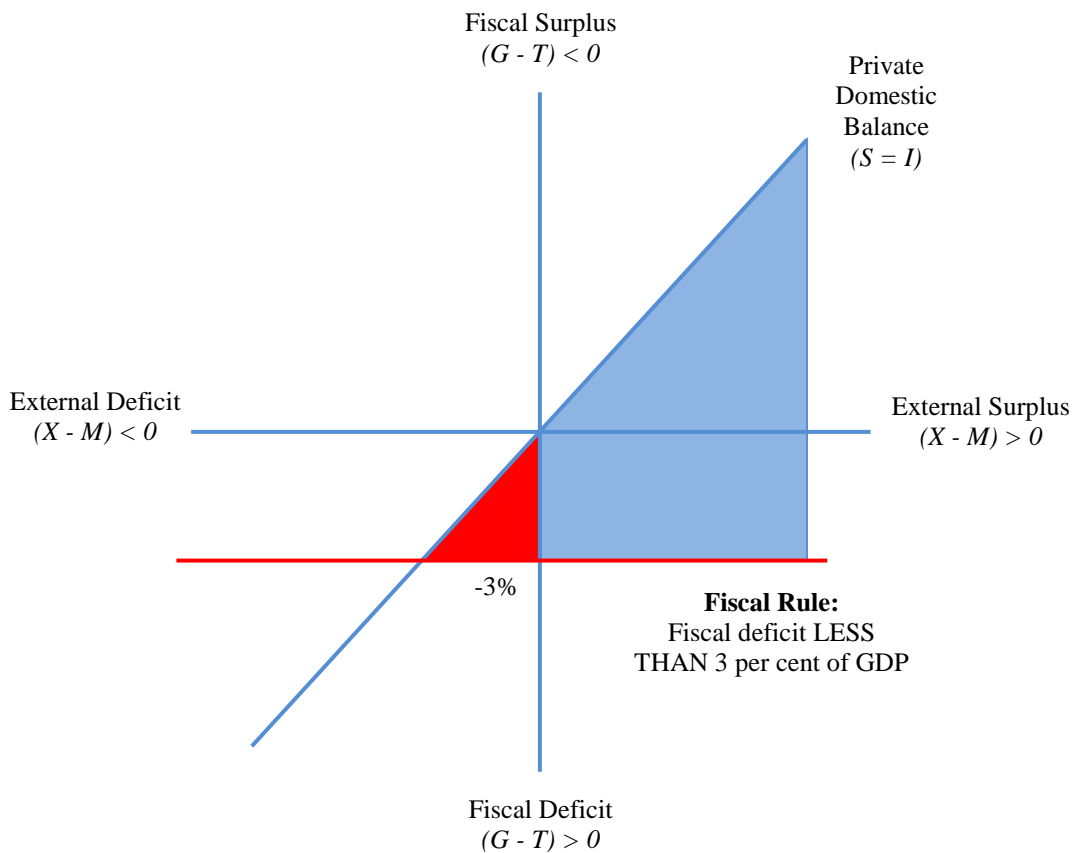
We now must consider what the fiscal rule means for both permissible and sustainable spaces available to a macroeconomic policy maker.

Clearly any point above the 3 per cent of GDP fiscal deficit line in Figure 5.8 is permissible. However, using the same logic as before, the sustainable space requires that the private domestic sector be in surplus overall, even though short-term deviations from this status can occur from time to time.

Figure 5.8 shows the sustainable space for such an economy (the combination of red and blue areas). The blue-shaded area shows the sustainable space available to policy makers in nations that run external surpluses. The red-shaded area shows the sustainable space available to policy makers in nations that run external deficits.

Thus the policy space that governments have to operate within when fiscal rules are imposed is very limited relative to the options available to a sovereign, currency-issuing government, which operates without any direct quantitative restrictions on the deficits they can run.

**Figure 5.8 Sustainable space for governments constrained by fiscal rules**



Why is this important? A rule-free government can always utilise the available space to ensure aggregate demand is sufficient to maintain full employment and price stability.

By definition, not every nation can run an external surplus because a surplus has to be matched by a deficit or deficits in other nations. While the external surplus nations have more policy flexibility when operating under a fiscal rule of the type shown in Figure 5.8, it still remains that the allowable fiscal deficits may be insufficient to maintain the aggregate demand necessary to sustain full employment.



The policy inflexibility facing nations which run external deficits and simultaneously have to operate under fiscal rules as shown in Figure 5.8 become even more restrictive, as shown by the small red triangle. When such an economy experiences a negative economic shock which leads the private sector to seek to reduce its spending and target a sectoral surplus, the extent to which the fiscal deficit can move to absorb the loss of overall aggregate demand is very limited.

It is highly likely that such an economy will experience enduring recessions as a result of the artificial fiscal rules (restrictions) that are placed on the government. Note that such a situation is representative of Greece and some other Mediterranean European nations.

The sustainable goal for a government should be to maintain full employment and price stability and allow its fiscal balance to adjust accordingly to ensure aggregate demand is consistent with those goals. A sovereign, currency-issuing government can always meet those goals if it chooses.

However, the imposition of fiscal rules restricts the government from achieving these goals and makes the fiscal outcome the goal rather than the more significant macroeconomic goals of full employment and price stability.

The lesson is that the government should never specifically target any particular fiscal outcome, but rather, should target employment growth and price stability.

## References

Australian Bureau of Statistics (ABS) (2014) *Australian System of National Accounts: Concepts, Sources and Methods*, ABS: Canberra.

OECD (2015) Organisation for Economic Co-operation and Development, Economic outlook annex tables. Available at: <http://www.oecd.org/eco/outlook/economicoutlookannextables.htm> (accessed 1 October 2015).

Ritter, L.W. (1963) "An Exposition of the Structure of the Flow-of-Funds Accounts", *The Journal of Finance*, 18(2), May, 219-230.

Watts, M.J. and Sharpe, T.P. (2016) *Sustaining Surpluses: Prospects and Implications*, mimeo, The University of Newcastle, NSW, January.

Wray, L.R. (1999) *Surplus mania: A reality check*, Levy Economics Institute, *Policy Note 3*, available: [http://www.levyinstitute.org/pubs/pn99\\_3.pdf](http://www.levyinstitute.org/pubs/pn99_3.pdf).

# Chapter 6: Introduction to Sovereign Currency: The Government and its Money

## Chapter Outline

- 6.1 Introduction
- 6.2 The National Currency (Unit of Account)
  - One nation, one currency
  - Sovereignty and the currency
  - What 'backs up' currency?
  - Legal tender laws
  - Fiat currency
  - Taxes drive the demand for money
  - Financial stocks and flows are denominated in the national money of account
  - The financial system as an electronic scoreboard
- 6.3 Floating versus Fixed Exchange Rate Regimes
  - The gold standard and fixed exchange rates
  - Floating exchange rates
- 6.4 IOU's Denominated in National Currency: Government and Non-Government
  - Leveraging
  - Clearing accounts extinguishes IOUs
  - Pyramiding currency
- 6.5 Use of the Term 'Money': Confusion and Precision

## **Learning Objectives**

1. Explain why a fiat currency is valued and is acceptable in domestic transactions.
2. Recognise the distinction between fixed and floating exchange rate regimes and their significance for the conduct of macroeconomic policy.
3. Understand how IOUs are created and extinguished.

## 6.1 Introduction

In this chapter we will examine in more detail several of the concepts briefly introduced in earlier Chapters of this textbook. We first turn to the money of account and the nation's currency, and note that the latter is not backed by a precious metal, such as gold. We argue that the so-called fiat currency is valued and widely used in transactions because it is required as the means to relinquish tax obligations levied by the state. All financial stocks and flows are denominated in the national money of account. In this context the financial system can be viewed as a record of transactions, that is a scoreboard. We then examine the difference between floating and fixed exchange rate systems. Government and non-government IOUs are denominated in the money of account. After defining leveraging, we argue that these different types of IOUs can be conceived of as a financial pyramid, with government IOUs at the top. Finally we emphasise the need to use the term 'money' very carefully to avoid confusion.

## 6.2 The National Currency (Unit of Account)

Let us look at money as the **unit of account** in which stocks and flows are denominated.

### One nation, one currency

In Chapter 1 we introduced the concept of the money of account. The Australian dollar, the US dollar, the Japanese yen, the British pound, and the European euro are all examples of a money of account. The first four of these monies of account are each associated with a single nation. Throughout history, the usual situation has been 'one nation, one currency', although there have been a few exceptions to this rule, including the modern euro, which is a money of account adopted by a number of countries that have joined the Economic and Monetary Union of the European Union (EMU). When we address the exceptional cases, such as the EMU, we will carefully identify the differences that arise when a currency is used, but not issued, by a nation.

Most of the discussion that follows will be focused on the more common case in which a nation adopts its own money of account. The government of the nation issues a currency (usually consisting of metal coins and paper notes of various denominations) denominated in its money of account. Spending by the government as well as tax liabilities, fees, and fines owed to the government are denominated in the same money of account. These payments are enforceable by law. More generally, broad use of a nation's money of account is ensured by enforcing monetary contracts in the court of law, such as the payment of wages.

In many nations there are private contracts that are written in foreign monies of account. For example, in some Latin American countries it is common to write some kinds of contracts in terms of the US dollar. It is also common in many nations to use US currency in payment. According to some estimates, the total value of US currency circulating outside America exceeds the value of US currency used at home. Much of this is thought to be involved in illegal activities, including the drug trade. Thus, one or more foreign monies of account as well as the corresponding foreign currencies might be used in addition to the domestic money of account and the domestic currency denominated in that unit. Sometimes this is explicitly recognised by, and permitted, by the authorities, while other times it is part of the underground economy that tries to avoid detection by using foreign currency.

## **Sovereignty and the currency**

The national currency is often referred to as a **sovereign currency**, that is, the currency issued by the sovereign government. The sovereign government retains for itself a variety of powers that are not given to private individuals or institutions. Here, we are only concerned with those powers associated with money. The sovereign government alone, has the power to determine which money of account it will recognise for official accounts. Further, modern sovereign governments, alone are invested with the power to issue the currency denominated in each nation's money of account. For example, if any entity other than the US government tried to issue US currency it would be prosecuted as a counterfeiter, with severe penalties resulting. As noted above, the sovereign government imposes tax liabilities (as well as fines and fees) in its money of account, and decides how these liabilities can be paid - that is, it decides what it will accept in payment so that taxpayers can fulfil their obligations. Finally, the sovereign government also decides how it will make its own payments - when it purchases goods or services, or meets its own obligations, such as pensions to retirees. Most modern sovereign governments make payments in their own currency, and require tax payments in the same currency. For reasons that we will examine later, requiring tax payments in the government's currency ensures that the same currency will be accepted in payments made by government.

### **What 'backs up' the currency?**

There is, and historically has been, some confusion surrounding sovereign currency. For example, many policy makers and economists have had trouble understanding why the private sector would accept currency issued by the government when it made purchases. Some have argued that it is necessary to 'back up' a currency with a precious metal in order to ensure acceptance in payment. Historically, governments have sometimes maintained a reserve of gold or silver (or both) against its currency. It was thought that if the population could always return currency to the government to obtain precious metal instead, then the currency would be accepted because it would be thought to be 'as good as gold'. Sometimes the currency itself did contain precious metal - as in the case of gold coins. In the US, the Treasury did maintain gold reserves equal to 25 per cent of the value of the issued currency until the late 1960s, but American citizens were not allowed to redeem currency for gold; only foreign holders of US currency could do so. However, the US and most nations have long since abandoned this practice. And even with no gold backing, the US currency is still in high demand all over the world, so that the view that currency needs precious metal backing is erroneous.

### **Legal tender laws**

Another explanation that has been offered is legal tender laws. Historically, sovereign governments have enacted legislation requiring their currencies to be accepted in payments. Indeed, paper currency issued in the US proclaims 'this note is legal tender for all debts, public and private'; Canadian notes say 'this note is legal tender'; and Australian paper currency reads 'This Australian note is legal tender throughout Australia and its territories.' By contrast, the paper currency of the UK simply says 'I promise to pay the bearer on demand the sum of five Pounds' (in the case of the five pound note). On the other hand, the euro paper currency makes no promises. Further, throughout history there are many examples of governments that passed legal tender laws, but still could not create a demand for their currencies - which were not accepted in private payments, and sometimes even rejected in payment by the government. In

some cases, the penalty for refusing to accept a king's coin included the burning of a red hot coin into the forehead of the recalcitrant. Hence, there are currencies that readily circulate without any legal tender laws as well as currencies that were shunned even with legal tender laws. Further, as we know, the US dollar circulates in a large number of countries in which it is not legal tender (and even in countries where its use is discouraged by the authorities).

### **Fiat currency**

Modern currencies are often called **fiat currencies** because there is no promise made by government to redeem them for precious metal - their value is proclaimed by 'fiat' (the government merely announces that a coin is worth a half-dollar without holding a reserve of precious metal equal in value to a half dollar). Many students in economics courses are shocked when they are first told that there is 'nothing' backing the currency in their pockets. While they had probably never contemplated actually taking the currency down to the treasury to exchange it for gold, they had found comfort in the erroneous belief that there was 'something' standing behind the currency - perhaps a reserve of precious metal available for redemption. The UK currency's 'promise to pay the bearer on demand the sum of five Pounds' appears to offer a sound basis, implying that the treasury holds something in reserve that it can use to make the promised payments. However, if one were to actually present to the UK government a five pound note, the treasury would simply offer another five pound note, or a combination of notes and coins that sum to five pounds! Any citizen of the US or Australia would experience the same outcome at their own treasuries: a five dollar note can be exchanged for a different five dollar note, or for some combination of notes and coins to make five dollars. That is the extent of the government 'promise to pay'!

If currency cannot be exchanged for precious metal, and if legal tender laws are neither necessary nor sufficient to ensure acceptance of a currency, and if the government's 'promise to pay' really amounts to nothing, then why would anyone accept a government's currency? Let us try to determine why.

### **Taxes drive the demand for money**

One of the most important powers claimed by sovereign government is the authority to levy and collect taxes (and other payments made to government including fees and fines). Tax obligations are levied in the national money of account - for example, dollars in the US and Australia, yen in Japan, pounds in the UK and so on. Further, the sovereign government also determines what can be delivered to satisfy the tax obligation. In all modern nations, it is the government's own currency that is accepted in payment of taxes.

While taxpayers mostly use cheques drawn on private banks to make tax payments, when government receives these cheques, it debits the reserves of the private banks, which are held at the central bank. Reserves are just a special form of government currency used by banks to make payments to one another and to the government. Like all currency, reserves are the government's IOU. Effectively, private banks intermediate between taxpayers and government, making payment in currency on behalf of the taxpayers. Once the banks have made these payments, the taxpayer has fulfilled her obligation, so the tax liability is eliminated.

The tax payment reduces the worker's financial wealth because their bank deposit is debited by the amount of the tax payment. At the same time, the government's asset (the tax liability owed

by the worker) is eliminated when the taxes are paid, and the government's liability (the reserves held by private banks) is also eliminated. This is an example of the operation of the payments system, which will be analysed in greater detail in Chapter 13.

We are now able to answer the question posed above: why would anyone accept government's 'fiat' currency? The answer is because the government's currency is the main (and usually the only) thing accepted by government in payment of taxes. It is true, of course, that government currency can be used for other purposes: coins can be used to make purchases from vending machines; private debts can be settled by offering government paper currency; and government money can be hoarded in piggy banks for future spending. However, these other uses of currency are all subsidiary, deriving from government's willingness to accept its currency in tax payments. It is because anyone with tax obligations can use currency to eliminate these liabilities that government currency is in demand, and thus can be used in purchases or in payment of private obligations. The government cannot easily force others to use its currency in private payments, or to hoard it in piggybanks, but government can force use of currency to meet tax obligations that it imposes.

For this reason, neither reserves of precious metals nor legal tender laws are necessary to ensure acceptance of the government's currency. All that is required is the imposition of a tax liability to be paid in the government's currency. The 'promise to pay' that is engraved on UK pound notes is superfluous and really quite misleading. We know that the UK treasury will not really pay anything (other than another note) when the five pound paper currency is presented. However, it will and must accept the note in payment of taxes. This is really how government currency is redeemed - not for gold, but in payments made to the government. We will go through the accounting of tax payments later. It is sufficient for our purposes now to understand that the tax obligations to government are met by presenting the government's own IOUs to the tax collector.

We can conclude that taxes drive money. The government first creates a money of account (the dollar), and then imposes tax obligations in that national money of account. In all modern nations, this is sufficient to ensure that most debts, assets, and prices, will also be denominated in the national money of account. The government is then able to issue a currency that is also denominated in the same money of account, so long as it accepts that currency in tax payment. When we talk about the government 'issuing' currency, the most usual way in which this occurs is through government spending. We say the government spends the currency into existence. It can also make loans.

It is not necessary to 'back' the currency with precious metal, nor is it necessary to enforce legal tender laws that require acceptance of the national currency. For example, rather than engraving the statement 'This note is legal tender for all debts, public and private', all the US government needs to do is to promise 'This note will be accepted in the payment of taxes' in order to ensure general acceptability within the US and even abroad.

### **Financial stocks and flows are denominated in the national money of account**

Financial stocks and financial flows are denominated in the national money of account. While working, the employee earns a flow of wages that are denominated in money, effectively accumulating a monetary claim on the employer (see Chapter 5). On payday, the employer eliminates the obligation by providing a pay cheque that is a liability of the employer's bank.

Again, that is denominated in the national money of account. If desired, the worker can cash the cheque at their bank, receiving the government's currency - again an IOU of the government. Alternatively, the cheque can be deposited in the worker's bank, leaving the worker with an IOU of their bank, denominated in the money of account.

Any disposable income that is not used for consumption purchases represents a flow of saving, accumulated as a stock of wealth. In this case, the saving is held as a bank deposit, that is, as financial wealth. These monetary stocks and flows are conceptually nothing more than accounting entries, measured in the money of account. We can easily imagine doing away with coins and paper notes as well as cheque books, with all payments made through electronic entries on computer hard-drives. All financial wealth could similarly be accounted for without use of paper.

In Chapter 5, we carefully examined the definitions of stocks (for example, wealth) and flows (for example, income, spending and saving), as well as the relationships between them.

### **The financial system as an electronic scoreboard**

The modern financial system can be seen as an elaborate system of record keeping, a sort of financial scoring of the game of life in a capitalist economy. Financial scoring can be compared with a scoreboard at a sporting event, say a game of football. When a team scores a goal, the official scorer awards points, and electronic pulses are sent to the appropriate combination of LEDs so that the scoreboard will show the appropriate number of points depending on the football code being played. As the game progresses, point totals are adjusted for each team. The points have no real physical presence, they simply reflect a record of the performance of each team according to the rules of the game. They are not 'backed' by anything, although they are valuable because the team that accumulates the most points is deemed the 'winner' - perhaps rewarded with fame and fortune. Further, in some codes, points can be taken away after review by officials who determine that rules were broken and that penalties should be assessed. The points that are taken away don't really go anywhere - they simply disappear as the scorekeeper deducts them from the score.

Similarly, in the game of life, earned income leads to 'points' credited to the 'score' that is kept by financial institutions. Unlike the game of football, in the game of life, every 'point' that is awarded to one player is deducted from the 'score' of another - either reducing the payer's assets or increasing their liabilities. However, accountants in the game of life are very careful to ensure that financial accounts always balance. The payment of wages leads to a debit of the employer's 'score' at the bank, and a credit to the employee's 'score', but at the same time, the wage payment eliminates the employer's implicit obligation to pay wages as well as the employee's legal claim to wages. So, while the game of life is a bit more complicated than the football game, the idea that record keeping in terms of money is a lot like record keeping in terms of points can help us to remember that money is not a 'thing' but rather is a unit of account in which we keep track of all the debits and credits - or, 'points'.



### 6.3 Floating versus Fixed Exchange Rate Regimes

In the previous sections we dealt with the case of governments that do not promise to convert their currencies on demand into precious metals or anything else. When a \$5 note is presented to the US Treasury, it can be used to pay taxes or it can be exchanged for five \$1 notes (or for some combination of notes and coins that total \$5) - but the US government will not convert it to anything else. Further, the US government does not promise to maintain the exchange rate of US dollars at any particular level. Most of this textbook will be concerned with sovereign currencies which operate with **floating exchange rates** against other currencies, so that they are not convertible at a fixed rate to another currency. Examples of such currencies include the US dollar, the Australian dollar, the Canadian dollar, the UK pound, the Japanese yen, the Turkish lira, the Mexican peso, the Argentinean peso, and so on. We will now make this important distinction between fixed and floating exchange rates clearer.

#### The gold standard and fixed exchange rates

A century ago, many nations operated with a gold standard in which the country not only promised to redeem its currency for gold, but also promised to make this redemption at a fixed exchange rate. An example of a fixed exchange rate is a promise to convert thirty-five US dollars to one ounce of gold. For many years, this was indeed the official US exchange rate. Other nations also adopted fixed exchange rates, pegging the value of their currency either to gold or, after WWII, to the US dollar. For example, at the inception of the post WWII system, known as the Bretton Woods system, the official exchange rate for the UK pound per US dollar was 0.2481 (on December 27, 1945). This is equivalent to a person receiving \$US4 for each UK pound presented for conversion. As all other currencies in the system were set relative to the US dollar, this also set their relative values with each other. So on December 27, 1945, 119.1 French francs exchanged for \$US1, which meant that it that 480 francs were required to purchase one UK pound. In Chapter 16, we will learn how to interpret exchange rate quotations and calculate various cross parities.

In order to make good on its promises to convert its currency at fixed exchange rates, each nation had to keep a reserve of foreign currencies (and/or gold). For example, if a lot of UK pounds were presented for conversion to \$US (for example, by foreign central banks to the Bank of England), the UK's reserves of foreign currency could be depleted rapidly. There were three strategies that could be adopted by the UK government to avoid running out of foreign currency reserves, but none of them was very pleasant. They included: a) alter the value of the pound against the US dollar – that is, devalue; b) borrow foreign currency reserves; or c) deflate the economy using higher interest rates and/or fiscal cutbacks to curtail imports and attract capital inflow.

#### Floating exchange rates

In August 1971, the US President Nixon abandoned US participation in the fixed exchange rate system because it was unable to continue to guarantee conversion of US dollars into gold at the agreed price. Many countries followed suit. This meant that these governments no longer promised to convert their currency to another currency (or gold) at a fixed rate. As a result, the relative values of currencies against each other were allowed to float and be determined hour by hour by forces of demand and supply. It didn't stop conversion of currencies into other

currencies. It just meant that the values governing that conversion would frequently fluctuate. It is easy to convert most currencies into any other major currency at private banks and at kiosks in international airports. Currency exchanges do these conversions at the current exchange rate set in international markets (less fees charged for the transactions). These exchange rates change day-by-day, or even minute-by-minute, fluctuating to match demand (from those trying to obtain the currency in question) and supply (from those offering that particular currency in exchange for other currencies).

The determination of exchange rates in a floating exchange rate system is exceedingly complex. The international value of the US dollar, for example, might be influenced by such factors as the demand for US assets, the US trade balance, US interest rates relative to those in the rest of the world, US inflation, and US growth relative to that in the rest of the world. So many factors are involved that no statistical model has been developed yet that can reliably predict movements of exchange rates.

What is important for our analysis, however, is that with a floating exchange rate, a government does not need to fear that it will run out of foreign currency reserves (or gold reserves) for the simple reason that it does not convert its domestic currency to foreign currency at a fixed exchange rate. Indeed, the government does not have to promise to make any conversions at all. In practice, governments operating with floating exchange rates do hold foreign currency reserves, and they do offer currency exchange services for the convenience of their financial institutions. However, the conversions are done at current market exchange rates, rather than keeping the exchange rate at a prescribed level.

Governments can also intervene into currency exchange markets to try to nudge the exchange rate in the desired direction. They also will use macroeconomic policy (including monetary and fiscal policy - as discussed later) in an attempt to affect exchange rates. Sometimes this works, and sometimes it does not. The point is that, with a floating exchange rate, attempts to influence exchange rates are discretionary. With a fixed exchange rate, government must use policy to try to keep the exchange rate fixed. On the other hand, the floating exchange rate ensures that the government has greater freedom to pursue other policy goals - such as maintenance of full employment, sufficient economic growth, and price stability. How it might do that is discussed in later chapters.

#### **6.4 IOUs Denominated in National Currency: Government and Non-Government**

In the sections above we have noted that assets and liabilities are denominated in a money of account, which is chosen by a national government and given force through the mechanism of taxation. With a floating exchange rate, the government's own IOUs - currency - are nonconvertible in the sense that the government makes no promise to convert them to precious metal, to foreign currency, or to anything else. Instead, it promises to accept its own IOUs in payments made to itself (mostly tax payments, but also payments of fees and fines). This is the necessary and fundamental promise made: the issuer of an IOU must accept that IOU in payment. So long as government agrees to accept its own IOUs in tax payments, the government's IOUs will be in demand (at least for tax payments, and probably for other uses as well).

Similarly, private issuers of IOUs also promise to accept their own liabilities. For example, if you have a loan with your bank, you can always pay principle and interest on the loan by writing a cheque on your deposit account at the bank. Indeed, all modern banking systems operate a cheque clearing facility so that each bank accepts cheques drawn on all other banks in the country. This allows anyone with a debt due to any bank in the country to present a cheque drawn on any other bank in the country for payment of the debt. The cheque clearing facility then operates to settle accounts among the banks - a topic to be discussed in detail in Chapter 10. The important point is that banks accept their own liabilities (cheques drawn on deposits) in payments on debts due to banks (the loans banks have made), just as governments accept their own liabilities (currency) in payments on debts due to government (tax liabilities).

### **Leveraging**

There is one big difference between government and banks, however. Banks do promise to convert their liabilities to something. You can present a cheque to your bank for payment in currency, what is normally called 'cashing a cheque', or you can simply withdraw cash at the Automatic Teller Machine (ATM) from one of your bank accounts. In either case, the bank IOU is converted to a government IOU. Banks normally promise to make these conversions either 'on demand' (in the case of 'demand deposits', which are normal cheque accounts) or after a specified time period (in the case of 'time or term deposits', including savings accounts and certificates of deposits, known as CDs - perhaps with a penalty for early withdrawal).

Because banks make this promise to convert on demand, they must either hold reserves of currency, or have quick access to them. Their reserves take the form of vault cash plus deposits held at the central bank. Note that they need to hold only small amount of reserves against their deposits because they know that redemptions (withdrawals) over any short period will be a tiny fraction of their total deposits. The fraction of reserves against deposits is called the reserve ratio. We can think of deposits as **leveraging** the reserves. For example, in the USA, the ratio of reserves against bank deposits is around 1 per cent. This means the leverage ratio is 100-to-1.

Banks hold a relatively small amount of currency in their vaults to handle these conversions, but most of their reserves take the form of deposits at the central bank. If they need more currency, they ask the central bank to send an armoured truck with the desired notes and coins. Banks don't want to keep a lot of cash on hand, nor do they need to do so in normal circumstances. For our purposes here, bank reserves (deposits at the central bank) are equivalent to vault cash because a bank can immediately convert them to currency to meet cash withdrawals. There is no functional difference between cash held in bank vaults and reserve deposits held at the central bank. We can include both as currency – government liabilities with zero time to maturity.

Lots of cash could increase the attractiveness to robbers, but the main reason for minimising holdings is because it is costly to hold currency. The most obvious cost is the vault and the need to hire security guards. However, more important to banks is that holding reserves does not earn profits. Banks would rather hold loans as assets, because debtors pay interest on these loans. For this reason, banks operate with high leverage ratios, holding a very tiny fraction of their assets in the form of reserves against their deposit liabilities. So long as only a small percentage of their depositors try to convert deposits to cash on any given day, this is not a problem. However, in the case of a bank run in which a large number of customers try to convert their deposits to cash on the same day, the bank will have to obtain currency from the central bank. This can even lead

to a lender of last resort action by the central bank that lends currency reserves to a bank facing a run. These are issues that we will address later.

### **Clearing accounts extinguish IOUs**

There is another reason that banks hold reserves. When you write a cheque on your bank account to pay a bill, the recipient of the cheque will deposit it in their own bank - which is probably a different bank. Their bank will present the cheque to your bank for payment. This is called **clearing accounts**. Banks clear accounts using government IOUs, and for that reason banks maintain reserve deposits at the central bank. More importantly, they have access to more reserves should they ever need them, both through borrowing from other banks through the interbank market for reserves (an overnight market where banks lend to and borrow from each other), or through borrowing them from the central bank. All modern financial systems have developed procedures that ensure banks can get currency and reserves as necessary to clear accounts among themselves and with their depositors. The central bank is duty bound to provide banks with sufficient reserves should it fall short on any particular day.

When First National Bank receives a cheque drawn on Second National Bank, it asks the central bank to debit the reserves of Second National and to credit its own reserves. This is now handled electronically. Note that while Second National's assets will be reduced (by the amount of reserves debited), its liabilities (cheque deposit) will be reduced by the same amount. Similarly, when a depositor uses the ATM to withdraw currency, the bank's assets (cash reserves) are reduced, and its IOUs to the depositor (the liabilities in the deposit account) are reduced by the same amount.

Other business firms use bank liabilities for clearing their own accounts. For example, the retail firm typically receives products from wholesalers on the basis of a promise to pay after a specified time period (for example, this period is usually 30 days in the US). Wholesalers hold these IOUs until the end of the period, at which time the retailers pay by a cheque drawn on their bank account (or, increasingly, by an electronic transfer from their account to the account of the wholesaler). At this point, the retailer's IOUs held by the wholesalers are cancelled.

Alternatively, the wholesaler might not be willing to wait until the end of the period for payment. In this case, the wholesaler can sell the retailer's IOUs at a discount (for less than the amount that the retailer promises to pay at the end of the period). The discount is effectively interest that the wholesaler is willing to pay to get the funds earlier than promised. In this case, the retailer will finally pay the holder of these IOUs at the end of the period, who effectively earns interest (the difference between the amount paid for the IOUs and the amount paid by the retailer to extinguish the IOUs). Again, the retailer's IOU is cancelled by delivering a bank liability (the holder of the retailer's IOU receives a credit to their own bank account). As we will see later, discounting is the basis of both commercial banking and of interest rates.

### **Pyramiding currency**

This brings up another important point. Private financial liabilities are not only denominated in the government's money of account, but they also are, ultimately, convertible into the government's currency. As we have discussed, banks explicitly promise to convert their liabilities to currency (either immediately in the case of demand deposits, or with some delay in the case of time deposits). Other private firms mostly use bank liabilities to clear their own

accounts. Essentially, this means they are promising to convert their liabilities to bank liabilities, 'paying by cheque' on a specified date (or, according to other conditions specified in the contract). For this reason, they must have deposits, or have access to deposits, with banks to make the payments.

Things can get even more complex than this, because there is a wide range of financial institutions (and, even, non-financial institutions that offer financial services) that can provide payment services. These can make payments for other firms, with net clearing among these 'non-bank financial institutions' occurring using the liabilities of banks. Banks in turn, clear accounts using government liabilities. There could thus be 'six degrees of separation' (many layers of financial leveraging) between a creditor and debtor involved in clearing accounts.

We can think of a pyramid of liabilities, with different layers according to the degree of separation from the central bank. Perhaps the bottom layer consists of the IOUs of households, held by other households, by firms engaged in production, by banks, and by other financial institutions. The important point is that households usually clear accounts by using liabilities issued by those higher in the debt pyramid - usually financial institutions.

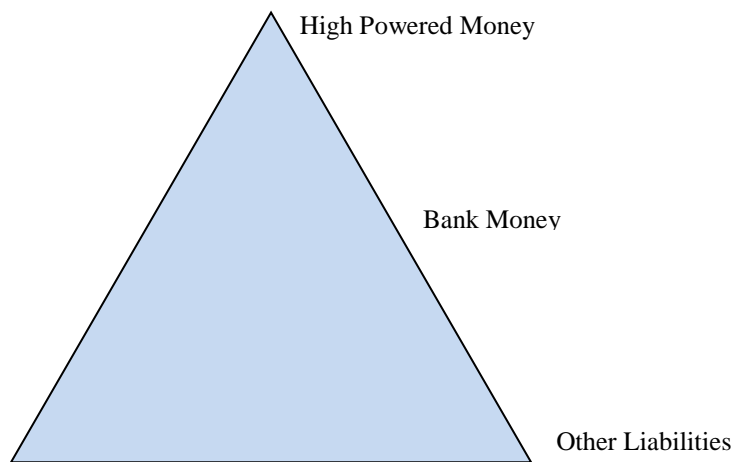
The next layer up from the bottom consists of the IOUs of firms engaged in production, with their liabilities held mostly by financial institutions higher in the debt pyramid (although some are directly held by households and by other production firms), and who mostly clear accounts using liabilities issued by the financial institutions, sometimes called shadow banks.

At the next layer we have non-bank financial institutions, which in turn clear accounts using the banks whose liabilities are higher in the pyramid. Banks use government liabilities for net clearing.

Finally, the government is at the top of the pyramid - with no liabilities higher than its non-convertible IOUs. The shape of the pyramid is instructive for two reasons. First, there is a hierarchical arrangement whereby liabilities issued by those higher in the pyramid are more generally more acceptable. In some respects, this is due to higher credit worthiness (the government's liabilities are free from credit risk; as we move down the pyramid through bank liabilities, toward non-financial business liabilities and finally to the IOUs of households, risk tends to rise - although this is not a firm and fast rule). Second, the liabilities at each level typically leverage the liabilities at the higher levels. In this sense, the whole pyramid is based on leveraging of (a relatively smaller number of) government IOUs. This is a concept we will return to in the next section.

The following 'pyramid' (developed by Hyman Minsky and Duncan Foley, and extended by Stephanie Bell) provides a nice visual representation of the concept of leveraging. At the top of the pyramid are the government's liabilities; below this are the liabilities of banks, normally made convertible into government's high powered money, which is also called the monetary base and constitutes the sum of all bank reserves held in the central bank clearing accounts and outstanding currency (notes and coins). At the bottom of the pyramid we include all other money-denominated liabilities (these could include the IOUs of non-financial firms as well as those of households).

**Figure 6.1 The Minsky – Foley pyramid**



## 6.5 Use of the Term ‘Money’: Confusion and Precision

Before concluding this chapter, we will briefly distinguish between our use of the term ‘money’ and the way this term is often used. The term ‘money’ is often used colloquially to refer to income, as in ‘how much money do you make at your job’. As was discussed in Chapter 5, income is a flow that is measured in nominal terms, that is, in the money of account. In this book, we will always carefully distinguish flows from stocks, and will not use the term ‘money’ in place of ‘income’.

The term ‘money’ is also often used to indicate a particular liability, such as the demand deposit liability of a bank, or the currency IOU of the government. In fact, as we have discussed above, all financial liabilities are denominated in a money of account. It is thus rather arbitrary to call some of these ‘money’ and to exclude others. Further, each time one uses the term money to generally refer to money-denominated liabilities, one must provide a list of those that are included as ‘money’ or a list of those that are excluded. Otherwise, we can never be sure what the speaker means.

Throughout this book, we will carefully distinguish between the money of account (the US dollar or the Australian dollar, for example), and specific money-denominated liabilities (demand deposits issued by banks or currency issued by the government, for example). The term ‘money’ simply refers to the unit of account chosen by government to denominate tax liabilities and payments made to government - the dollar in both the US and Australia. As we have discussed, this does not have any physical existence but rather is the unit in which we can keep track of debts and credits - much as a ‘point’ is the unit of account used in American football to keep track of touchdowns and field goals. Just as a touchdown is denominated in points, a coin is denominated in dollars (or fractions of a dollar). A touchdown takes a physical form (a player carrying the football crosses the goal line), but the six points used to ‘account’ for the touchdown do not have any physical presence. In the same manner, a ten dollar note issued by the treasury has a physical form (a piece of paper imprinted with ink), but the ten dollars owed by the treasury that it ‘accounts’ for do not.

# Chapter 7: The Real Expenditure Model

## Chapter Outline

- 7.1 Introduction
- 7.2 Aggregate Supply
- 7.3 Aggregate Demand
- 7.4 Private Consumption Expenditure
- 7.5 Private Investment
- 7.6 Government Spending
- 7.7 Net Exports
- 7.8 Total Aggregate Expenditure
- 7.9 Equilibrium National Income
  - Special Topic: Inventory movements and planned investment
- 7.10 The Expenditure Multiplier
  - An algebraic treatment
  - A graphical treatment
  - Numerical example of the expenditure multiplier at work
  - Changes in the magnitude of the expenditure multiplier

## Learning Objectives

1. Acknowledge the distinction between identities and behavioural relationships.
2. Understand the components of aggregate demand and their determinants.
3. Analyse how the components of aggregate demand interact to determine equilibrium national income
4. Understand why the multiplier effect typically exceeds unity.

## 7.1 Introduction

In Chapter 4, we outlined the measurement of GDP via the expenditure approach in the National Accounts (NIPA), which underlines the fundamental premise in macroeconomics that it is total spending that drives output (GDP) and employment in the economy. However National Accounts principles are founded on definitions which give rise to an identity equating total spending to GDP (National Income).

In this Chapter, we shall develop a model of output (GDP) determination, which will provide us with an understanding as to how the different components of expenditure interact and determine total output. This means we must outline the key variables which influence the components of total spending that is develop behavioural relationships. We will consider total expenditure in **real** terms and relate that to a simple model of aggregate supply.

There are two aspects to conceiving of the income-expenditure framework in real terms. First we consider that when they spend, consumers and governments, desire to achieve real outcomes in terms of the command on real resources (output). Thus a household spending, say \$100 on consumption goods is making a decision to purchase a real quantity of goods and services.

Second we must formulate the supply-side of the economy in real terms too, so that it is consistent with the demand side. We can do that by using the National Accounting concept of GDP at constant rather than current prices as our measure of economic activity.

Thus, we abstract from price changes and assume that firms react to changes in aggregate spending by adjusting the quantity of output rather than the price and quantity. Firms are thus assumed to respond dollar-for-dollar to increased demand by increasing output and income. There are various ways in which we can justify considering firms to be **quantity adjusters**.

First, firms use mark-up pricing principles, whereby they add a profit mark-up to their unit costs and face roughly constant unit costs over the output range within which they normally produce. Typically they maintain some excess capacity and can thus increase output relatively easily without further investment in productive capacity, which would take time. If they face insufficient capacity relative to demand, firms are likely to raise prices to ration demand which inevitably leads to the loss of customers to competitors.

Second, firms face various costs when adjusting prices and as a result only periodically make such adjustments. It has been said that firms use 'catalogue pricing', whereby they make their prices known to their prospective customers through advertising and other means and then are prepared to sell goods and services at those prices irrespective of demand (up to their full productive capacity). At the end of the current catalogue period, they will then make any necessary adjustments to prices based on expected future demand and any recent and expected movements in unit costs. We consider mark-up pricing models in Chapter 8 when we develop the detailed aggregate supply framework.

In this chapter we will arrive at three important propositions about the macroeconomy:

1. Total spending drives total output (GDP) and employment in the economy;
2. There is no guarantee that equilibrium output will be associated with full employment;  
and



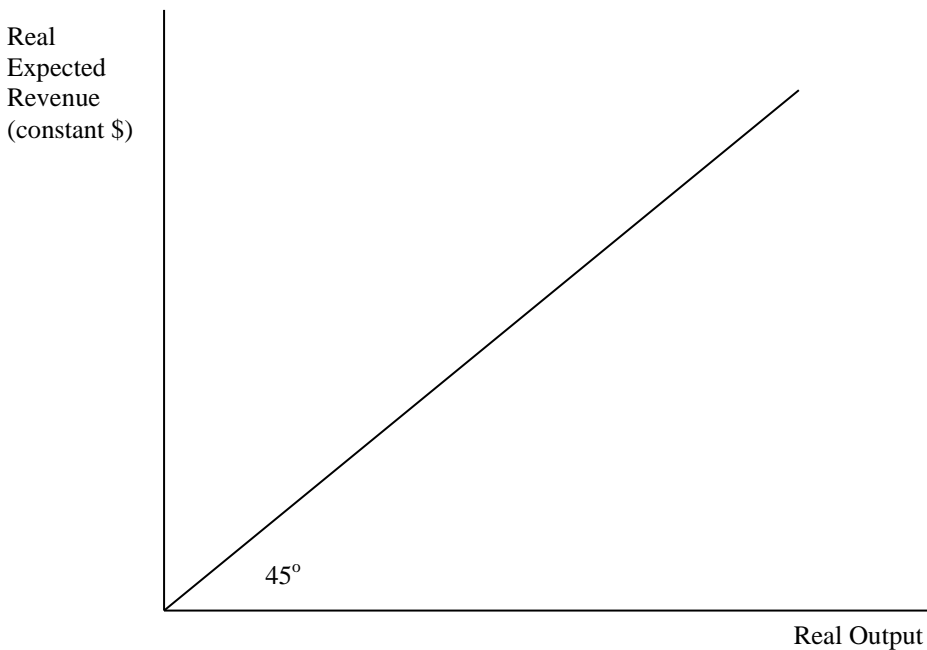
3. A change in autonomous expenditure, such as investment or government expenditure, will lead to a multiplied (larger) change in real GDP (National Income).

## 7.2 Aggregate Supply

Figure 7.1 depicts the simplified constant price aggregate supply relationship that we will work with in this Chapter to concentrate on the way the economy adjusts to changes in aggregate demand. It is drawn as a 45° line emanating from the origin with total expected revenue (in constant dollars) on the vertical axis and real output on the horizontal axis.

Obviously, without considering any economic meaning, the 45° line shows all points where real income (real expected proceeds) is equal to real output (Aggregate Supply). We also measured planned real expenditure (Aggregate Demand) on the vertical axis, so the 45° line shows all points where output and (expected) expenditure are equal (in constant prices).

**Figure 7.1** Aggregate supply



In Chapter 4 we explained the different perspectives that we can take in measuring aggregate economic activity. The expenditure, income and output approaches provide different views of the national accounting framework but these approaches yield the same aggregate outcome. The total value of goods and services produced in any period is equal to the total spending and the total income generated (wages, profits, rent and profit) in that same period.

As we will see in Chapter 8 when we consider aggregate supply in more detail, firms supply a particular level of output (and incur costs of production) as long as it can generate enough revenue to cover the costs and realise their desired profits.

Note that here, as opposed to the accounting identities in Chapter 4, the vertical axis shows *expected* real revenue, to be generated by employing workers and producing output. In other words, it is forward looking. However, by producing output, incomes will be generated, most of

which will be spent. The 45-degree line shows us the points where the revenues generated are equal to the expenditures - which of course must be true because spending must generate revenue.

The vertical axis provides a measure of total revenue or expenditure, which also equals to real income. From the perspective of the firms, the vertical axis tells them the real expected proceeds that can be generated by selling the different levels of output.

The other point to note is that we have imposed no full capacity point on the graph. At some point, when the economy is operating at full capacity, firms are unable to continue expanding real output in response to additional spending. When we formally introduce the expenditure side of the economy in the next section we will also impose a full employment output condition beyond which firms cease to be quantity adjusters.

### 7.3 Aggregate Demand

In Chapter 4 we learned that firms generate additional productive capacity through new investment in order to produce additional goods and services to satisfy demand. Here, for simplicity, we are assuming no depreciation of the existing stock of productive capacity. Once the capital stock is in place, firms will respond to increases in spending for the goods and services they supply by increasing output up to the productive limits of their capital and the available labour and other inputs. Beyond full capacity, they can only increase prices when increased spending occurs. We assume that potential output is fixed for the period we are analysing.

By adopting the assumption in this chapter that prices do not adjust to changes in demand, we thus say that higher aggregate demand will lead to increased production, which in turn increases national income.

**The basic macroeconomic rule then is that, subject to the existing productive capacity, total spending drives output and national income, which, in turn, drives employment.**

In Chapter 4 we introduced the National Accounting framework used by national statistical agencies and learned that total expenditure in any period is expressed as the sum of the following sources of spending:

- Consumption by households or persons ( $C$ ).
- Investment spending by firms ( $I$ ).
- Government spending ( $G$ ).
- Export spending by foreigners ( $X$ ) minus import spending by domestic residents ( $M$ ), which we denote as net exports ( $NX$ ) = ( $X - M$ ).

The sum of these expenditures equals GDP as a matter of accounting. Total expenditure sums to total output and total income.

From the National Accounting framework, we know that total expenditure ( $E$ ) in the domestic economy in any particular period can be expressed as:

$$(7.1) \quad E = C + I + G + (X - M)$$

You will see that while exports add to total spending in the domestic economy, imports lead to a drain in spending because it represents the spending of local residents, firms and governments on goods and services produced by other nations.

The equilibrium level of real national income ( $Y$ ) is determined by aggregate demand.

$$(7.2) \quad Y = E$$

You should note that the level of real GDP that is produced by the current period's expenditure does not necessarily have to equal the full employment output level. Keynes among others demonstrated that full employment was not guaranteed by the market system.

We consider those issues in detail in Chapter 9 when we study the labour market.

In the remainder of this Chapter we will develop a more detailed understanding of the behaviour of each of these components of total spending and explain how they interact to determine total output (GDP) and national income. We will also derive an expression for equilibrium national income.

#### **7.4 Private Consumption Expenditure**

To gain an understanding of the determinants of aggregate demand, we have to focus on private sector decision-making, which occurs within a broad set of constraints resulting from the interaction between government and non-government.

Private consumption spending is the largest component of total spending on GDP in most economies. Consumption is the sum of household spending on non-durable goods (for example, food), durable goods that provide benefits beyond a single year (for example, cars and white-goods like refrigerators) and services (for example, restaurants, theatres and the like).

Table 7.1 shows the ratio of private consumption expenditure to total GDP for most of the OECD nations. While there are notable exceptions, the outcomes in most nations are close to the OECD average of 60.7 per cent. The ratios across countries are also relatively stable over time.

**Table 7.1 Consumption ratios, OECD nations, 2005-2010, per cent**

	<b>2005</b>	<b>2010</b>
Australia	54.6	52.4
Austria	53.2	52.7
Belgium	49.7	51.0
Canada	51.7	56.3
Chile	54.4	56.0
Czech Republic	410.6	50.9
Denmark	46.8	46.4
Estonia	50.4	52.0
Finland	49.7	53.0
France	53.6	51.0
Germany	55.8	56.4
Greece	66.3	74.9
Hungary	52.8	53.6
Iceland	54.4	52.8
Ireland	42.7	49.9
Italy	56.6	60.6
Japan	54.6	56.5
Korea	49.8	50.4
Luxembourg	34.8	32.9
Mexico	60.8	62.1
Netherlands	46.6	44.5
New Zealand	56.5	51.0
Norway	39.9	41.2
Poland	61.1	58.1
Portugal	59.2	63.9
Slovak Republic	52.0	59.0
Slovenia	52.1	56.1
Spain	54.0	56.6
Sweden	410.8	46.7
Switzerland	58.3	55.9
Turkey	62.8	65.0
United Kingdom	62.5	61.5
United States	65.8	68.3
OECD - Total	59.0	60.7

Source: OECD Statistics

What determines total private consumption expenditure?

The most elementary theory of private consumption ( $C$ ) is that it is a stable function of disposable national income ( $Y_d$ ). Disposable income is the flow of income that remains after taxes have been paid.

We define disposable income as:

$$(7.3) \quad Y_d = Y - T$$

A simple model of the government's tax policy is that it levies a proportional tax rate ( $t$ ) from total national income, which means that the total tax revenue ( $T$ ) at any level of income is given as:

$$(7.4) \quad T = tY$$

Assume that the proportional tax rate ( $t$ ) is 0.20. This means that for every dollar of national income generated in the economy the government takes 20 cents out in the form of taxation. The 80 cents, which remains, is disposable income.

We consider taxation to be a 'leakage' from the expenditure system because it is income that is not available for private spending.

Disposable income can be written as:

$$(7.5) \quad Y_d = Y - T = Y - tY = (1-t)Y$$

In our specific example, this would be written as  $Y_d = (1 - 0.2)Y = 0.8Y$ .

In macroeconomics, the **aggregate consumption function** is defined as a relationship between total consumption ( $C$ ) and total disposable income ( $Y$ ).

$$(7.6) \quad C = C_0 + cY_d$$

$C_0$  is a constant and is a base level of consumption which is independent of disposable income. The coefficient  $c$  is called the marginal propensity to consume (MPC) and measures the fraction of every additional dollar of disposable income that is consumed.

The MPC is generally presumed to have a value between 0 and 1. If, for example,  $c = 0.75$  then for every extra dollar of disposable income that the economy generates, consumption would rise by 75 cents.

It is important to understand that the MPC in this model is an aggregate, which is an average of all the individual household consumption propensities. Lower income households tend to have MPC values close to 1 whereas the higher income households have much lower than average consumption propensities.

This arises because lower income families find it harder to purchase enough goods and services to maintain basic survival given their income levels. Higher income earners not only consume more in absolute terms but also have more free income after they have purchased all the basic essentials.

As we will learn later in this chapter, the distribution of income is an important consideration when seeking to understand changes in aggregate demand. For example, a change in tax policy that increased disposable income for low-income consumers would have a greater positive

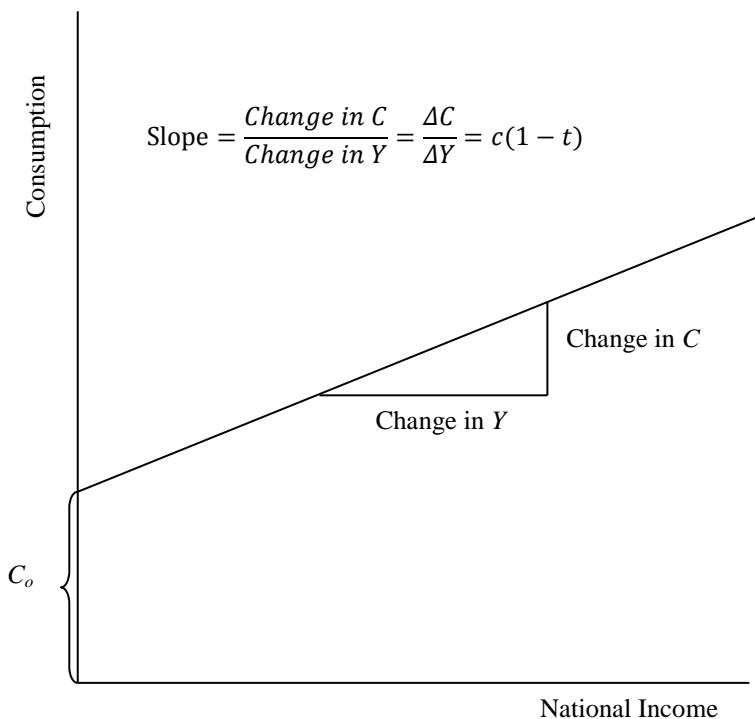
impact on final consumption than a tax cut aimed at giving high-income earners the same absolute increase in disposable income.

By substituting the expression for disposable income in Equation (7.5) into the consumption Equation (7.6), we can show the direct influence of the tax rate on private consumption:

$$(7.7) \quad C = C_0 + cY_d = C_0 + c(I - t)Y$$

Figure 7.2 shows the consumption function in graphical terms. Note here and in the graphs to follow that the vertical axis is specified in terms of actual rather than expected expenditure (or revenue from the perspective of firms), which described the vertical axis in Figure 7.1. For simplicity, we will assume that expected and actual revenue is equal, although in the real world that is not likely to be the case. Later we will consider what happens when expectations are not met.

**Figure 7.2** The consumption function



The consumption function cuts the vertical axis at  $C_0 > 0$ . The consumption function is then upward sloping because we have postulated that consumption rises with national income.

In the Appendix of the book, entitled *Methods, Tools and Techniques*, we learn how to derive a slope graphically. The slope of a line is the ratio RISE over RUN. Rise in this case is the change in consumption spending ( $\Delta C$ ) and run is the change in national income ( $\Delta Y$ ) and we have drawn a little triangle underneath the consumption function to illustrate this.

In fact,  $\Delta C = c(I-t)\Delta Y$  and RISE over RUN =  $\Delta C/\Delta Y = c(I-t)\Delta Y/\Delta Y = c(I-t)$ . The slope of the consumption function is thus given by the coefficient  $c(I-t)$ , which is lower than the MPC because a \$1 rise in national income translates into less than a \$1 rise in disposable income given that the marginal tax rate ( $t$ ) is positive.

You should be able to work out what would happen if the MPC ( $c$ ) increased. The result would be an increase in the slope of the consumption function such that at every level of disposable income, total consumption would be higher.

What determines aggregate saving ( $S$ )?

$$(7.8) \quad S = Y_d - C = Y - T - C$$

Saving at the macroeconomic level is thus a **residual** that remains after households have made their spending decisions.

Given the MPC out of disposable income which takes the value ( $c$ ), we can define a related concept – the marginal propensity to save ( $s$ ) which is just the 1 minus the marginal propensity to consume ( $c$ ):  $s = 1 - c$ . Note that since disposable income can only be consumed or saved,  $I = s + c$  or:  $MPS + MPC = 1$ .

When national income rises, the government takes out some taxes leaving an increase in disposable income, which is then the source of increased consumption (via the marginal propensity to consume) with the remainder of the increase in disposable income being saved.

If the MPC is 0.75 and the tax rate is 0.2, then if national income increases by \$100, total tax revenue rises by \$20, so disposable income rises by \$80, consumption then rises by \$60 and \$20 is saved.

## 7.5 Private Investment

When a macroeconomist uses the term investment they are referring to a very specific type of spending, which does not accord with the common usage of the term. For example, a lay person might think of investment as a person putting some money in a fixed-term deposit at a bank or the purchase of some shares in a company.

The National Accounting meaning of investment is any spending that adds to the productive capacity of the economy, that is adds to the capital stock. Capital in this context is productive plant and equipment or other capacity, which defines the potential output of an economy. Thus, when a firm builds a new factory or purchases a new piece of machinery, they are considered to be investing.

Generally, firms invest but households consume and save. The one major exception is residential real estate investment, which is included as investment. But here again, the economist's definition differs somewhat from the lay-person's, because only *newly constructed* housing counts. Purchases of existing ('used') housing do not count as investment for the purposes of GDP accounting because it is not newly produced.

Changes to the stock of inventories, which are unsold goods, are also considered to be a component of business investment in each period because they add to the potential of the economy to meet current aggregate demand for goods and services. As we will see, the dynamics of inventories provide important information about the state of the business cycle.

Table 7.2 shows a snapshot of the Australian National Accounts for the June quarter 2012. The organisation of the expenditure components is based on the standard National Accounts framework that we discussed in Chapter 4, which is broadly shared across the world.

Under the heading, Gross fixed capital formation, a number of individual line entries are listed which comprise the separate categories that the statistician uses to estimate total investment spending. The *Change in inventories* is classified as a separate category.

Economists distinguish between gross and net investment. Gross investment is the total spending by firms on new plant and equipment and on inventories. However, in each period the existing capital stock depreciates. For example, machines wear out and/or become obsolete; buildings require maintenance; and car fleets require updating.

Some of the gross investment in each period merely covers the depreciation of the existing capital stock.

Net investment is the component of gross investment that adds **new** productive capacity – that is, which increases the overall capital stock. Net investment is thus gross investment less total depreciation. We assume that there is no depreciation of existing capital stock and no planned changes to the stock of inventories.

## 7.6 Government Spending

We have already introduced one element of the government's interaction with the non-government sector, namely the proportional tax rate ( $t$ ), which is our simplified expression for what is a complex tax structure in real world economies.

The National Accounts framework shows that government spending takes a variety of forms. First, all levels of government purchase a range of goods and services from the non-government sector as a means of fulfilling their social and economic goals. Some of the purchases are for consumption goods and services, while other spending is categorised as public investment or public capital formation. The latter category of spending generates the valuable public infrastructure that enhances the welfare and profitability of the non-government sector.

Second, governments directly employ workers to provide a range of services to the public.

Third, the government provides a range of **transfer payments** to the non-government sector in the form of pension and welfare entitlements and other transfer payments. In Chapter 4, we learned that the National Accounts framework did not include transfers as a component of government spending because they do not constitute a final demand for goods and services by government.

In our models, which are stylisations of the national accounting measures, the flow  $T$  (total taxes) represents **net taxes**. That is, they are total tax revenue minus total transfers to the non-government sector.

Net government spending (that is, government spending minus taxes net of transfers) is determined by two broad forces. First, the decisions that the government takes in setting its fiscal policy (that is, levels of expenditure and tax rate(s)) will be important. Second, the state of the overall economic cycle impacts on net taxes and hence net government spending. For example, when the economy is performing badly, net taxes will fall as a result of both lower taxes and higher welfare payments even without any explicit change in government policy. The opposite will be the case when the economy is growing strongly and unemployment is falling. We call these effects cyclical because they vary with the state of the economic cycle.



For the purposes of the following discussion, we will assume away these cyclical effects on government spending (*G*) and assume that its level is exogenous to national income. We learned about the meaning of exogenous variables in the *Methods, Tool and Techniques* Appendix.

**Table 7.2 Expenditure chain volume measures in national accounts (Australia)**

	<b>% Change June 15 to September 15</b>	<b>% Change September 14 to September 15</b>	<b>% Points contributions to growth in GDP June 15 to September 15</b>
<i>Final consumption expenditure</i>			
General Government	1.0	3.7	0.2
Households	0.6	2.5	0.3
<i>Gross fixed capital formation</i>			
<u>Private</u>			
Dwellings	1.2	8.8	0.1
Ownership transfer costs	0.8	5.0	-
Non-dwelling construction	-2.6	-10.5	-0.2
Machinery and equipment	-3.9	-9.3	-0.2
Cultivated biological resources	1.7	2.1	-
Intellectual products	-1.4	-2.8	-
<u>Public</u>	-1.8	-2.8	-0.1
Changes in inventories	-	-	NA
Gross national expenditure	0.1	0.8	0.1
Exports of goods and services	0.8	5.0	0.2
Imports of goods and services	-0.3	0.5	0.1
Statistical discrepancy (E)	NA	NA	0.2
Gross domestic product	0.6	2.3	0.6

Source: ABS (2015)

## 7.7 Net Exports

In Chapter 16 we will consider the determinants of net exports in detail when we introduce the exchange rate and measures of international competitiveness.

Exports are goods and services produced in the local economy, which are then sold to the rest of the world. The National Accounts includes them in the measure of national production and income because they are produced locally. Export spending is thus an injection of expenditure into the economy from abroad and increases national output and income.

While export spending boosts national income, we consider exports to be a cost in the sense that they deprive the domestic population of the use of the real resources that are used up in the production of the goods and services sold abroad.

Imports comprise expenditure on goods and services by households, firms and government, which are produced by the rest of the world. In other words, some of the consumption expenditure, investment expenditure and government expenditure in each period does not increase local production and is thus considered to be 'lost' because it has 'leaked out' of the domestic expenditure-income loop.

The National Accounts framework thus deducts imports from the estimates of national production and income to recognise that some spending provides demand for goods and services produced by the rest of the world.

By separating the consumption, investment and government spending that leaks out of the economy via imports into a separate category we can clearly appreciate the net effect of external trade. We thus know that the National Account measures of consumption spending ( $C$ ), private capital formation ( $I$ ) and government spending ( $G$ ) represent total expenditure on goods and services, including any imports.

Even though import expenditure is a leakage from the expenditure system, we consider them to provide benefits to the domestic economy by allowing households, firms and government to enjoy access to goods and services not otherwise available or available on competitive (qualitative and/or price) terms.

The difference between exports ( $X$ ) and imports ( $M$ ) is called the **net exports** ( $NX$ ) of a nation. A trade surplus would occur when exports are greater than imports. A trade deficit occurs when the opposite is the case. We will consider these issues in more detail in Chapter 16.

In this chapter we simplify the model by assuming that exports ( $X$ ) are given in any year and determined by national income in the rest of the world, which is outside the influence of the domestic economy.

What determines import spending?

We assume that import spending ( $M$ ) rises with national income. The higher is the national income, the greater will be the flow of imported consumer goods and services and imports of capital equipment.

We will consider the impact of exchange rate changes on import volumes and values in Chapter 16. For now we assume that a nation imports a fixed proportion of every dollar of national income.

That proportion is called the **marginal propensity to import** ( $m$ ) and has a meaning similar to the MPC. The marginal propensity to import is the extra import spending that occurs as a result of a dollar increase in national income.

Our simplified import expenditure model is given as:

$$(7.9) \quad M = mY$$

For example, if  $m = 0.2$ , then if national income ( $Y$ ) rises by \$100, import spending will increase by \$20. The higher is the marginal propensity to import, the higher is the leakage in import spending at every level of national income.

## 7.8 Total Aggregate Expenditure

Now that we have considered all the components of aggregate expenditure (or demand) we can write the equation for aggregate demand ( $E$ ) as:

$$(7.10) \quad E = C + I + G + (X - M)$$

This is an accounting statement, which is derived from the National Accounting framework.

However, we have already developed some simple behavioural theories about the individual components of total expenditure. We assumed that consumption and imports were positive functions of national income and that investment, exports and government spending were determined in each period by factors invariant to national income. We also adopted a simple tax rule.

We can expand the aggregate demand equation to reflect those behavioural assumptions, which, in turn, will allow us to derive an expression for equilibrium income.

$$(7.11) \quad E = C_0 + c(I - t)Y + I + G + X - mY$$

You can see that some components of total expenditure in the economy are dependent on national income levels and some ( $I$ ,  $G$  and  $X$ ) are, by assumption, independent of or autonomous to national income.

We can use the techniques you learned in the *Methods, Tools and Techniques* Appendix to modify Equation (7.11).

$$(7.12) \quad E = C_0 + I + G + X + [c(I - t) - m]Y$$

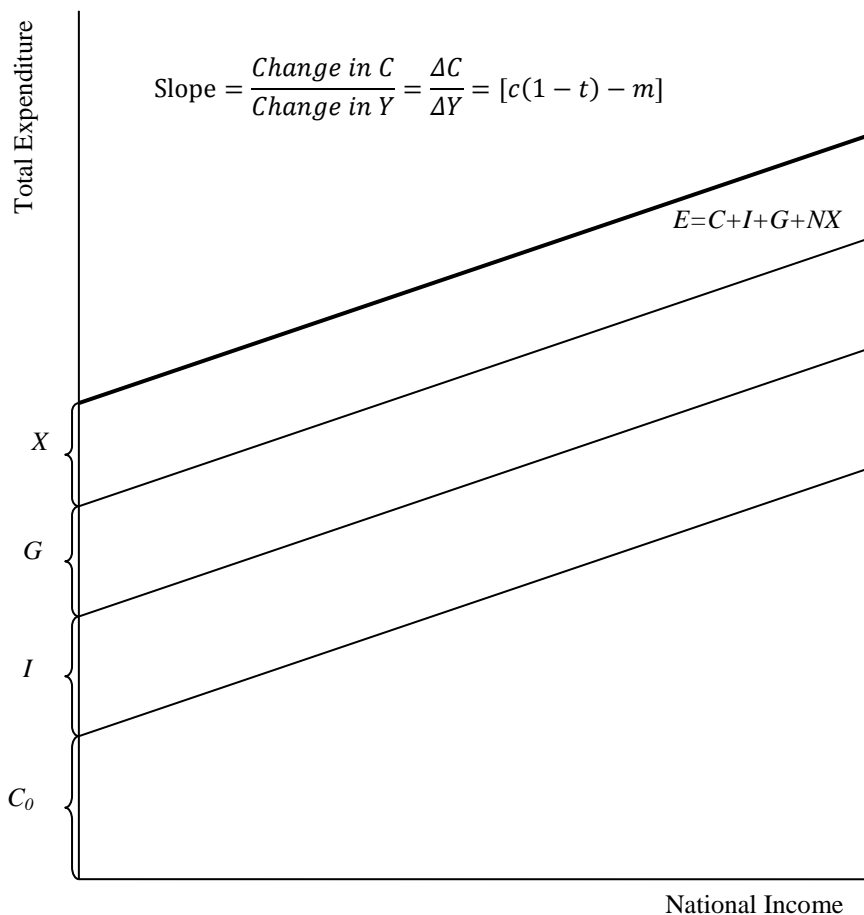
We could usefully simplify this expression by adding together all the components of total expenditure that are autonomous (such that  $A = C_0 + I + G + X$ ) to write the **Aggregate Demand Function** as:

$$(7.13) \quad E = A + [c(I - t) - m]Y$$

The slope of this function, namely  $\Delta E/\Delta Y = [c(I - t) - m]$ , tells us that the change in aggregate spending for a given change in national income is larger, the higher is the marginal propensity to consume ( $c$ ) and the lower are the tax rate ( $t$ ) and the marginal propensity to import ( $m$ ).

See if you can explain this result to yourself at this point. Think about a positive marginal propensity to consume as leading to an induced consumption expenditure increase when national income rises and the tax rate and marginal propensity to import being leakages from the expenditure system for each dollar rise in national income. We will provide a full analysis of this in the next section when we consider the **expenditure multiplier**. Figure 7.3 shows the Aggregate Demand Function and its individual autonomous components.

**Figure 7.3 The aggregate demand function**



The Aggregate Demand Function is drawn with national output/income on the horizontal axis and aggregate demand or planned expenditure on the vertical axis.

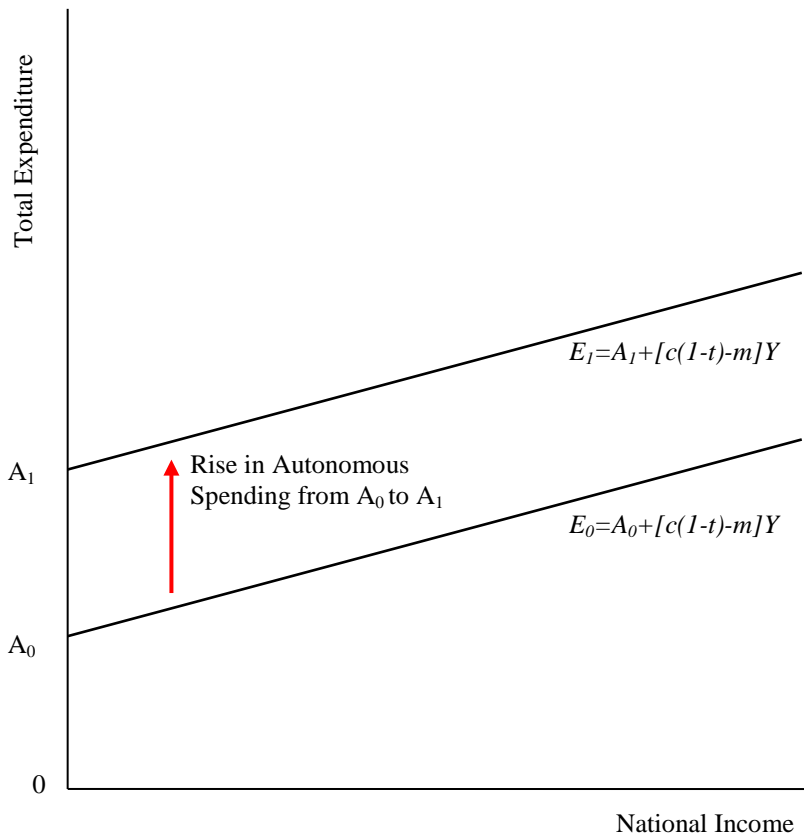
It shows the relationship between national income and planned expenditure and two aspects are important:

- Total autonomous spending,  $A$ , is the horizontal intercept; and
- The slope of the function depends on the marginal propensity to consume ( $c$ ), the marginal propensity to import ( $m$ ) and the tax rate ( $t$ ).

The aggregate demand (planned expenditure) schedule will shift if any of the components that make up autonomous spending change. Also the slope of the schedule will change, if ( $c$ ), ( $t$ ) or ( $m$ ) changes.

Referring to Figure 7.3, if planned government spending or private investment were to rise, then the Aggregate Demand Function would shift upwards – the vertical intercept would rise by the rise in planned spending. As a result, planned spending would be higher at each income level. The opposite would be the case if, for example, government spending was to be cut. In each case the slope of the function would be unchanged. Figure 7.4 shows the impact of a rise in autonomous spending from  $A_0$  to  $A_1$ .

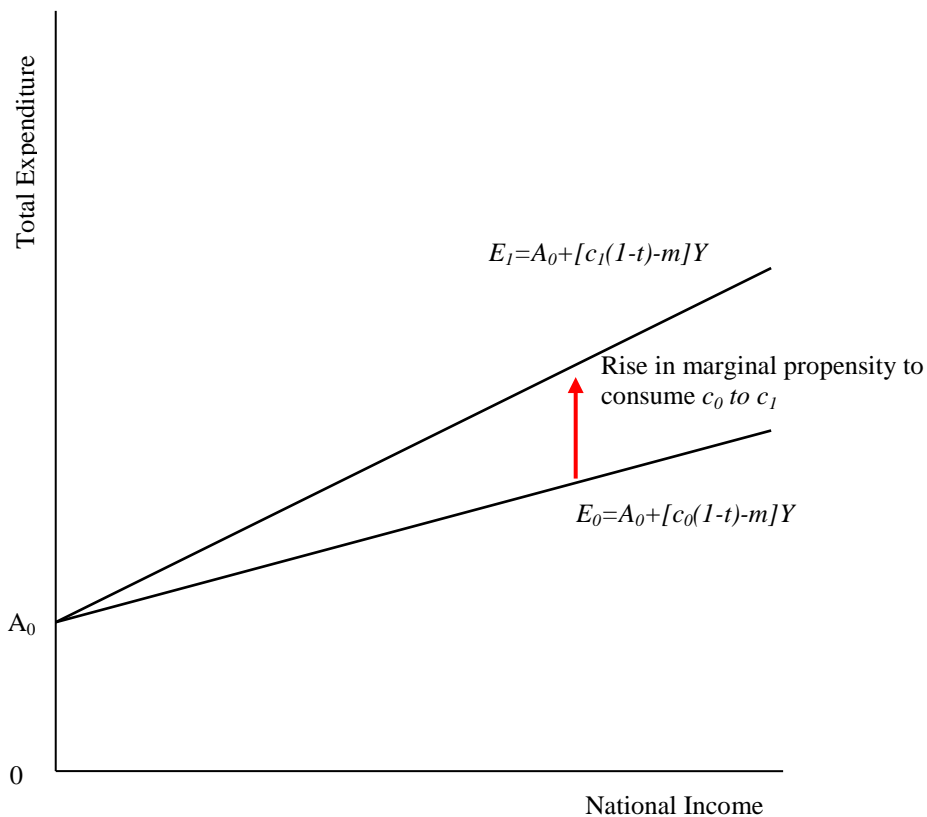
**Figure 7.4** An increase in the intercept of the aggregate demand function



The slope of the Aggregate Demand Function may also change. A rise in the marginal propensity to consume ( $c$ ) and/or a fall in the tax rate ( $t$ ) and/or the marginal propensity to import ( $m$ ) will increase the slope of the function.

Figure 7.5 shows the impact of a rise in the marginal propensity to consume ( $c$ ). At every national income level, total planned spending is now higher.

**Figure 7.5** Changing slope of the aggregate demand function



See if you can sketch the impacts and explain the meaning of a rise in the tax rate ( $t$ ), a rise in the marginal propensity to import ( $m$ ) and a fall in the marginal propensity to consume ( $c$ ).

You should also try to articulate what factors will influence the actual values of the marginal propensity to consume ( $c$ ) and marginal propensity to import ( $m$ ) in the real world. For example, poorer nations with basic financial systems might be expected to have a higher marginal propensity to consume than nations with high disposable income.

## 7.9 Equilibrium National Income

The term equilibrium in macroeconomics is used to refer to a situation where there are no forces present which would alter the current level of spending, output and national income. At that point, firms are selling all the output they produced based on their expectations of planned expenditure. Equilibrium is associated with a position of rest.

You should be very careful not to confuse equilibrium with full employment. A macroeconomic equilibrium can occur at times when there is very high involuntary unemployment as we will learn in Chapter 11. Full employment is only one possible point of equilibrium.

Equilibrium occurs when planned expenditure is equal to national income and output.

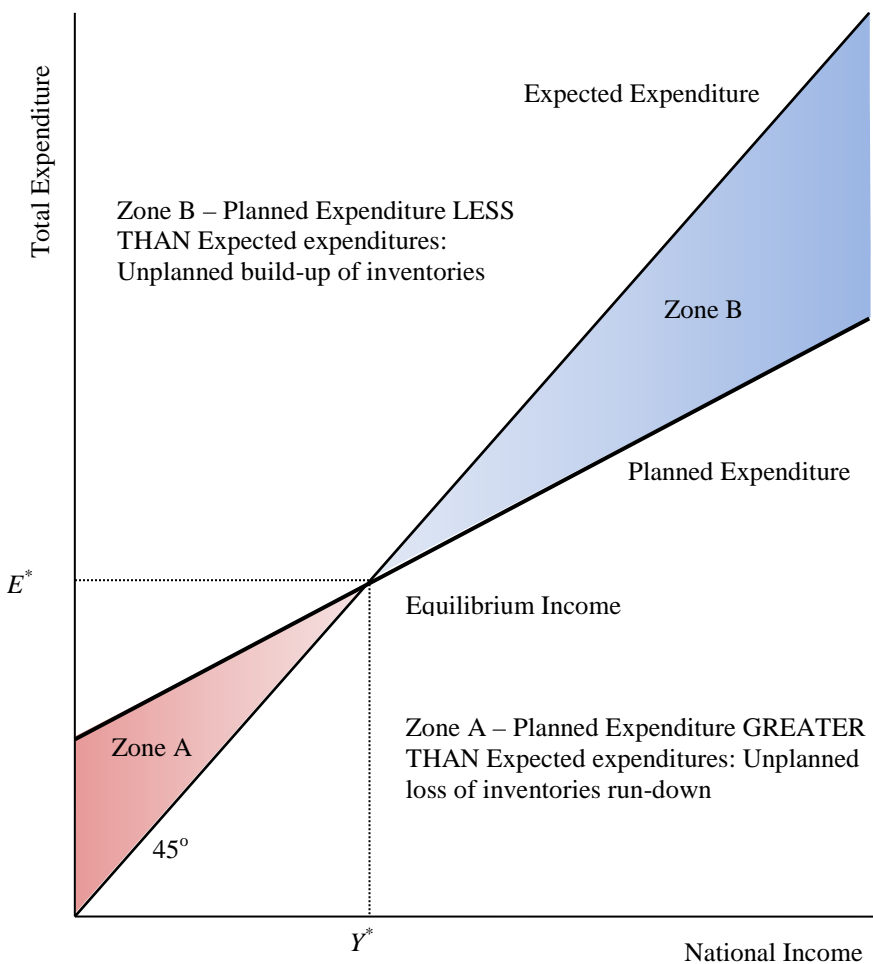
Under our current assumption that firms in the economy are quantity-adjusters and prices are fixed in the short-term, Figure 7.1 showed us that the 45-degree line was the aggregate supply curve, since the real level of Aggregate Supply equals National Income.

We are implicitly assuming that there are idle resources available for firms to deploy in expanding output. At some point, when the economy is operating at full capacity, firms are unable to continue expanding real output in response to additional spending. At that point, price rises are inevitable.

Equilibrium thus occurs when the Aggregate Demand Function cuts the 45-degree line because, at this point, the aggregate demand expectations formed by the firms, which motivated their decisions to supply –  $Y^*$  – are consistent with the total planned expenditure –  $E^*$  – by consumers, firms, government and the external economy.

Figure 7.6 shows the equilibrium income ( $Y^*$ ) and expenditure ( $E^*$ ) combination. This point defines the effective demand in the economy at this point in time.

**Figure 7.6 Planned expenditure and equilibrium income**



Note the two areas that lie between the Aggregate Demand Function and the 45-degree line. Zone A is characterised by planned expenditure being greater than expected demand and hence actual output and income. Firms have supplied insufficient output and generated less national income than would be consistent with actual planned expenditure.

In this situation, there is an unplanned reduction in the stock of inventories, which provides the signal to firms that they have been mistaken in their expectations. Firms would react to this unplanned run-down in inventories by increasing output and national income.

On the other hand, Zone B is characterised by planned expenditure being less than expected demand and hence actual output and income. Firms have been too optimistic and over-supplied output and generated more national income through their production of output than would be consistent with actual planned expenditure.

In this situation, an unplanned increase in the stock of inventories provides the signal to firms that they have been mistaken in their expectations. Firms would react to this increase in inventories by decreasing output and national income.

The inventory cycle is an important part of the cyclical adjustments that quantity-adjusting firms make to bring their expectations and production decisions into line with planned expenditure.

The importance of inventory cycles is that they lead to changes in production and income, which bring the economy back into equilibrium. The decision to increase production means that more employment will be created and the higher national income leads to an increase in planned expenditure.

Firms will continue increasing output, income and employment until their expectations are matched by planned expenditure and there are no further unplanned reductions in inventories. Equilibrium is reached at the  $Y^*-E^*$  combination, that is, where the Aggregate Demand Function cuts the 45-degree Aggregate Supply line.

Similarly, if firms find out they have produced too much, so that there is undesired inventory, they will cut back employment and production. Firms will continue decreasing output, income and employment until there are no further unplanned increases in inventories and the  $Y^*-E^*$  combination is achieved. At that point, planned inventories are being held and firms are producing in line with planned expenditure.

#### **Special Topic: Inventory movements and planned investment**

We have learned that unplanned changes in the stock of inventories lead to real GDP and national income adjustments because they signal to firms that their expectations with respect to current aggregate demand, that they formed in the past and on which they based their current production decisions, were wrong.

If the stock of inventories starts to increase, beyond the normal level that firms maintain to meet the flux in spending, it signals that firms were overly optimistic about the level of aggregate demand. Once they form the view that the discrepancy is not a random event, they will cut back on production and national income will fall.

Conversely if inventories start to be depleted below the normal level and firms think this is not an ephemeral episode, then real GDP will rise because firms will revise their expectations of aggregate demand upwards. Output, employment and national income will rise as a result.

There is an interesting disjuncture between this view of inventories and the concept of planned aggregate demand or expenditure, which pervades our analysis.



We have defined national income equilibrium as occurring when planned aggregate demand equals real GDP or national income. In Chapter 4, we learned that the national accounts always set aggregate spending equal to real GDP or national income.

However, the accounting concept of total expenditure is slightly different to our macroeconomic concept of planned aggregate demand. The difference is that the flow of spending on inventories in any given period need not accord with the planned expenditure on inventories that firms choose to make to meet the normal fluctuations in their sales.

The way the national accounts deal with this discrepancy is to classify all inventory expenditure in a period as part of gross capital formation or investment.

In macroeconomics we conceptualise the discrepancy by differentiating between planned ( $p$ ) and unplanned ( $u$ ) aggregates. So total investment,  $I = I_p + I_u$ , where the second term,  $I_u$  is the unplanned build-up (or loss) of inventories, which leads to changes in real GDP and national income. Thus  $I_u$  can be positive (negative), so if sales are lower (higher) than expected, the stock of inventories rises (falls) so that total investment will be higher (smaller) than planned.

So in the national accounts for a period,  $I_u$  would be included as part of inventory investment. But from a macroeconomic theory perspective, we would consider a positive or negative value for  $I_u$  in any period as providing evidence that the firms' expectations have been inaccurate and there is dynamic process in the economy, whereby firms change output, real GDP and national income. Equilibrium thus implies that  $I_u = 0$ .

## 7.10 The Expenditure Multiplier

There is an additional feature of this income adjustment mechanism that is important to understand. The Aggregate Demand Function (total planned expenditure), as depicted in Equation 7.13, is composed of two components: (a) the autonomous spending component ( $A = C_0 + I + G + X$ ); and (b) expenditure induced by the level of national income,  $[c(I - t) - m]Y$ .

What would be the impact on real GDP (and national income) if one of the components of autonomous spending changed? We know that real GDP and national income will rise if planned spending rose and will fall if planned spending falls. The question of interest now is by how much will real GDP and national income change after a change in planned spending driven by a change in autonomous spending (for example, an increase in government spending).

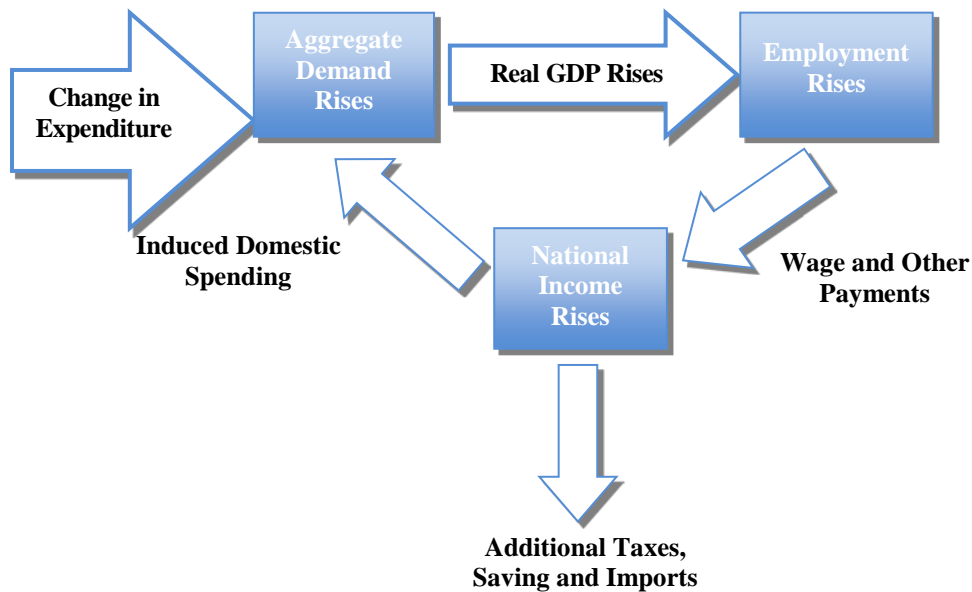
Economists have developed the concept of the **expenditure multiplier** to estimate how much national income ( $Y$ ) will change for a given change in autonomous spending ( $A$ ).

Figure 7.7 sketches the expenditure multiplier process. From an initial equilibrium position, an increase in autonomous expenditure provides an instant boost to aggregate demand. Firms respond to the increased planned expenditure and raise employment to produce the increased output (real GDP), as explained in the last section. National income increases.

This rise in national income induces further consumption spending which leads to a further rise in aggregate demand. A proportion of the rise in national income leaks out in the form of higher tax payments and imports and increased saving.

The process continues until the induced spending becomes so small that there are no further real GDP increases. The process works in reverse for a fall in autonomous expenditure.

**Figure 7.7 The multiplier process**



### An algebraic treatment

The formal expression for the expenditure multiplier is derived directly from the equilibrium national income and expenditure relationship.

The Aggregate Demand Function was expressed as  $E = A + [c(1 - t) - m]Y$ . The national income equilibrium condition is given as:

$$(7.14) \quad Y = E$$

If we substitute the equilibrium condition into the Aggregate Demand Function (7.14) we get:

$$(7.15a) \quad Y = E = A + [c(1 - t) - m]Y$$

and solving for  $Y$  (by collecting  $Y$  terms on the left-hand side) gives:

$$(7.15b) \quad Y[1 - c(1 - t) + m] = A$$

Thus equilibrium income is:

$$(7.15c) \quad Y = 1/[1 - c(1 - t) + m]A$$

The expenditure multiplier ( $\alpha$ ) is the coefficient next to the  $A$  term in Equation (7.15).

$$(7.16) \quad \alpha = \Delta Y/\Delta A = 1/[1 - c(1 - t) + m]$$

So if  $A$  changes by \$1 then  $Y$  changes by  $\alpha = 1/[1 - c(1 - t) + m]$  or  $\alpha$  times the change in  $A$ . The denominator of the expression for  $\alpha$  is less than unity, so the multiplier exceeds one. This is to be expected since, after the initial increase of autonomous expenditure,  $\Delta A$ , any induced increase in the consumption of domestically produced goods and services will lead to  $\Delta Y$  exceeding  $\Delta A$ .

By inspecting the terms that define the expenditure multiplier, you can see that it is a ratio involving the marginal propensity to consume ( $c$ ), the marginal tax rate ( $t$ ) and the marginal propensity to import.

Applying the tools you learned in the Appendix, you can observe the following:

- Other things equal, the higher is the marginal propensity to consume, the higher is the expenditure multiplier.
- Other things equal, the lower is the tax rate, the higher is the expenditure multiplier.
- Other things equal, the lower is the marginal propensity to import, the higher is the expenditure multiplier.
- The opposite is the case if  $c$  is lower and  $t$  and  $m$  are higher.

The task now is to explain the economic processes that lead to these conclusions.

We start with the essential insight that aggregate demand drives output, which generates incomes (via payments to the productive inputs). Accordingly, what is spent generates output and income in that period. The income is then available for use. There are various ways in which the income derived from the payments arising from output production can be used, namely:

- Consumption expenditure.
- Saving.
- Meeting tax obligations to government.
- Spending on imports.

For example, workers, who are hired by firms to produce goods and services, spend part of their wage income that they earn on consumption. They also meet their tax obligations and may save a portion of their disposable income.

### A graphical treatment

In Figure 7.4 we learned that if any of the components of autonomous aggregate expenditure change, the Aggregate Demand Function shifts up or down, with the extent of the shift being measured by the change in the vertical intercept.

Assume that government spending rises as a result of the government being concerned that the rate of unemployment is too low. In Chapter 12 we will learn that mass unemployment is always the result of deficient aggregate demand relative to the productive potential of the economy and a simple remedy is for governments to increase total spending.

Figure 7.8 shows the change in equilibrium expenditure and income when government spending increases ( $\Delta G$ ). Point A is the initial equilibrium real GDP and national income level,  $Y_0^*$ , which corresponds to aggregate expenditure of  $E_0^*$ . The Aggregate Demand Function is given as  $E = C + I + G_0 + NX$ .

At this point there are no unplanned inventory changes and firms' production decisions are based upon expected aggregate demand being realised.

Now government spending increases by  $\Delta G$ , which increases the Aggregate Demand Function (where  $E = C + I + G_1 + NX$ ) and real GDP and national income increases to  $Y_1^*$ , which corresponds to aggregate expenditure of  $E_1^*$ . The new equilibrium national income is at Point B.

The reason that equilibrium real GDP and national income increase relates to the firms' revision of expected expenditure. When the government injects the new autonomous spending in to the economy, aggregate spending at the current equilibrium is greater than real output. The difference is the line segment AA'.

This distance indicates the excess aggregate demand (relative to current real GDP) and the stock of inventories would be falling. Firms would soon revise their expectations of aggregate demand upwards and start to produce more real output and generate higher levels of national income.

They would continue to increase production and national income until their aggregate demand expectations were consistent with actual aggregate demand, a state which occurs at Point B (where the new Aggregate Demand Function cuts the 45° aggregate supply line).

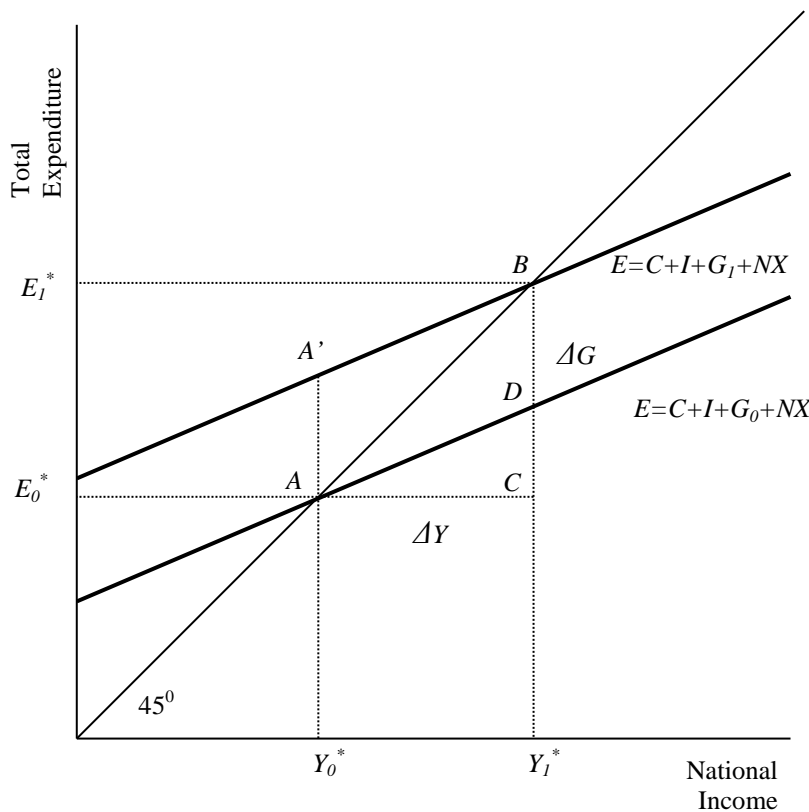
Note that the change in equilibrium national income,  $\Delta Y$  is greater than the initial change in autonomous expenditure,  $\Delta G$ . The difference between the two changes is given by the line segment CD.

**How do we explain this difference?**

The expenditure multiplier indicates by how much real GDP and national income changes when there is a change in autonomous expenditure. The larger is the multiplier, the larger is the change in real GDP and national income for a given change in autonomous expenditure.

The total change in aggregate demand ( $\Delta E$ ) following a change in autonomous expenditure (in this case,  $\Delta G$ ) is the sum of  $\Delta G$  (segment BD) and the induced consumption spending (segment DC) that follows the initial rise in national income. Refer back to Figure 7.7 if you are unsure about this point.

**Figure 7.8 Impact of a change in government spending on equilibrium expenditure and income**



The induced consumption spending is shown as ‘Induced Domestic Spending’ in Figure 7.7. As firms react to the initial disequilibrium at Point A (the excess aggregate demand AA’) by increasing real GDP and national income, households, in turn, increase their consumption expenditure. But at the same time, imports are rising by  $m\Delta Y$ , tax revenue is rising by  $t\Delta Y$  and households save a portion of each extra dollar of disposable income,  $(1-c)\Delta Y_d$ .

These leakages mean that each subsequent round of induced spending is smaller than the last and eventually becomes zero. At that point, the economy reaches the new equilibrium at Point B in Figure 7.8.

So the total change in real output and national income,  $\Delta Y$  is equal to the total change in aggregate expenditure,  $\Delta E$ , which is equal to the initial change in autonomous spending,  $\Delta A$  plus the induced consumption  $\Delta C$ .

### Numerical Example of the expenditure multiplier at work

In our example there is an initial spending increase of \$100, which might have been a government order for new public school buildings. This extra \$100 in government spending leads construction firms to produce more output and increase total income payments by \$100 (under the assumption in this Chapter that firms are quantity adjusters), some of which is earned by construction workers in the form of wages. These workers and other income recipients then spend some of the additional income on goods and services produced locally and further afield.

Refer back to Figure 7.7 to reinforce your understanding of the sequence of events.

Assume that the marginal propensity to consume ( $c$ ) is 0.75, the current tax rate is 0.20, and the marginal propensity to import is 0.20. This means that for an extra \$100 of national income:

- \$20 goes to tax revenue and is drained from the domestic economy.
- Disposable income thus rises by \$80 and household consumption rises by \$60 with the residual being increased saving of \$20.
- \$20 is spent on additional imports and is lost to the domestic economy.
- Total leakages from the initial \$100 of extra income that is generated, namely taxes plus saving plus imports, are thus \$60, leaving additional consumption on **domestically produced** goods and services at \$40.

The way to think of the second-round expenditure injection is to note that national income rises by \$40 in response to the additional consumption spending on domestically produced goods and services - which is referred to as **induced consumption** - and then to focus on the additional leakages. After taxation is taken out, consumers determine how much they wish to spend on increased consumption, with saving then being the residual.

Table 7.3 shows the process for 10 rounds of additional induced domestic spending (with the last round of induced consumption being close to zero) following the initial rise in government spending by \$100. We deliberately use the term ‘round’ so as not to give the impression that the adjustment follows an orderly process across actual time. The rounds are of indeterminate length and may be irregular with respect to each other.

Each successive induced spending increase is smaller than the last because of the leakages.

The initial spending ‘fans out’ or spreads throughout the entire economy. The initial spending **multiplies** into a much larger increase in spending.

**Table 7.3 The expenditure multiplier process**

	$\Delta$ GDP	$\Delta$ Taxes	$\Delta$ Disposable Income	$\Delta$ Consumption	$\Delta$ Saving	$\Delta$ Imports	$\Delta$ Total Leakages
	[1]	[2]	[3]	[4]	[5]	[6]	[7]
Round 1	100.0	20.0	80.0	60.0	20.0	20.0	60.0
Round 2	40.0	8.0	32.0	24.0	8.0	8.0	24.0
Round 3	16.0	3.2	12.8	9.6	3.2	3.2	9.6
Round 4	6.4	1.3	5.1	3.8	1.3	1.3	3.8
Round 5	2.6	0.5	2.0	1.5	0.5	0.5	1.5
Round 6	1.0	0.2	0.8	0.6	0.2	0.2	0.6
Round 7	0.4	0.1	0.3	0.2	0.1	0.1	0.2
Round 8	0.2	0.0	0.1	0.1	0.0	0.0	0.1
Round 9	0.1	0.0	0.1	0.0	0.0	0.0	0.0
Round 10	0.0	0.0	0.0	0.0	0.0	0.0	0.0
<b>Total change</b>	166.60	33.30	133.30	100.00	33.30	33.30	100.00

Note the **leakages** are getting smaller with each spending period following the initial injection of \$100. In the bottom row we see that the sum of all the period-by-period changes in GDP (Column [1]) equals \$166.60. The total increase in tax revenue (Column [2]) is \$33.30. The total increase in disposable income (Column [3]) is thus \$133.30, which leads to total induced consumption (Column [4]) of \$100 and the total increase in saving (Column [5]) is \$33.30. Total Imports (Column [6]) rise by \$33.30.

The total leakages from the expenditure system – taxes, saving and imports – sum to \$100.00 at the end of the adjustment period, which is the amount of the total initial injection in autonomous expenditure. At the point when additional leakages equal the additional injections, the system is at rest and the multiplied impact of the initial injection in autonomous expenditure is complete.

Think back to our algebraic definition of national income equilibrium shown in Equation (7.15) and the related expression for the expenditure multiplier, Equation (7.16).

If we substitute the assumed values in this example into the multiplier formula we get:

$$\text{Multiplier} = \Delta Y / \Delta G = 1 / [1 - c(1 - t) + m] = 1 / [1 - 0.75(1 - 0.20) + 0.20] = 1.666$$

The multiplier is the total change in GDP for a one-dollar initial increase in aggregate demand. In this example, given the values of the marginal propensity to consume, the tax rate and the marginal propensity to import, the multiplier is calculated to be 1.666 (rounded).

That means, that if autonomous expenditure (for example, government spending) rose by \$100, the total change in GDP, after the economy adjusts to the higher production and income levels, would be \$166.60 (rounded).

### Changes in the magnitude of the expenditure multiplier

Equation (7.16) defined the expenditure multiplier as  $\Delta Y/\Delta G = 1/[1 - c(1 - t) + m]$ , which means that its size depends on the marginal propensity to consume ( $c$ ), the tax rate ( $t$ ) and the marginal propensity to import ( $m$ ).

The following conclusions can be drawn at this stage:

- The multiplier is larger (smaller) the larger (smaller) is the marginal propensity to consume ( $c$ ).
- The multiplier is larger (smaller) the smaller (larger) is the marginal propensity to import ( $m$ ). The more open the economy to trade, the lower is the multiplier.
- The multiplier is larger (smaller) the smaller (larger) is the tax rate ( $t$ ).

A higher marginal propensity to consume means that each successive round of induced consumption spending is larger, other things equal. Given that the marginal propensity to save ( $s$ ) is just  $(1-c)$ , the multiplier is higher when the marginal propensity to save is lower.

In general, the lower are the leakages (taxes, saving and imports) from the expenditure system, the higher will be multiplier. This is because the lower the leakages from each spending round, the larger is the induced consumption

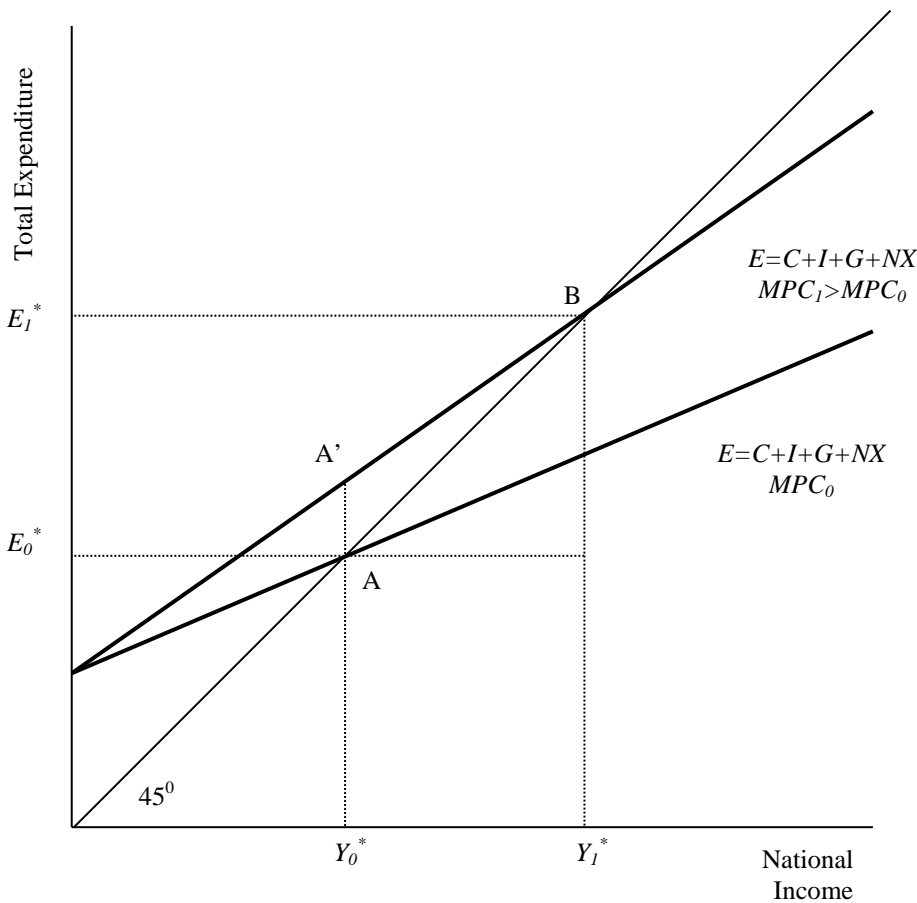
Earlier in the Chapter, we learned that the slope of the Aggregate Demand Function (see Figure 7.4) was  $[c(1-t) - m]$ , which tells us that the change in aggregate spending for a given change in national income is larger, the larger is the marginal propensity to consume ( $c$ ) and the lower is the tax rate ( $t$ ) and the marginal propensity to import ( $m$ ).

Consider an increase in the marginal propensity to consume ( $c$ ). Figure 7.9 depicts this case. The initial Aggregate Demand Function (for  $MPC_0$ ) is associated with a national income equilibrium at Point A where  $E_0^*$  generates production and national income of  $Y_0^*$ .

When the marginal propensity to consume rises to  $MPC_1$ , the Aggregate Demand Function pivots upwards at the vertical intercept. At the current equilibrium income level,  $Y_0^*$ , households choose to spend an increased proportion of each dollar of disposable income on consumption. The initial change in aggregate demand ( $E$ ) is measured by the distance A to A'.

The economy responds to the increased consumption spending by increasing production and national income. The higher multiplier (as a result of the higher MPC) then drives national income up further and the economy reaches a new equilibrium at Point B. At that point the economy comes to rest again.

**Figure 7.9** Impact of a change in the marginal propensity to consume on equilibrium expenditure and income



The following conclusions can be drawn at this stage:

- The slope of the Aggregate Demand Function is steeper (shallower) the larger (smaller) is the marginal propensity to consume ( $c$ ). This means that for a given flow of autonomous spending, aggregate demand and national income will be higher if the marginal propensity to consume rises, and vice versa.
- The slope of the Aggregate Demand Function is steeper (shallower) the smaller (larger) is the marginal propensity to import ( $m$ ). This means that for a given flow of autonomous spending, aggregate demand and national income will be higher if the marginal propensity to import falls, and vice versa.
- The slope of the Aggregate Demand Function is steeper (shallower) the smaller (larger) is the tax rate ( $t$ ). This means that for a given flow of autonomous spending, aggregate demand and national income will be higher if the tax rate falls, and vice versa.

Table 7.4 shows the impact of varying the parameters that determine the size of the multiplier ( $c$ ,  $m$  and  $t$ ) on national income for a given injection of autonomous expenditure. You might like to simulate multiple changes in the parameters (for example, a rising MPC and a rising tax rate) to gain a greater understanding of how these influences interact.



**Table 7.4 Simulating changes in the multiplier components**

---

<b>Varying the Marginal Propensity to Consume</b>			
Marginal Propensity to Consume (c)	0.6	0.7	0.8
Marginal Propensity to Import (m)	0.2	0.2	0.2
Tax Rate	0.2	0.2	0.2
Multiplier	1.39	1.56	1.79
Autonomous Spending	100	100	100
National Income	139	156	179
<b>Varying the Marginal Propensity to Import</b>			
Marginal Propensity to Consume (c)	0.8	0.8	0.8
Marginal Propensity to Import (m)	0.1	0.2	0.3
Tax Rate	0.2	0.2	0.2
Multiplier	2.17	1.79	1.52
Autonomous Spending	100	100	100
National Income	217	179	152
<b>Varying the Tax Rate</b>			
Marginal Propensity to Consume (c)	0.8	0.8	0.8
Marginal Propensity to Import (m)	0.2	0.2	0.2
Tax Rate	0.1	0.2	0.3
Multiplier	2.08	1.79	1.56
Autonomous Spending	100	100	100
National Income	208	179	156

---

## References

ABS (2015) "Expenditure Chain Volume Measures", Australian National Accounts: National Income, Expenditure and Product, September (Cat.: 5206.0).

# Chapter 8: Introduction to Aggregate Supply

## Chapter Outline

- 8.1 Introduction
- 8.2 Some Introductory Concepts
  - Schedules and functions
  - The employment-output function
  - Money wages
- 8.3 Price Determination
- 8.4 The General Aggregate Supply Function
- 8.5 Some Properties of the General Aggregate Supply Function (AS)
- 8.6 Factors Affecting Aggregate Output per Hour
  - How does this affect our understanding of production costs?
  - How would a firm react to an increase in aggregate demand?
  - What factors might explain the observed pro-cyclical movement in labour productivity?

## Learning Objectives

1. Understand the mark-up pricing model and its underlying assumptions.
2. Explain why the pricing model is consistent with firms acting as quantity adjusters.
3. Recognise that labour productivity can be pro-cyclical.

## 8.1 Introduction

In Chapter 7 our theory of expenditure and income determination linked aggregate spending to the generation of income. The focus on the demand drivers of aggregate income and output abstracted from any spending impacts on the price level and assumed that the most firms in the economy were rather passive. They simply responded to growth in nominal spending by increasing real output up to the full capacity level in the economy. In doing so, we ignored the complexity of the supply-side. We also abstracted from what might happen after the economy reached its full capacity level.

In this Chapter we shall develop a widely used model of mark-up pricing. This model provides a rationale for the claim that over a range of output the price level is more or less constant. This means that, as a first approximation, treating firms as quantity adjusters in response to changing levels of total expenditure is a reasonable assumption.

The theory we develop in this Chapter will therefore complete the demand-side model that we developed in Chapter 7 to allow us to determine the real output level, the price level and total employment.

## 8.2 Some Introductory Concepts

### Schedules and functions

In the Appendix *Methods, Tools and Techniques* we introduced the essential analytical and introductory techniques that students should learn in order to grasp macroeconomics.

As a reminder, economic models use schedules or curves to depict behaviour, which can either be *ex ante* (prior to action and reflects planned or desired action by households, firms, government etc); or *ex post*, which represents actual outcomes that are the result of action.

In the simplest macroeconomic model of expenditure, income and employment we encounter an aggregate demand schedule and an aggregate supply schedule. These schedules depict *ex ante* behaviour and tell us what the outcomes will be given other conditions in the economy. In this Chapter we will consider aggregate supply schedules.

The terms, schedule and function, are used interchangeably in the economics literature. We prefer to use function to depict a relationship between variables – such as spending and income.

### The employment-output function

To develop a theory of employment – that is, explain its level and movement over time – in relation to a monetary economy operating under capitalist conditions, we need to develop an understanding of how employment is related to output determination. This relationship is also important because **per unit labour costs** (total labour costs divided by total output) underpin the pricing of output via a **price mark-up**. In this context we develop the concept of the Employment-Output function, which shows the how much labour is required to produce a given volume of real output.

Given the output that the firm plans to produce in response to the expected demand for its good or service, employment will be determined by the productivity of labour. This production decision is made in an environment of stable wage rates and capital-labour ratios. The capital-

labour ratio depicts the combination of productive capital (machines, equipment, etc) and labour that defines the current productive technology.

For example, an excavation firm might provide a hand shovel to each worker engaged in digging foundations for a new building. This would be a low capital-labour ratio production technology. Sometimes this is referred to as a labour-intensive technique.

Alternatively, it could use mechanical digging equipment and employ fewer workers to produce the same output. In this instance, the production process would employ higher capital-labour ratio techniques – sometimes referred to as capital-intensive production.

We can write the **Employment-output function** as:

$$(8.1) \quad Y = \gamma N$$

where  $N$  is the total number of workers employed,  $\gamma$  is the rate of labour productivity, and  $Y$  is planned real output (based on expected spending).

What is labour productivity? Labour productivity is defined as output per unit of labour input per period of time, for example, per hour. So we could solve Equation 8.1 for  $\gamma$  to get  $Y/N$ , which is the algebraic equivalent of our definition.

The higher is labour productivity ( $\gamma$ ) the less employment is required to produce a unit of output for a given production technique (implicit in  $\gamma$ ).

Factors, which influence the magnitude of  $\gamma$ , include: technology (whether it is best-practice, capital- or labour-intensive); worker skill and motivation; and management skill and business organisation. Often in the public discussions about slowing productivity growth there is an undue focus on the worker with claims such as poor motivation and skill gaps. Rarely is management skill the focus of enquiry despite evidence that poor management decision-making is a cause of slow productivity growth. For example, the failure to invest in the latest technology will lower the growth in labour productivity.

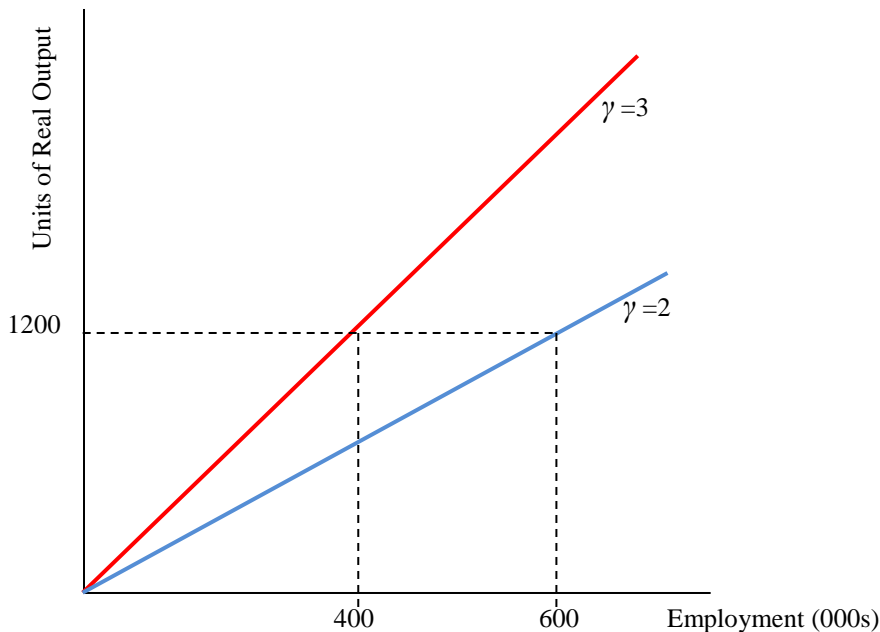
As an example, the Australian airline Qantas dominated the international travel market for Australians travelling abroad. In the late 1970s the airline carried around 42 per cent of Australian travellers abroad. By 2012, this proportion had dropped to 18 per cent as competition from airlines such as Emirates and Singapore Airlines had cut into its market share. There are many reasons for this decline in market share, but one of the major explanations was that Qantas management made poor decisions with respect to its fleet upgrades and refused to invest in the latest jets, which were more fuel-efficient and hence could operate at lower cost.

If  $\gamma$  is stable in the short-run (within the current investment cycle) then once the firm decides on the level of output to produce to satisfy expected demand, it simultaneously knows how many workers must be employed. As an example, if it takes 10 workers to produce 1000 units of output per day, then daily labour productivity would be 100 per worker. So if the firm anticipated an increase in output to say 1500 per day, it would require an additional 5 workers to ensure it could supply the new higher level of output.

Figure 8.1 shows the two different employment-output functions for the economy, each associated with a constant  $\gamma$ , so these functions are positively sloped straight lines.

If firms expected aggregate demand would be 1200 thousand units in the current production period, then given the state of technology (represented by  $\gamma$ ) it would employ 600 thousand workers if  $\gamma = 2$  (lower productivity) and 400 thousand workers if  $\gamma = 3$  (high productivity).

**Figure 8.1** The employment-output function



Firms produce based on expected aggregate spending and once all the sectors have made their spending decisions (that is, once aggregate demand is actually realised), the firms discover whether their expectations were accurate or not. That is, they find out whether they have overproduced, under-produced or produced the right amount, once spending has occurred.

### Money wages

A useful short-run assumption is to assume that money wage rates are exogenous in the short-run. This is not the same as assuming that money wage rates never change. It merely says that, in terms of the parameters of our aggregate supply model (that is, the different influences that we consider will impact on aggregate supply), the money wage rate will be assumed to be invariant in the short-run.

Before we discuss the possible factors, which make this a reasonable assumption to make, we must first clarify some, often-confused concepts relating to wages.

We distinguish between the **money wage** rate and the **real wage** rate.

The money wage rate is determined in the labour market and is the amount, in nominal (current dollar) terms, that the workers receive per hour when they sell their labour power to the capitalist business firms or other employers (for example, government). The actual money wage at any point in time is the outcome of agreements reached between employers and workers, either on a decentralised, negotiated basis or through sector- or economy-wide negotiations. In some nations, such as Australia, there has been a history of wage setting tribunals (courts) which led to the practice of industrial relations in general and the determination of wages in particular to

become a specialised judicial process, reflecting the adversarial nature of relations between workers and capital.

The money wage outcome at any point in time is heavily dependent on the bargaining strengths of the parties involved. Wage changes occur at infrequent intervals and condition the behaviour of the parties concerned for the ensuing economic period (sometimes months, usually years). It is this infrequent nature of wage setting via institutional structures (such as, employer-union negotiations) and the implied contractual nature of the wage relationship existing between employers and workers over some future period, which is used to justify the assumption that money wages are exogenous and fixed in the short-run for the purpose of developing an explanation of aggregate supply.

The real wage rate is the money wage rate deflated by some price index. We learned how deflators are constructed and are used to convert current price variables into constant price (real) variables in Chapter 4

The choice of deflator depends on the context. The real wage, from the perspective of the worker, would be the money wage expressed in terms of real consumption good equivalents. So we would consider this rate to be the money wage rate divided by a measure of consumer prices (the Consumer Price Index).

From the employer's perspective, the real (product) wage is more accurately measured by the money wage paid to workers divided by the specific price the firm receives for its output, which is a narrower concept than the real wage considered from the perspective of the worker.

Importantly, contrary to what most mainstream textbooks will suggest, the real wage is not determined in the labour market and can only be influenced by the workers in as much as they can influence the money wage rate outcome.

The **real wage is a ratio of two prices** – the money wage (determined in the labour market) and the consumer price level (determined in the goods and services market). The two prices, which form the real wage rate, are determined by different forces in different markets in the economy.

As we will see, prices are largely set by business firms in the goods and services (product) market according to desired mark-ups on cost. Prices are not fixed by workers.

Often economists and others suggest that workers should cut their real wages to improve the employment prospects of the unemployed. We argue that the policy suggestion is without merit. But as a precursor to that discussion, even if the proposition was based on a causal understanding of how mass unemployment occurs, there are several preliminary, but critical questions that such proposals fail to answer.

How can workers achieve a cut in their real wage when they can only influence the money wage outcome?

How might a money wage change influence price changes? In particular, a money wage cut may lead to a price cuts due to the fall in the costs of production, and thus leave the real wage unchanged.

These initial queries are quite apart from the dispute among economists as to whether a real wage cut would influence employment growth independent of changes in effective demand.

What is the basis of the money wage inflexibility assumption? First, negotiations over money wages typically occur at infrequent intervals, as noted above. Second, there is strong evidence that workers resist cuts in money wages and firms, generally prefer not to offer such cuts. Only in extraordinary circumstances relating to the imminent collapse of the enterprise in which they are employed and the existence of very high levels of unemployment have we observed workers agreeing to money wage reductions.

The downward rigidity of money wages is also the result of employer preferences. Even when the unemployment rate approaches double figures (a rate considered high by historical standards), the absolute number of workers not in employment relative to those who retain their jobs is small. As such, employers are reluctant to risk jeopardising convivial industrial relations with the majority of workers to possibility improve the employment prospects of a small proportion of employed workers.

We consider these issues in more detail in Chapter 11 Unemployment and Inflation.

### 8.3 Price Determination

Clearly, a firm seeks to generate a profit over and above the costs of production. How does it go about setting the price that it will accept for its output?

Firms are assumed to operate in a **non-competitive** economy. You may have considered the case of perfect competition in a microeconomics course where firms are assumed to have no price setting discretion because the market is so large and firms are assumed to be so small.

We are thus introducing oligopoly as a basic assumption rather than the orthodox use of perfect competition as the benchmark. Hence firms are considered to be **price-setters rather than price-takers**. Firms are assumed to fix their prices as a mark-up over costs. Economists are divided about the determinants of the mark-up and the costs considered relevant in the pricing decision by firms.

Further, debate remains as to whether the mark-up is invariant to the state of demand. However, the use of the mark-up as a basic description of firm behaviour in the real world is difficult to dispute.

In the real world, firms typically have discretionary price-setting power and seek a rate of return on the capital employed, which necessitates that they generate a profit margin over their total costs of production.

The total price per unit sold must therefore cover its (variable) costs of production per unit of output, such as labour and raw material costs, plus the profit margin. We will assume that the profit margin covers overheads and other fixed costs plus net profit.

Firms are thus assumed to employ a mark-up pricing model such that:

$$(8.2) \quad P = (1 + m)[W/\gamma]$$

where  $P$  is the price of output,  $m$  is the per unit mark-up on unit labour costs,  $W$  is the money wage per hour and  $\gamma$  is labour productivity per hour. At this stage we abstract from raw material costs. Thus  $\gamma$  is defined as the units of output per unit of labour input per hour.

If  $\gamma = 0.5$  then 2 labour hours are required to produce one unit of output. If the money wage ( $W$ ) was \$5 per hour, then the unit labour costs (that is, labour cost per unit of output) would be \$10.

As noted, the mark-up ( $m$ ) is set to provide a surplus above the direct unit labour costs to account for fixed (overhead) labour and other fixed costs, including interest payments on loans, in addition to a provision for profits (return on equity). The amount of profit desired is related, in part, to the amount of investment that the firms plan to undertake because retained earnings are an important source of internal finance that the firm draws on to reduce its exposure to higher costs of externally funding new projects.

In the short-run, the price will be rigid with the firm supplying output according to demand. Price changes would occur when there were changes in the money wage rate or other variable costs, the mark-up (margin), or trend labour productivity. Trend labour productivity is used here to differentiate it from the cyclical swings that occur in labour productivity, which we consider in Section 8.6 of this Chapter.

The mark-up or margin ( $m$ ) is a reflection of the market power of the firm. The higher the market power, the higher will be the margin. Thus, in more competitive sectors, the margin will tend to be lower than in less competitive sectors. Also changes in competitiveness of a sector will, over time, lead to changes in the size of the mark-up.

If in our example, the mark-up ( $m$ ) is set at 40 per cent, then the firms will price its output at \$14 per unit (\$10 multiplied by 1.40).

The features of this approach are as follows:

1. Prices are unambiguously a function of costs.
2. Firms use their price-setting discretion to generate a monetary surplus above average variable costs. This monetary surplus is designed to cover profits. Importantly, profits are considered to be influenced in the short-run by the ability of firms to realise the mark-up on unit costs. Factors which may squeeze the mark-up (down to say 30 per cent) will accordingly also squeeze profits per unit of output.
3. The mark-up impacts directly on the real wage that workers receive. Assume that total marked-up costs only include (for simplicity) wage costs. Total wage costs are the product of the money wage rate  $W$  and the number of workers employed  $N$ , that is,  $WN$ .

A simplified price mark-up model would be in this case:

$$(8.3) \quad P = (1 + m)WN/Y$$

where all the terms are as defined previously.  $WN/Y$  is wage costs per unit of output, in other words per unit labour costs, which we defined above.

We can re-write this equation as:

$$(8.4) \quad Y/(1 + m) = WN/P$$

and further re-arrangement yields:

$$(8.5) \quad W/P = (Y/N)/(1 + m) = \gamma/(1 + m)$$

which says that the real wage ( $W/P$ ) is dependent on the average productivity of labour ( $Y/N$ ) and the size of the mark-up. The larger the mark-up ( $m$ ), other things being equal, the lower is the real wage.

Fourth, the volume of profits (as distinct from the per unit profit) depends on the size of the mark-up – which influences profit per unit of output– and the actual volume of output sold in any



period. The latter is determined by the state of aggregate demand in the economy and, as we saw in Chapter 7, is determined by the level of household consumption expenditure, private investment expenditure, net exports and government spending.

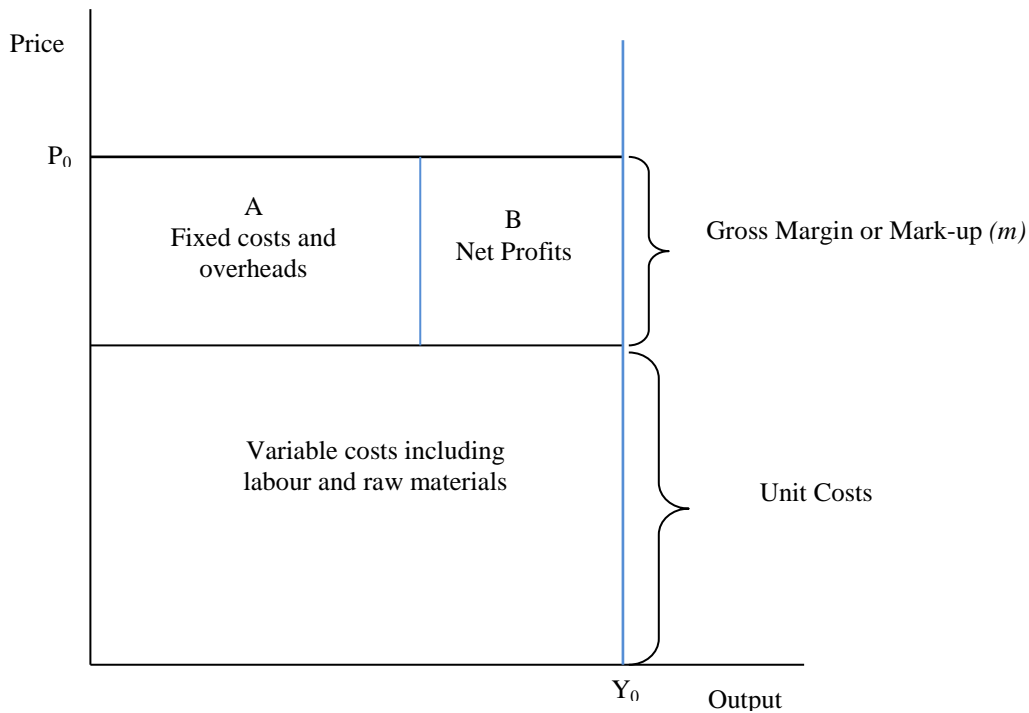
Figure 8.2 shows the way in which the price set by all firms ( $P_0$ ) at a point in time is distributed as incomes. Here the current level of output being produced is  $Y_0$ . The price  $P_0$  is a mark-up on total unit variable costs which covers fixed costs (including labour overheads) and an allowance for profit.

Total revenue for the economy as a whole is the area defined by  $P_0$  times  $Y_0$  and the distribution of that level of output as income is shown by the areas below the price line. Fixed costs are represented by the rectangle A, whereas rectangle B represents net profit.

Firms thus supply the output that is demanded at the price  $P_0$ . They produce a given level of output according to their expectation of total spending in the economy. The diagram below makes no presumption that the level of output  $Y_0$  is consistent with that expectation. It is, in fact, total output sold and may or may not satisfy the firms' expectations.

In that sense, the net profits generated may be below or above the level that the firm aimed to achieve at the beginning of the production period.

**Figure 8.2 Output, sales and national income**



Fifth, usually mark-up theories assume that the immediate impact of changes in demand on the mark-up and hence prices is small. For the planning period ahead, firms calculate their costs and desired profits on the basis of an expected level of output, which they believe they can sell. Deviations in this expected level of demand promote output changes rather than price changes.

Firms may plan to increase profits by raising the mark-up.

## 8.4 The General Aggregate Supply Function

Before we consider complicating factors, such as changes in productivity and competitiveness, it is useful to consider what the price determination rule means for the shape of the aggregate supply function.

If we assume that  $m$ ,  $W$  and  $\gamma$  are constant in the short-run then the aggregate supply curve would be a horizontal line in the price-real income graph up to some full capacity utilisation point ( $Y^*$ ). Economists sometimes refer to a horizontal line in this context as being perfectly elastic. Firms in aggregate will supply as much real  $Y^*$  output (goods and services) as is demanded at the current price level set according to the mark-up rule described above.

Figure 8.3 is similar to Figure 8.2 but adds the full capacity utilisation level of real output ( $Y^*$ ) to derive the **General Aggregate Supply Function (AS)**. This shaped AS function is sometimes referred to as a reverse-L shape for obvious reasons.

The horizontal segment has been explained by the price mark-up rule and the assumption of constant unit costs. But why does it become vertical after full employment?

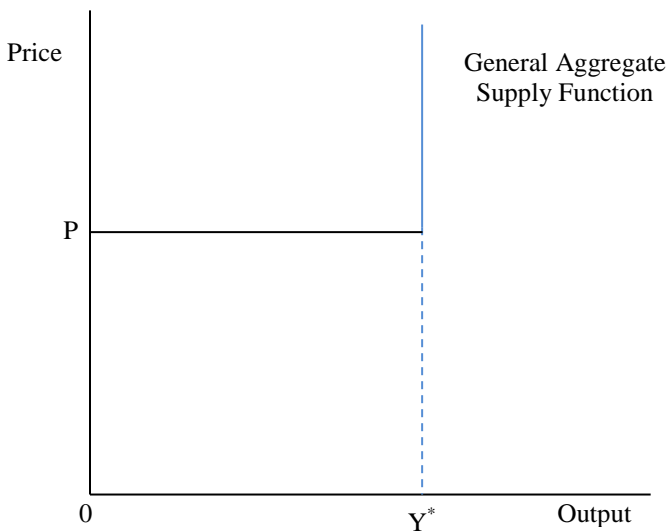
After this point, the economy exhausts its capacity to expand short-run output due to shortages of labour and capital equipment. At that point, firms will be trying to outbid each other for the already fully employed labour resources and in doing so would drive money wages up. We will return to this possibility later in this Chapter.

Under normal circumstances, the economy will rarely approach the output level ( $Y^*$ ) which means that for normal utilisation rates the economy faces constant costs.

There is some debate about when the rising costs might be encountered given that all firms are unlikely to hit full capacity simultaneously. The **reverse-L shape** simplifies the analysis somewhat because it assumes that the capacity constraint is reached by all firms at the same time. In reality, bottlenecks in production are likely to occur in some sectors before others and so cost pressures will begin to mount before overall full capacity output is reached.

This could be captured in Figure 8.3 by some curvature near  $Y^*$ , thus eliminating the right-angle. We consider this issue in more detail in Chapter 11.

**Figure 8.3 The general aggregate supply function (AS)**



### 8.5 Some Properties of the General Aggregate Supply Function (AS)

The AS equation is simply the price determination model Equation (8.2), which shows that in the short-run, the behaviour of the aggregate supply in the economy depends on  $m$ ,  $W$  and  $\gamma$ .

Accordingly:

- If the money wage rate rises, other things equal, the unit cost level rises and the firms would translate this in time into a price rise thereby restoring the previous mark-up.
- If there is growth in labour productivity ( $\gamma$ ) as a result of say, increased labour force morale, increased skill levels, more technologically-based production techniques, better management, and the like, then unit costs ( $W/\gamma$ ) will fall. This means that the firms can generate the same profit margin at lower prices. The AS function would thus shift downwards by the extent of the decline in unit costs.
- Variations in the mark-up ( $m$ ) will cause the price level to change. Increases in industrial concentration, more advertising etc may lead to firms being able to increase the overall profit margin that can be sustained. Tight conditions in the goods and services market, where sales are constrained, may lead firms to reduce the mark-up as they all struggle for market share. This could also occur as a result of strong trade unions pushing (successfully) for wage increases. Thus to avoid losing market share, the firms may choose to absorb some of the cost rises into the margin.
- If employment is below full employment, then actual output is less than  $Y^*$ , which means there is an output gap. Increases in aggregate demand (spending), which are seen by firms to be permanent, will result in an expansion of output without any price increases occurring. If the firms are unsure of the durability of the demand expansion, they may resist hiring new workers and utilise increased overtime instead. That is, they initially respond to the increased aggregate spending by increasing hours of work rather than persons employed. The higher costs associated with paying overtime rates are likely to be absorbed in the profit margin because firms desire to maintain their overall market share.

The Aggregate Supply Function is a useful vehicle for exploring an inflationary process arising from conflict between groups over the distribution of income. We postpone this analysis until Chapter 11.

## 8.6 Factors Affecting Aggregate Output per Hour

What factors determine the impact of change in hours of employment on aggregate output? Over time, many influences are at work. These include improvements in technology, changes in the average quality of labour from increased education and health, and changes in organisational and management skills, which will lead to a steady increase in the level of output that is produced from a given quantity of inputs.

In seeking to understand short-run employment and output determination, we adopt the view that these influences work slowly over time and so we abstract from them in our short-run analysis.

The neo-classical production function analysis, which is standard in most textbooks, assumes that in the short-run, with all other productive inputs (capital, land etc) fixed, output will increase at a decreasing rate as more hours of employment are used by firms.

This is the so-called **Law of Diminishing Marginal Productivity**, which allows economists of this persuasion to postulate increasing marginal costs as output increases (costs increase at an increasing rate as more output is produced). In turn, this leads to an inverse relationship between labour demand and the real wage. These relationships are derived from the assumption that firms produce and employ labour such that their profits are maximised at a given price and money wage.

The validity of ‘the Law’ has been the subject of considerable controversy. In essence it is a theoretical construct – an unproven assertion. No conclusive empirical evidence has ever been assembled to substantiate ‘the Law’ as a reasonable generalisation of production relationships in modern monetary economies.

On the contrary, there is a mass of empirical evidence available, derived from actual studies of business firms, to support the view that costs of production are constant in the relevant or normal range of output and that the Law of Diminishing Marginal Productivity is not applicable.

In fact, a strong positive relationship between output per hour and the business cycle is observed in the real world. We call this a **pro-cyclical movement** in output per hour, which means that output per unit of labour input increases as the level of production and employment increases.

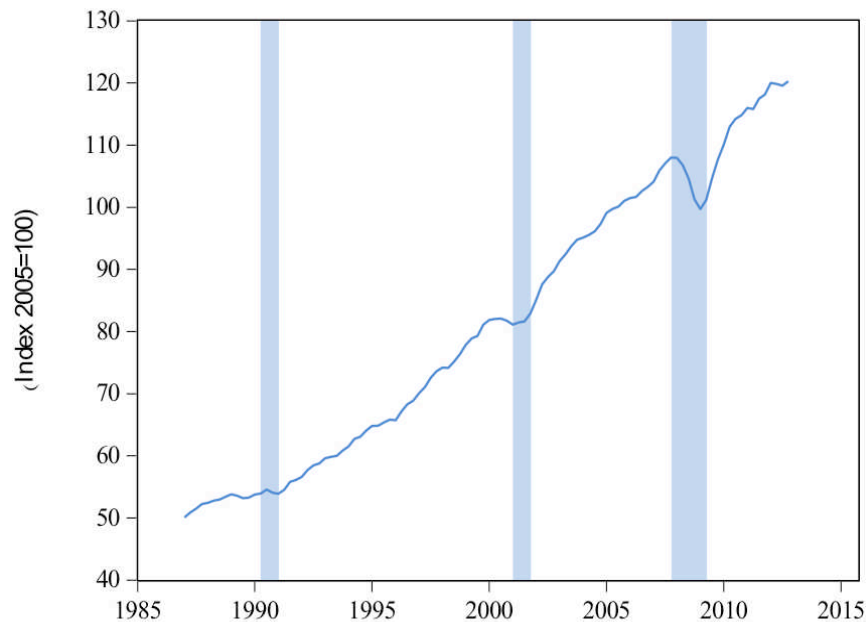
The pro-cyclical pattern of labour productivity (output per hour) means that costs per unit of output will not increase as output increases. Total costs will obviously rise but the per-unit costs will decline as the economy approaches full capacity.

Consider Figures 8.4 and Figure 8.5, which show real output per person and real output per hour in the US manufacturing sector, respectively. The shaded areas are the recessions defined by the US National Bureau of Economic Research (NBER).

Both measures of labour productivity are pro-cyclical. During a recession, when output is falling, productivity falls. This is in contradistinction to the Law of Diminishing Marginal Productivity.

The US behaves in a similar way to all advanced economies with respect to pro-cyclical movements in labour productivity in the manufacturing sector.

**Figure 8.4 US manufacturing output per person employed**



Source: US Federal Reserve Bank of St Louis. Shaded areas indicate US recessions.

If  $W$  is the money wage rate and  $N$  is total employment measured in hours, then total labour costs in any period are:

$$(8.6) \quad C = WN$$

$Y$  is real output and so **unit labour costs** ( $ULC$ ) which are the costs incurred for each extra unit of output, are given by:

$$(8.7) \quad ULC = WN/Y$$

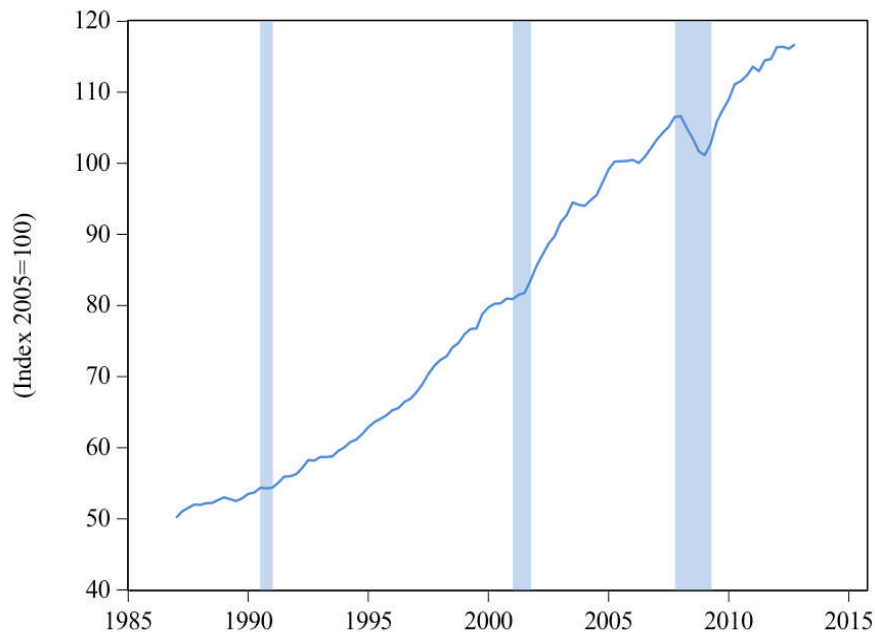
Noting that  $(Y/N)$  is output per unit of input (or labour productivity), we can re-arrange Equation (8.7) as:

$$(8.8) \quad ULC = W/(Y/N)$$

The expression for  $ULC$  shows that if the money wage rate is fixed, then changes in  $ULC$  will be driven by changes in labour productivity. In particular, if labour productivity is constant, rather than pro-cyclical, and other direct production costs (for example, raw materials) are constant per unit of output, the aggregate supply curve will be elastic at the (constant) price associated with marking-up per unit direct (labour) costs. Also the labour demand curve at both the firm and aggregate level will be elastic at the going money wage, subject to the level of aggregate demand.

This means that all the results that depend on the operation of the Law of Diminishing Marginal Productivity are no longer valid approximations of the way the economy works.

**Figure 8.5 US manufacturing output per hour of all persons**



Source: US Federal Reserve Bank of St Louis. Shaded areas indicate US recessions.

The theory of production we present here is based on several stylised facts from the real world:

1. Economies are rarely at full employment and the existing capital stock is rarely fully utilised. Idle machines typically accompany idle workers when the economy goes into a downturn.
2. The capacity of firms to substitute one input (say, labour) for another (say, capital) in the production process is limited. In the real world, a typical firm employs a number of machines and types of equipment, which have more or less fixed labour requirements.

For example, say a firm provides services from an office and each worker requires a desk, a chair and a computer to perform their duties. In the usual course of events (except when introducing an extra shift) the firm has to increase its capital and labour in the proportions defined by the technology being used to expand output.

It doesn't violate reality too much to simplify this stylised fact by assuming what economists refer to as fixed input coefficients technology.

Take a trivial example of a cleaning firm, which uses brooms as its principle technology. It services a contract to sweep rooms in office blocks each day. It is hard to imagine two workers pushing one broom or one worker pushing two brooms. So to start production, the firm needs to combine its productive inputs in a fixed ratio (in this case, 1 to 1).

If the firm gained contracts for more office cleaning which exceeded the capacity of one cleaner, then it would have to (given the technology being used) add another broom for the second worker to use and so on. So the productive inputs are added in fixed proportions defined by the technology being used.

What will the level of employment in this firm depend upon?

The firm will hire according to the demand for its services and its demand for labour will not be very sensitive to wage changes. However, it will make decisions about the viability of its operations based, in part, on wage costs. But on a day-to-day basis, if it is profitable at the current wage rates, then it will increase or decrease its demand for labour based on the revenue it can anticipate from sales.

In other words, **effective demand drives labour demand**.

Cutting wages would only redistribute total revenue towards profits, which might damage aggregate demand as workers will have less income to spend.

The neo-classical production theory, based on the Law of Diminishing Marginal Productivity, considers that firms are able to substitute labour and capital freely and if the price of labour increases in real terms, the firms will quickly use less labour and more capital.

The problem with this conception is that firms are rarely able to substitute inputs quickly and to use more capital and less labour typically requires a total change in technology. Real wage movements would have to be very large to justify the firm scrapping their existing technology.

While the fixed input ratio assumption is extreme, it shows the essential relationship between effective demand and employment. The story is not fundamentally changed if we consider the more realistic case of limited substitution possibilities.

Consider how firms might act. Based on the available technologies and the projected relative costs of labour and capital into the future, a typical firm will choose the lowest cost technology it can afford. In turn, this will set the capital-labour input ratio that it will be more or less bound by in the coming production periods. In making that decision, the firm is also committing to a certain labour demand given the relationship between the technology being used and the associated input proportions.

When demand for its product or service rises, the firm will first try to utilise existing staff and capital more fully. If the rise of demand is sustained, the firm will then increase staff and invest in additional capital equipment. The reverse sequence occurs when demand for its output falls.

In the graphs above we observed that measured labour productivity rises in expansion and falls in recession. To some extent this results because firms do not fully adjust employment to every change of demand. In expansions, workers are pressed to work faster as an alternative to hiring; in downturns, firms do not fully adjust the workforce for the fall of demand because they do not wish to lose experienced workers.

Further, if we return to our example of the cleaning firm, as it obtains more contracts for servicing it may first require the existing workforce to clean more offices per day - perhaps by a 'speed-up'. We would observe labour productivity rising. At some point, the firm must hire an additional worker (and add a broom). Measured labour productivity might be lower for a while until the firm adds enough contracts for servicing to fully utilise the larger workforce.

For these reasons, we observe a cyclical component to labour productivity that is not consistent with neoclassical production function theory. Equally the 'Law of diminishing marginal productivity' is inconsistent with the simplifying assumption of constant labour productivity, which has no cyclical component.

As a result of capital being specifically embodied in the form of machines, equipment, buildings and the like – once installed there are very few substitution possibilities.

The firm knows that if it needs to produce more output to meet the market demand then it will have to increase its demand for labour and capital, in the proportions governed by the technology in use.

Adding more labour alone will not increase output, just as adding more capital alone will not increase output.

### **How does this affect our understanding of production costs?**

In this economy, firms will adjust their input use to meet the fluctuations in demand for output. If orders decline, then their demand for inputs will decline. Both capacity utilisation and labour utilisation will decline.

But for the firm, capital becomes what economists call a free good. Relative to its purchase and installation costs the variable costs of running the capital are usually low. Economists call the major costs involved sunk, which means that the firm has already incurred them whether they run the plant or not.

Accordingly, the firm will use as much capital as is required to produce the current output that is being demanded. When demand falls, the firms simply leave some proportion of their capital stock idle.

But in doing so, they shed labour because the variable costs of the labour input are relatively high when compared to the fixed hiring and related costs.

What role does the real wage play in this? Even if the real wage fell to zero the firms would not employ more workers if aggregate demand didn't justify it. Firms will not produce if there is not a prospect of sale (barring the small proportion of production they keep as inventories to smooth out orders).

### **How would a firm react to an increase in aggregate demand?**

If there has been a prolonged downturn then we would observe idle capital and labour (unemployment). The unemployed workers are willing to work at the current wage rates but there is no demand for their services because effective demand is too low.

While we have reason to believe that unit costs decline as capacity utilisation increases, fixed factor input proportions mean that firms face constant unit costs in normal ranges of production. That is, we are assuming that money wages are fixed in the short-run and labour productivity is constant.

If the firm received increased orders for its output then it will seek to maintain its market share by increasing output. Assuming constant unit costs, the firm will bring its idle capital back into production and hire more workers.

There would be no pressure on the firm to raise prices because there would be no upward pressure on per unit costs. As output rises, the demand for labour increases at the constant real wage.

This suggests that the Aggregate Supply curve is very flat over the normal range of output. Increases in nominal demand will be met by increases in real output (income).



There are several reasons why firms might be reluctant to increase prices (even though costs might rise temporarily as we explain below) or reduce them when aggregate demand falls.

First, industries are characterised by a few dominant firms that exercise market power.

Second, consumer loyalty to products of other firms means that they will not react to a price fall in other similar products.

Third, a price cut would reduce revenue if it did not induce a sufficient number of consumers to switch brands. Further, competitors might match lower prices, to retain their consumers.

Fourth, there are significant costs involved in adjusting prices. Firms have to produce new price tags and catalogues.

### **What factors might explain the observed pro-cyclical movement in labour productivity?**

The following factors help to explain the observed pro-cyclical pattern of labour productivity.

First, a dimension of the aggregation problem appears when we consider that the value of output per hour of employment varies considerably among the range of firms and industries that comprise the total economy.

For example, the manufacture of high-tech electrical goods would have a much greater output per hour of labour input than say the provision of hairdressing services.

It can be shown that even if diminishing returns were operating at the individual firm level (by assumption, not a fact) such a constraint need not be functional at the aggregate level.

If the proportions of output attributable to individual industries change as output increases (a fact observed in the real world), and the changes are such that the industries where diminishing returns are most apparent lose a disproportionate amount of their share in total output, then labour productivity can increase. This hypothetical example merely indicates the dangers that are involved in adding up a set of non-linear relations operating at the micro level. This is beyond the level of understanding that is required to master the material in this textbook.

Second, and less esoteric, as far as the individual firm is concerned, the actual real world relationship between changes in labour hours and changes in output may not exhibit diminishing returns because other productive inputs may vary in the same proportion as the labour input.

Thus the neo-classical assertion that capital, in the form of specific plant and equipment, is always held constant confuses the distinction between the stock of capital in value terms, that is its monetary worth and the flow of services that the stock produces which is revealed by the rate of capacity utilisation.

While the stock of capital changes only slowly over time, utilisation rates can vary in the short-run. Firms will leave machines idle as their production plans are changed in the face of declining demand for their products.

Unused machines and idle factory space will not raise the productivity of the remaining machines and equipment in use.

When the firm believes that it can sell more output, production is increased and unused machines are turned back on and unemployed workers are assigned to them. There is no reason to assume

that output per unit of labour input on these machines would be any different to that derived from the plant that was kept active during the downturn.

Third, and of great practical importance, is the observation that most firms desire to maintain long-term relations with their labour forces. The reason for this behaviour by firms relates to the fixed costs of hiring (recruiting, training and redundancy provisions) and to the need to maintain morale among the workers.

Efficiency is crucially dependent on the feelings that the workers have towards job security and the like. Firms are also reluctant to dismiss specialised workers for fear of losing them permanently.

As a consequence, employment tends to fluctuate less violently than output or production. Labour productivity therefore falls in a recession and rises in booms.

In booms, output grows quickly, but the firm, which has hoarded labour in the recession, does not immediately expand employment. It merely works its existing labour force more intensively – that is, adjusts hours of work rather than persons employed.

These adjustments are reinforced by the fact that hiring and firing decisions depend on future or expected sales. Firms will increase employment of new workers and incur the fixed costs if it expects to maintain a higher sales level.

If the rise in demand is not expected to be permanent (or the firm is not sure of its durability) then it will use overtime as the cheapest means of increasing output.

So in the short-run, costs might rise as overtime premiums are paid as the firm decides whether the increase in demand is permanent or transitory.

Once it realises that the sales will remain at the higher level, costs fall again as new staff are hired and overtime declines.

# Chapter 9: Labour Market Concepts and Measurement

## Chapter Outline

- 9.1 Introduction
  - 9.2 Measurement
    - Labour force framework
    - Impact of the business cycle on the labour force participation rate
  - 9.3 Categories of Unemployment
  - 9.4 Broad Measures of Labour Underutilisation
  - 9.5 Flow Measures of Unemployment
  - 9.6 Duration of Unemployment
  - 9.7 Hysteresis
- Appendix: Advanced Material
- The collection and publication of labour market statistics
  - Labour market stocks and flows

## Learning Objectives

1. Gain an understanding of the main features of the labour force framework and the definitions of employment and unemployment.
2. Recognise the deficiencies of measuring labour underutilisation by the official unemployment rate.
3. Understand the relationships between labour market stocks and flows.
4. Recognise the importance of unemployment duration in understanding labour market hysteresis.

## 9.1 Introduction

Chapter 3 provided an outline of the evolution of economic systems from tribal societies, to slavery, feudalism and eventually to modern capitalism. An introduction to the concept of a (labour) market as a social construct with embedded power relations was developed to provide an intrinsic understanding of what happens when someone gets a job and receives a wage.

This Chapter is largely devoted to definitional and measurement issues associated with modern labour markets. We outline the labour market framework, which incorporates definition of the states of employment, unemployment and not in the labour force, which are stocks. We classify types of unemployment and argue that the rate of unemployment is an inadequate measure of labour underutilisation. We explain the relationships between our stock measures and the flows between the labour market states. We conclude with an exploration of the average duration of unemployment and its role in the process of labour market hysteresis.

## 9.2 Measurement

While up until now we have been concerned with developing a theoretical framework to explain how real GDP and national income are determined, macroeconomics is also concerned with understanding the dynamics of employment and relatedly, unemployment.

Many a textbook will say that ‘Macroeconomics is the study of the behaviour of employment, output and inflation’. Further, a central idea in economics whether it be microeconomics or macroeconomics, is efficiency, which is getting the best out of the available resources. The concept is extremely loaded and is the focus of many disputes – some more arcane than others.

At the macroeconomic level, the efficiency frontier is normally summarised in terms of full employment, which has long been a central focus of economic theory, notwithstanding the disputes that have emerged about what we mean by the term.

However, most economists would agree that an economy cannot be efficient if it is not using the resources available to it to the limit. In recent decades, the emergence of issues relating to climate change have focused our attention on what that limit actually is. In this Chapter, we focus on the use of labour resources.

The concern about **full employment** was embodied in the policy frameworks and definitions of major institutions in most nations at the end of the Second World War. The challenge for each nation was how to turn its war-time economy, which had high rates of employment as a result of the prosecution of the war effort, into a peace-time economy, without sacrificing the high rates of labour utilisation.

In this section, we outline key concepts and consider issues relating to measurement. How do we know how much employment there is at any point in time? What is unemployment? Is it a measure of wasted labour resources or are there other considerations that should be taken into account?

### Labour force framework

The Labour Force Framework constitutes a set of definitions and conventions that allow the national statisticians to collect data and produce statistics about the labour market. These statistics include employment, unemployment, economic inactivity, underemployment, which

can be combined with other survey data covering, for example, job vacancies, earnings, trade union membership, industrial disputes and productivity to provide a comprehensive picture of the way the labour market is performing.

The Labour Force Framework is a classification system, governed by a set of rules and categories. It forms the foundation for cross-country comparisons of labour market data. The framework is made operational through the International Labour Organization (ILO) and its International Conference of Labour Statisticians (ICLS). These conferences and expert meetings develop the guidelines or norms for implementing the labour force framework and generating the national labour force data.

The Australian Bureau of Statistics publication – *Labour Statistics: Concepts, Sources and Methods* – describes the international guidelines that have been agreed by the national statistical agencies. The guidelines outline the organising principles that define the Labour Force Framework. National statistical agencies work within internationally agreed standards when publishing labour statistics. In the US, the Bureau of Labor Statistics collects and publishes labour force data.

The rules contained within the labour force framework have the following features:

- an activity principle, which is used to classify the population into one of the three basic categories, namely employed, unemployed and not in the labour force.
- a set of priority rules, which ensure that each person is classified into only one of the three basic categories.
- a short reference period to reflect the labour supply situation at a specified moment in time.

The priority rules are applied to ensure that labour force activities take precedence over non-labour force activities and working or having a job (employment) takes precedence over looking for work (unemployment). Also, as with most statistical measurements of activity, employment in the informal sectors, or the black-market economy, is outside the scope of activity measures.

There is a long-standing concept of ‘gainful work’, which shapes these priorities, but this has proven controversial. Gainful work is typically seen as work for profit or pay. One can work in government or in the non-profit sector where a payment is received.

Thus a person who does ironing for a commercial laundry would be considered to be pursuing gainful work, whereas if the same person was ironing for their family they would be considered inactive. Clearly, with economic and non-economic roles being biased along gender lines, this distinction leads to an undervaluation of a substantial portion of work performed by females, as we noted in Chapter 4.

Thus paid activities take precedence over unpaid activities. Thus, for example, in Australia, persons who were keeping house, on an unpaid basis, are classified as not in the labour force, while those who receive pay for this activity are in the labour force and employed. Similarly persons who undertake unpaid voluntary work are not in the labour force, even though their activities may be similar to those undertaken by the employed.

Figure 9.1 summarises the Labour Force Framework as it applies in Australia but this is a common organising structure across all nations. National statistical agencies conduct a **Labour**

**Force Survey** (LFS) on a regular basis, usually monthly, which collects data using the concepts and definitions provided for in the Labour Force Framework.

The **Working Age Population** (WAP) typically refers to all citizens above 15 years of age. In several countries, the lower age threshold is 16 years of age. In the past, the age span was 15 years to retirement age, usually around 65 years of age. However, as social changes have seen age discrimination laws come into force in many nations, the upper age limit has been accordingly abandoned in several nations. Also, the age at which retirees can access a government provided pension has increased in a number of countries.

The WAP is then decomposed into the **Labour Force** (the ‘active’ component) and Not in the Labour Force (the ‘inactive’ component). A worker is considered to be active if they are **employed** or **unemployed**.

The proportion of the adult population who comprise the labour force is governed by the **Labour Force Participation Rate**, which is defined as:

... the ratio of the labour force to the WAP, expressed in percentage.

We will consider the cyclical behaviour of the participation rate later in the Chapter.

The ILO defines a person as being employed if:

... during a specified brief period such as one week or one day, (a) performed some work for wage or salary in cash or in kind, (b) had a formal attachment to their job but were temporarily not at work during the reference period, (c) performed some work for profit or family gain in cash or in kind, (d) were with an enterprise such as a business, farm or service but who were temporarily not at work during the reference period for any specific reason. (Current International Recommendations on Labour Statistics, 1988 Edition, ILO, Geneva, page 47).

What constitutes ‘some work’ is controversial. In Australia and the USA, for example, a person who works one or more hours a week for pay is considered employed. So the demarcation line between employed and unemployed is in fact, very thin.

Within the employment category further sub-categories exist, which we will consider later. Most importantly, significant numbers of employed workers might be classified as being underemployed, if they are not able to work as many hours as they desire, because there is insufficient aggregate demand in the economy at that point in time.

What constitutes unemployment? According to ILO concepts, a person is unemployed if they are over a particular age, they do not have work, but they are currently available for work and are **actively seeking work**. Unemployed people are generally defined to be those who have no work at all.

Unemployment is therefore defined as the difference between employment and the economically active population (civilian labour force).

Two derivative measures capture a lot of public attention. First, the **Unemployment Rate** is defined as:

... the number of unemployed persons as a percentage of the civilian labour force.

The US unemployment rate in January 2016 was 4.9 per cent. This was derived from a labour force estimate of 158.335 million and total estimated unemployment of 7.791 thousand.

Second, statisticians publish the **Employment-Population ratio**, which is:

... the proportion of an economy's working-age population that is employed.

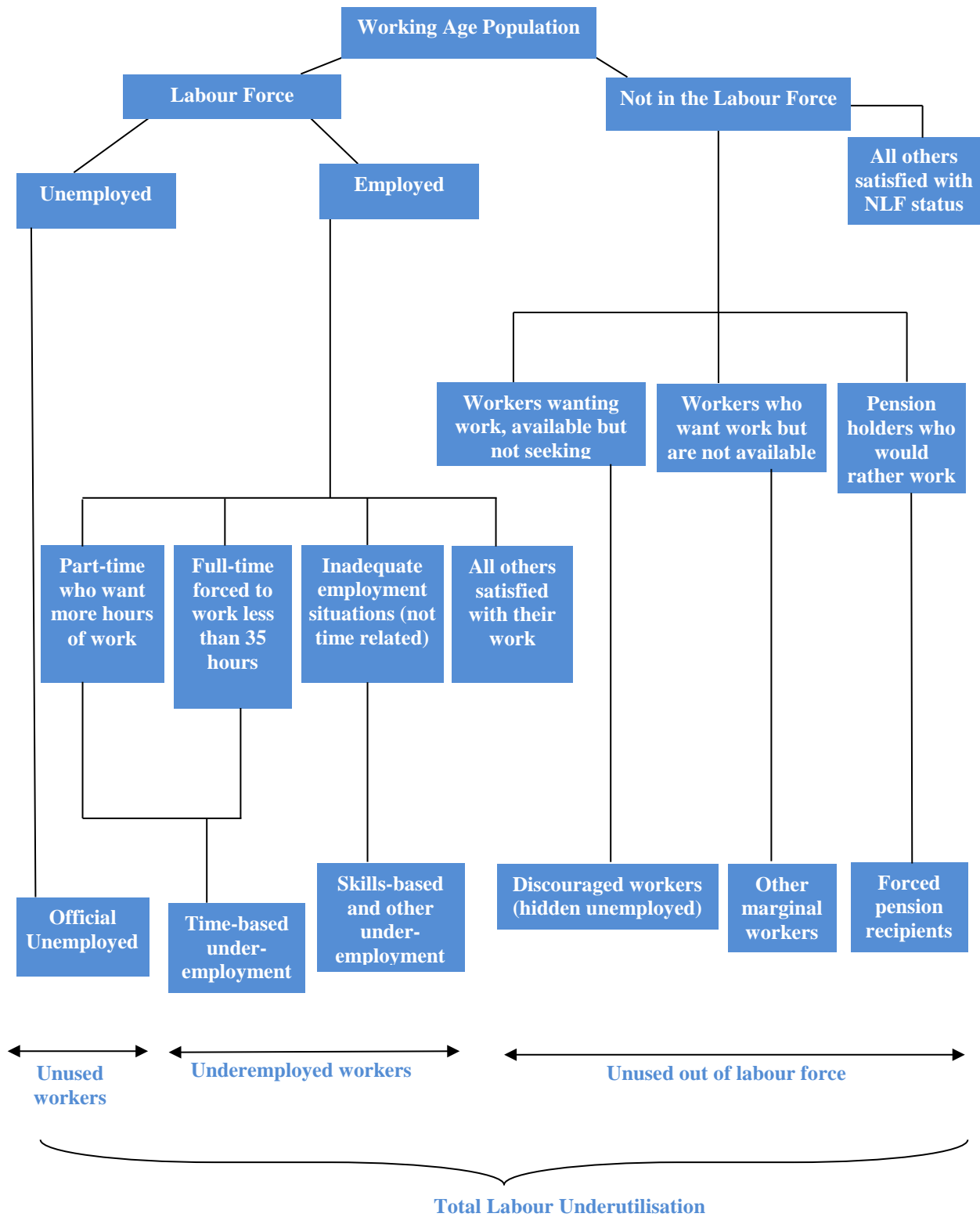
Note that the denominator of these two ratios is different. The unemployment rate uses the labour force while the employment-population ratio uses the WAP.

In the USA the employment to working-age population ratio was 59.6 per cent in January 2016.

We will see why this difference matters later when we consider the way the labour market adjusts over the economic cycle and how this impacts on our interpretation of the state of the economy as summarised by the unemployment rate and the employment-population ratio.

The unemployment measure noted above is what economists refer to as a stock measure. The unemployment rate is defined as a ratio of two stocks – the number of unemployed (numerator) and the labour force (denominator). The stock measure of the unemployment rate is compiled by the national statistician at a point in time, usually monthly.

**Figure 9.1 The labour force framework**





The category of ‘permanently unable to work’ as used in Australia would be identified with the category, not in the labour force, even though there is evidence to suggest that increasing ‘disability’ rates in some countries reflect an attempt to disguise the unemployment problem.

In terms of those out of the labour force, but marginally attached to it, the ILO states that persons marginally attached to the labour force are those who are not economically active under the standard definitions of employment and unemployment, but who, following a change in one of the standard definitions of employment or unemployment, would be reclassified as economically active.

Thus for example, changes in criteria used to define availability for work (whether defined as this week, next week, in the next 4 weeks etc ...) will change the numbers of people classified to each group. This also provides a great potential for volatility in the series and thus there can be endless argument about the limits applied to define the core series.

### **Impact of the business cycle on the labour force participation rate**

The working-age population is the population aged above the minimum working age, which is usually set at 15 years old. You have learnt that the proportion of the working-age population that offers themselves for work, which measures the labour force, is called the labour force participation rate. A change in the participation rate leads to a change in the size of the labour force.

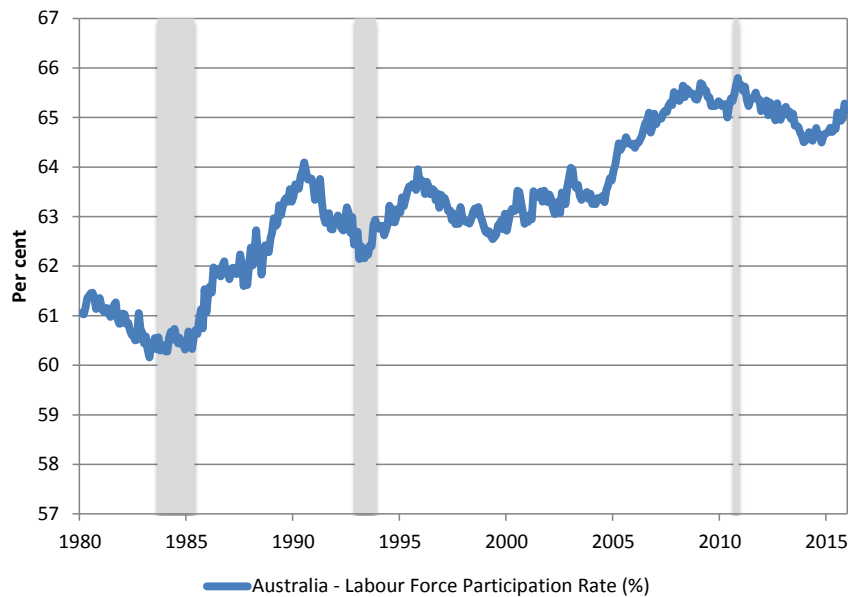
The labour force participation rate is a pro-cyclical variable – it rises in good economic times and falls when job opportunities are scarce. This means that in bad times there is likely to be a number of workers who would be willing to take job offers if they were made, but who have stopped looking for work and are classified by the national statistician as being not in the labour force. These workers who are discouraged from job search by the apparent lack of job opportunities, are considered to be hidden unemployed. From the perspective of availability, these workers are no different to the officially recorded unemployed. If a job offer was made to them they would take it immediately. This suggests that in bad times, the official unemployment rate understates to the ‘true’ underlying unemployment rate in the economy, due to the lower rate of labour force participation.

Figure 9.2 shows the Labour Force Participation Rate for Australia from the January 1980 to December 2015. The grey columns denote recessions. The pattern shown has two elements, which is common in the participation rate of most nations. First, it is clear there has been an overall upward trend in participation over this time, largely the result of the increased involvement of married women in the labour market. Second, there are distinct cyclical episodes coinciding with fluctuations in real GDP growth.

For example, in the early 1990s there was a severe recession in Australia, which precipitated a major decline in the participation rate. Participation then grew in the early 2000s with the growth in employment opportunities.

With the onset of the Global Financial Crisis in early 2007 and the slowing employment growth, the participation rate fell because job opportunities became scarcer. In late 2008, the Australian government reacted to the crisis by introducing two large fiscal stimulus packages, which promoted growth and an improvement in labour market conditions.

**Figure 9.2 Labour force participation rate, Australia, 1980 to 2015, per cent**



Source: Australian Bureau of Statistics, Labour Force Survey.

### 9.3 Categories of Unemployment

Economists have long used taxonomies to organise their thoughts about unemployment. Two often used categorisations focus on the distinctions among frictional, structural, cyclical (demand-deficient), and seasonal categories, on the one hand; and, the distinction between voluntary and involuntary unemployment on the other hand.

These taxonomies can cut across each other and no single category is better than the others. The categorisation system depends on the purpose of the analysis. In general, economists have married these categorisation frameworks into broader theoretical discussions which seek to explain why unemployment arises, whether it is a problem or not, and what can be done about it via policy interventions should we consider it to be a problem.

The most popular typology used to describe unemployment distinguishes among frictional, structural, cyclical (demand-deficient), and seasonal unemployment.

**Frictional unemployment** – recognises that the labour market is in a constant state of flux. Jobs are continually being created and destroyed, which means that workers who have been laid-off or quit are moving between jobs while firms are seeking workers for new jobs created or to fill existing jobs where the previous incumbent have left.

Further, new entrants into the labour force seek work while retirees leave jobs. Frictional unemployment arises because the matching of these demand and supply flows is not instantaneous. It takes time for workers and employers to gather relevant information and move between labour force states. Frictional unemployment is considered to be a short-term phenomenon and part of the normal functioning of the labour market. While it is debatable, this category would be expected to comprise around 1 to 2 per cent of the labour force.

**Seasonal unemployment** – arises when certain occupational skill groups and industry sectors experience fluctuations over the course of the year, which is of a systematic (seasonal) nature. For example, in certain regions, workers who are engaged in harvesting of agricultural crops will experience seasonal unemployment as they move between crops and localities. This category is considered to be small in magnitude when assessed on a macroeconomic scale. It is also difficult to distinguish from frictional unemployment.

**Structural unemployment** – is said to arise when there are enough jobs available overall to match the total pool of unemployment but there are mismatches between the skill demanded and the skills supplied and/or between the location of the jobs available and the location of the unemployment. This category of unemployment is often discussed in the context of industrial restructuring (for example, the decline of the manufacturing sector or deindustrialisation). Changes in the composition of industry employment create job losses in declining sectors and new job opportunities in emerging sectors. Further, given that industry employment is not spread evenly across regional space, the decline of a major firm in one region will have significant implications for the local labour market.

Changes in technology also have structural impacts in the sense that new skills become relevant, while old skills cease to be in demand by firms.

All of these disruptions to the pattern of employment take time to resolve. The relocation and re-training of workers displaced by structural change is sometimes a lengthy process. It is the changing pattern of required skills, the changing location of jobs and the extended time taken to resolve the resulting demand and supply imbalances that distinguishes the concept of structural unemployment from frictional unemployment.

However, there are two important qualifications to the normal conceptualisation of structural unemployment, which are not often considered in the mainstream textbooks.

First, the concept of a skills shortage is a relative concept, implying some distance from an optimal state, which begs the question: according to whom? Unsurprisingly, analyses of skills shortages by industry and governments invariably consider the issue from the perspective of business and profitability, which places the emphasis on containment of labour costs both in terms of wages and conditions, and hence, whenever possible, externalising the costs associated with developing the skills that firms require in their workers.

Within this context the notion of structural unemployment arising from ‘skills mismatch’ can be understood as implying an unwillingness of firms to offer jobs, with attached training opportunities, to unemployed workers that they deem to have undesirable characteristics. When the labour market is tight, the willingness of firms to indulge in their prejudices is more costly. However, when labour underutilisation is high, firms can easily increase their hiring standards, that is, broaden the desired characteristics they demand from workers, and the training dynamism driven by labour shortages is then absent. In this case, we observe, in a static sense, ‘skill mismatches’ which are really symptoms of a ‘low pressure’ economy.

Second, hiring standards and the willingness of firms to provide training opportunities when making job offers vary with economic activity. This means that the concept of structural unemployment is difficult to distinguish from the next category of unemployment we define, which is related to a lack of aggregate demand in the economy. When job openings are plentiful

and hiring is difficult, employers will hire workers and then train them on-the-job. Thus ‘skills mismatch’ problems are cyclical in nature.

Hence, there are significant overlaps between these categories, which reduce their capacity to provide a definitive decomposition of total unemployment.

**Cyclical (demand-deficient) unemployment** – arises when there is a shortage of jobs overall relative to the willing supply of labour resources (persons and hours) at the current wage levels. This category is termed demand-deficient unemployment because it relates to a deficiency in aggregate demand. Unemployment thus varies over the economic cycle – rising when aggregate spending falls below the level needed to fully employ the available workforce and falling when aggregate spending moves closer to the level needed to fully employ the available supply of labour.

Cyclical unemployment is also known as mass unemployment and arises when the macroeconomic system fails to generate enough jobs to match the preferences of the available workforce. It is also related to the concept of an output gap, which measures the percentage deviation of real GDP from the potential production levels at any point in time.

During an economic downturn (which may become a recession), cyclical unemployment will be the dominant proportion of measured unemployment. When economic activity improves as a result of increased aggregate demand, cyclical unemployment falls.

In Chapter 12 we will see that the economic and social costs of unemployment (associated with output gaps) are enormous, which makes the elimination of cyclical unemployment a policy imperative. The solution to cyclical unemployment is thus to increase the growth rate of aggregate demand to close any output gaps.

## 9.4 Broad Measures of Labour Underutilisation

Figure 9.1 summarised the Labour Force Framework as applied in Australia, which, is made operational through the International Labour Organization (ILO) and its International Conference of Labour Statisticians (ICLS). Thus all national statistical agencies have broadly similar structures for collecting data about the labour market.

We focus on the unemployment as an indicator of labour market performance because it signifies a waste of productive resources, quite apart from the individual and social costs that accompany it. However, unemployment is a narrow measure of labour underutilisation.

Labour underutilisation arises for a number of different reasons that can be subdivided into two broad functional categories:

- **A category involving unemployment or its near equivalent** – In this group, we include the official unemployed under ILO criteria and those classified as being not in the labour force due to failing to search for employment (discouraged workers), unavailable to start work (other marginal workers), and more broadly still, those who take disability and other pensions as an alternative to unemployment (forced pension recipients). These workers share the characteristic that they are jobless and desire work if there were available vacancies. However they fail to satisfy all the criteria for being defined as unemployed; and

- **A category that involves sub-optimal employment relations** – Workers in this category satisfy the ILO criteria for being classified as employed but suffer time-related underemployment or inadequate employment situations.

We will consider the near equivalent states of unemployment in the next section. For now we will focus on underemployment.

Within the Labour Force framework, a person of working age is considered employed if they have worked a minimum number of hours in the reference week for pay. Otherwise, they are classified as unemployed or not in the labour force, depending on how they fit into the activity criteria. In Australia and the USA, for example, a person only has to work one hour a week to be classified as employed by the Australian Bureau of Statistics and Bureau of Labor Statistics. The hours requirement differs across countries.

Underemployment may be time-related, referring to employed workers who are constrained by the demand side of the labour market to work fewer hours than they desire, or to workers in inadequate employment situations, who undertake jobs which have skill demands below their qualifications and/or where workers are forced to work longer than they desire.

Clearly, if society invests resources in education, then the skills developed should be used appropriately. The concept of an inadequate employment situation is very difficult to quantify and there is a paucity of data available as a result to measure it. However, national statisticians have developed sophisticated measures of time-related underemployment or visible underemployment.

In conceptual terms, a part of an underemployed worker is employed and a part is unemployed, even though they are wholly classified among the employed.

An economy with many part-time workers who desire but cannot find full-time work is less efficient than an economy with workers' preferences for work hours being satisfied. In this regard, involuntary part-time workers share characteristics with the unemployed.

Time-related underemployment is similar to unemployment because it arises from a deficiency in aggregate demand. Unemployment is manifested as a lack of available jobs, whereas the presence of underemployment indicates that the demand constraint rations the hours of work that are offered by firms.

In both cases, willing labour resources are wasted.

Table 9.1 shows the evolution of underemployment in a selection of OECD nations since 1990, ranked highest to lowest as at 2014.

**Table 9.1 OECD underemployment, per cent of labour force, 1990 to 2014**

	1990	2000	2005	2010	2014
Italy	1.6	2.8	4.6	6.7	10.2
Australia	3.8	5.9	6.8	7.1	8.1
Spain	1.1	1.5	3.5	5.3	7.8
Sweden	2.2	3.2	2.9	7.3	6.7
European Union	2.0	3.5	3.5	4.6	5.6
Canada	3.1	4.3	4.4	5.0	4.9
Greece	1.1	1.7	2.2	3.0	4.6
New Zealand	4.1	5.5	3.5	4.2	4.5
Japan	1.1	1.9	4.4	5.7	4.5
Netherlands	5.4	1.4	1.7	2.3	4.4
United Kingdom	1.3	2.3	1.9	3.6	4.2
Portugal	1.3	2.4	2.4	3.1	4.2
OECD countries	1.7	2.2	2.8	3.5	3.8
G7 countries	1.4	2.0	2.7	3.6	3.7
Germany	0.7	8.6	4.3	4.9	3.5
Finland		3.1	2.9	3.0	3.3
Austria		1.7	2.3	2.7	2.9
Belgium	3.0	3.6	3.2	2.5	2.2
Luxembourg	0.4	3.0	1.7	1.1	2.1
United States		0.7	0.9	1.5	1.6
Norway	3.2	1.6	2.4	1.5	1.4

Source: OECD, Share of Involuntary part-time workers in labour force.

The two concepts of underemployment are also related. The rising incidence of underemployment over the last 20 years in many countries has been associated with a rising casualisation of the workforce as governments have tilted the industrial relations playing field towards employers and reduced workplace protections and restrictions on the use of non-standard hours of work.

As a result, the quality of employment has fallen for many workers. This trend has also coincided with the growth of the service sector and in many nations (such as the US, Australia and Britain) this growth has been concentrated in lower-skilled, less stable jobs. Underemployment is common in these sectors.

We will see later that a meaningful definition of full employment has to include zero underemployment. A worker cannot be considered fully employed if they are enduring underemployment.

## 9.5 Flow Measures of Unemployment

Each period there are large numbers of workers that flow between the labour market states – employment (E), unemployment (U) and not in the labour force (N). The stock measure of each state indicates the level at some point in time, while the flows measure the transitions between the states over two periods (for example, between two months).

National statisticians measure these flows in their monthly labour force surveys. The various stocks and flows are denoted as follows (single letters denote stocks, dual letters are flows between the stocks):

**E** = employment, with subscript  $t$  denoting the current period, and  $t+1$  the next period

**U** = unemployment

**N** = not in the labour force

**EE** = flow from employment to employment (that is, the number of people who were employed last period and who remain employed this period)

**UU** = flow of unemployment to unemployment (that is, the number of people who were unemployed last period and who remain unemployed this period)

**NN** = flow of those not in the labour force last period and who remain in that state this period

**EU** = flow from employment to unemployment

**EN** = flow from employment to not in the labour force

**UE** = flow from unemployment to employment

**UN** = flow from unemployment to not in the labour force

**NE** = flow from not in the labour force to employment

**NU** = flow from not in the labour force to unemployment

Table 9.2 provides a schematic description of the flows that can occur between the three labour force framework states.

**Table 9.2 Labour market flows matrix**

	Status in Period 1		
Status in Period 0	Employed	Unemployed	Not in the Labour Force
Employed	EE	EU	EN
Unemployed	UE	UU	UN
Not in the Labour Force	NE	NU	NN

To give you some idea of the magnitude of these flows between any given months, Figure 9.3 summarises the flows for the US labour market for the period between December 2015 and January 2016. The data comes from the US Bureau of Labor Statistics.

The data shows us that total US employment in December 2015 was 149.678 million, total unemployment was 7.541 million and the number of persons who were counted as being not in the labour force was 94.495 million. The sum of these stocks is equal to the WAP (the population above the age of 16 years) of 252.397 million.

The flows data show that between the months of December 2015 and January 2016, 1.594 million workers who were unemployed in December 2015 moved into employment (UE) by January 2016. Similarly, 2.105 million workers who were counted as being employed in December 2015 moved into the unemployment pool (EU) in January 2016.

In terms of flows between the labour force and not in the labour force, there were 4.818 million workers who were counted as being employed in December 2015 who exited the labour force (EN) in January 2016 and 1.859 million workers who were counted as being unemployed in December 2015 who left the labour force (UN) in January 2016.

Flowing into the labour market, were 4.444 million new entrants who became employed (NE) and 2.099 million new entrants who ended up in unemployment (NU) in January 2016.

**Table 9.3 Gross flows in the US labour market, December 2015–January 2016, millions**

	Status Current Period			
Status Last Period	Employed	Unemployed	Not in Labour Force	Stocks Last Period
Employed	142.755	2.105	4.818	149.678
Unemployed	1.594	4.088	1.859	7.541
Not in Labour Force	4.444	2.099	87.952	94.495
Stocks Current Period	148.793	8.292	94.629	252.397

Source: US Bureau of Labor Statistics.

The final column and row show the levels of Employment, Unemployment and Not in the Labour Force, corresponding to December 2015 and January 2016, respectively. They are obtained from the row sums and the column sums. Employment fell over the month, which reflects the seasonal nature of employment in the winter, when some outdoor work cannot be conducted.

It is important to recognise that the Table 9.3 tracks the labour force status of the 252.397 million US citizens who were part of the WAP in both December 2015 and January 2016. It does not include individuals who joined the WAP in January 2016, due to age or moving to the USA and those left the WAP, due to death or departure from the USA.



We can also calculate the total inflows and outflows from the three labour force states between any two periods of interest. Table 9.4 shows these calculations for the above data.

The total inflow into employment is measured by the sum,  $NE + UE$  and for the period shown equalled 6.038 million whereas the total outflow from employment, measured by the sum,  $EU + EN$  was 6.923 million. The net flow was thus negative and equals 0.885 million workers. This confirms that employment between December 2015 and January 2016 fell.

The total inflow into unemployment is measured by the sum,  $EU + NU$  and for the period shown equalled 4.204 million whereas the total outflow from unemployment, measured by the sum,  $UE + UN$  was 3.453 million. The net flow was thus positive (meaning that unemployment rose over the period) and was equal to 0.751 million workers.

**Table 9.4 Total inflow and outflow from labour force states, US, December 2015 to January 2016, millions**

Labour Market State	Total Inflow December 2015 to January 2016 millions	Total Outflow December 2015 to January 2016 millions
Employment	$UE + NE = 6.038$	$EU + EN = 6.923$
Unemployment	$EU + NU = 4.204$	$UE + UN = 3.453$
Not in the Labour Force	$EN + UN = 6.677$	$NE + NU = 6.543$

Source: US Bureau of Labor Statistics.

Finally, the total exits from the labour force (into Not in the Labour Force) is measured by the sum,  $EN + UN$  and for the period shown equalled 6.677 million whereas the total new entrants into the labour force, measured by the sum,  $NE + NU$  was 6.543 million. The net flow was thus positive and equal to 0.134 million workers.

## 9.6 Duration of Unemployment

As noted, the unemployment rate is considered to be a narrow measure of labour market performance. Another dimension of labour underutilisation, which it does not capture, is the duration of unemployment. As the discussion of flows indicated, the labour market is a very dynamic part of the economy with large flows between the labour force states occurring on a weekly basis. The magnitude of these flows is, however, highly cyclical and net flows into unemployment are larger during a recession than in other times.

It is therefore important to consider the average duration of unemployment as part of our assessment of the state of the labour market.

The Australian Bureau of Statistics – Labour Statistics: Concepts, Sources and Methods – provides the following definition:

Duration of unemployment is defined as the elapsed period to the end of the reference week since the time a currently unemployed person began looking for work, or since a person last worked for two weeks or more, whichever is the shorter. Brief periods of work (of less than two weeks) since the person began looking for work are disregarded.

This conceptualisation is representative across nations even if there are some country-by-country variations in how the Labour Force Survey is collected.

The duration of unemployment influences the way we assess the distributional impacts of a recession. If for example, individuals who become unemployed only endure short spells of unemployment – that is, average duration in weeks is low – then the impact on their income flow and accumulated saving will be lower than if the spells of unemployment are longer. A drawn-out recession typically has the effect of wiping out any savings that the unemployed person may have accumulated.

For a given unemployment rate, an economy might be characterised by a predominance of short spells of unemployment (many people flowing in and out of the unemployment pool) or at the other extreme, the same people enduring long spells of unemployment (low inflows into and low outflows to the unemployment pool).

While any unemployment above some irreducible minimum rate is problematic, clearly the situation where individuals experience unequal durations of unemployment is longer is more costly.

As an example, assume an economy has a labour force of 100 persons and is enduring an unemployment rate of 8 per cent. This might occur if 8 individuals had become unemployed at the beginning of the month but who will find work in the following month. Next month, 8 different individuals become unemployed. Each individual has a duration of unemployment of one month.

On the other hand, the same economy might have the same 8 individuals enduring unemployment month-after-month and still maintaining an unemployment rate of 8 per cent. Thus 8 individuals had 12 months of continuous unemployment, whereas the remaining 92 individuals remained employed for the whole year.

The duration of unemployment displays distinct cyclical patterns. As economic activity starts to slow and enters recession, there are large flows into the unemployment pool and so short-term unemployment surges. So the overall pool of unemployment is more weighted to individuals with short duration spells of unemployment.

As the recession endures and the net inflows into unemployment remain positive but start to decrease, more workers move into longer duration categories of unemployment and long-term unemployment increases. The average duration of unemployment starts to rise more sharply at this stage. The longer the recession the higher will be the long-term unemployment rate. In the USA long term unemployment is defined as a duration of unemployment of 6 months or more, whereas in Australia and the UK it is defined as a duration of 12 months or more.

This pattern endures even after the economy is recovering. As the flows into unemployment start to fall, the pool of unemployment is now more heavily weighted by individuals with longer spells of unemployment. As a result, the average duration of unemployment continues to increase even though the unemployment rate might start falling.

The problem is that in the early stages of the recovery, employment growth has to be strong enough to absorb the new entrants into the labour force (that is, keep pace with the underlying population growth) and start eating into the huge pool of unemployed. There is evidence, which we discuss later that suggests in the early stages of the recovery, firms prefer to employ workers

who have only endured short spells of unemployment. In other words, the longer a person has been unemployed, the lower will be the probability of them gaining work.

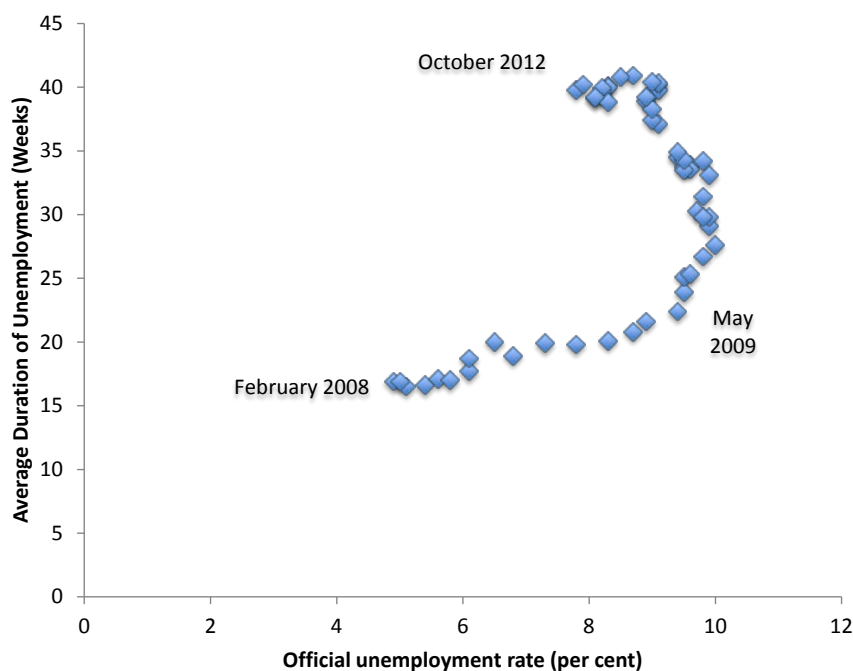
Figure 9.3 plots data from the US Bureau of Labor Statistics to illustrate the way that the average duration of unemployment behaves during a downturn and early stages of recovery.

In February 2008, the official US unemployment rate was 4.9 per cent and the average duration of unemployment was 16.9 weeks.

In the first 12 months of the downturn, the unemployment rate increased by 2.9 percentage points and the average duration of unemployment rose by 2.9 weeks.

However in the second-year of the downturn, the unemployment rate increased by 1.4 percentage points but the average duration of unemployment rose by 10.2 weeks. Even as the unemployment rate started to decrease in the third year of the crisis (by -0.7 percentage points), the average duration of unemployment continued to increase by 7.3 weeks.

**Figure 9.3** Unemployment rate and average duration of unemployment (weeks), US, February 2008 to October 2012



Source: US Bureau of Labor Statistics.

## 9.7 Hysteresis

One of the reasons we worry about situations where the duration of unemployment is high for extended periods relates to the concept of path-dependence or hysteresis.

Hysteresis is a term drawn from physics and is defined by the Oxford Dictionary as:

... the phenomenon in which the value of a physical property lags behind changes in the effect causing it, as for instance when magnetic induction lags behind the magnetizing force.

In economics, we sometimes say that where we are today is a reflection of where we have been. That is, the present is path-dependent or history matters. We will consider this effect in Chapter 11 Unemployment and Inflation because it has implications for how we conceptualise an unemployment rate that is consistent with stable inflation.

We will learn that the hysteresis effect describes the interaction between the actual and equilibrium unemployment rates. The significance of hysteresis is that the unemployment rate associated with stable prices, at any point in time should not be conceived of as a rigid non-inflationary constraint on expansionary macro policy. The equilibrium rate itself can be reduced by policies, which reduce the actual unemployment rate.

For the discussion in this Chapter we will confine ourselves to the way the economic cycle impacts on hiring in the labour market.

A recession causes unemployment to rise, and if it is prolonged, the short-term joblessness becomes entrenched long-term unemployment, as we noted in the previous section. Thus we would observe a rising average duration of unemployment as the number of long-term unemployed workers rises.

However, the unemployment rate behaves asymmetrically with respect to the economic cycle, which means that it jumps up quickly but takes a long time to fall again.

There is robust evidence pointing to the conclusion that a worker's chance of finding a job diminishes with the length of their spell of unemployment. When there is a deficiency of aggregate demand (and hence lots of unemployed workers seeking jobs), employers use a range of screening devices when they are hiring. These screening mechanisms effectively 'shuffle' the unemployed queue, with the least desired workers relegated to the back of the queue.

Among other things, firms increase hiring standards (for example, demand higher qualifications than are necessary) and may engage in petty prejudice. A common screen is called statistical discrimination whereby the firms will conclude, for example, that because, on average, a particular demographic cohort has higher absentee rates (for example), every person from that group must therefore share those negative characteristics. Personal characteristics such as gender, age, race and other forms of discrimination are used to shuffle the disadvantaged workers to the end of the queue.

In this context, the concept of hysteresis relates to how the labour market adjusts over the economic cycle. In a recession, many firms disappear altogether, particularly those which were using very dated capital equipment that was less productive and hence subject to higher unit costs than the best practice technology.

The skills associated with using that equipment become obsolete as it is scrapped. This phenomenon is referred to as skill atrophy. Skill atrophy extends beyond the specific skills needed to operate a piece of equipment or participate in a firm-specific process.

Long-term unemployment also erodes more general skills as the psychological damage of unemployment impacts on a worker's confidence and bearing. A lot of information about the labour market is gleaned informally via social networks and there is strong evidence pointing to

the fact that as the duration of unemployment becomes longer, the breadth and quality of an unemployed worker's social network falls.

New entrants to the labour force enter the unemployment pool because of a lack of jobs and are denied relevant skills (and socialisation associated with stable work patterns).

Further, as training opportunities are typically provided with entry-level jobs, it follows that the (average) skill of the labour force declines when vacancies fall.

As a result, both groups of workers – those who have lost their jobs and the new entrants – need to find jobs in order to update and/or acquire relevant skills. Skill (experience) upgrading also occurs through mobility between jobs, which is restricted during a downturn.

Therefore, workers who have endured shorter spells of unemployment, other things equal, will tend to be closer to the front of the queue. Firms form the view that those who are enduring long-term unemployment are likely to be less skilled than those who have just lost their jobs and with so many workers to choose from firms are reluctant to offer any training.

However, just as the downturn generates these skill losses, and imposes longer durations of unemployment on certain groups, a growing economy will start to provide training opportunities and the unemployment queue diminishes. This is one of the reasons that economists believe it is important for the government to stimulate economic growth when a recession is looming to ensure that the skill acquisition can occur more easily.

As demand picks up and the pool of unemployed workers shrinks, employers find they must be much less picky. 'Tight-full' employment has been defined as a position in which there are more vacancies than there are unemployed desiring jobs. In such a condition, even the long-term unemployed have a chance to obtain work. Groups that had unfairly faced racial, ethnic, or gender bias and who were at the back of the labour queue now have the opportunity to prove themselves. Maintaining tight-full employment helps to reduce the likelihood that employers will indulge their irrational biases.

## **Appendix: Advanced Material**

### **The collection and publication of labour market statistics**

At the end of the First World War, the ILO was established (1919) to set minimum labour standards. Each year, there is an International Labour Conference that makes decisions that determine what are called the International Labour Conventions and Recommendations.

One section of these conventions, the – Labour Statistics Convention (No. 160) –was adopted at the 71st International Labour Conference in 1985 and modernised the previous convention that was agreed in 1938.

Article 1 of the 1985 Convention requires all member states of the ILO (including Australia and the US) to:

- a) ... regularly collect, compile and publish basic labour statistics, which shall be progressively expanded in accordance with its resources to cover the following subjects:
- b) economically active population, employment, where relevant unemployment, and where possible visible underemployment;

- c) structure and distribution of the economically active population, for detailed analysis and to serve as benchmark data;
- d) average earnings and hours of work (hours actually worked or hours paid for) and, where appropriate, time rates of wages and normal hours of work;
- e) wage structure and distribution;
- f) labour cost;
- g) consumer price indices;
- h) household expenditure or, where appropriate, family expenditure and, where possible, household income or, where appropriate, family income;
- i) occupational injuries and, as far as possible, occupational diseases; and
- j) industrial disputes.

The ILO also publish very detailed technical guidelines about how these statistics should be collected and disseminated via one of its technical committees – the International Conference of Labour Statisticians (ICLS). This Committee meets about every five years and its membership comprises government officials who are ‘mostly appointed from ministries responsible for labour and national statistical offices’ and representatives from employer’s and worker’s organisations.

The ICLS agree on resolutions, which then determine the way in which the national statistical offices collect and publish data. While the national statistical agencies have some discretion as to how they undertake the task of preparing labour statistics, in general, there is widespread uniformity across agencies.

Labour statistics are often drawn into political controversies and government critics have been known to accuse the government of manipulating the official data for political purposes. But once you understand the process that governs the structure of the labour force statistical collection and the definitions outlined in the ICLS resolutions, it is hard to believe that argument.

This is not to say that there is not a lot of debate about what the official labour statistics measure and whether they can be improved, but it is important to understand how they are collected.

### **Labour market stocks and flows**

We can understand changes in the stock measures associated with the labour market states from one period to the next by considering the net flows between two periods.

Total employment at any point in time ( $E_t$ ) is given by the following expression, which is strictly an identity, since an adult member of the population can only be in one of three labour market states at a point in time:

$$(9.1) \quad E_t \equiv E_{t-1} + UE_t + NE_t - EU_t - EN_t$$

In terms of the actual flows in the US labour market between December 2015 and January 2016 summarised in Tables 9.3 and 9.4, Equation 9.1 is evaluated as (in millions):

$$(9.1a) \quad 148.793 = 149.678 + 1.594 + 4.444 - 2.105 - 4.818$$

The change in employment in any period,  $\Delta E$  is the total inflows minus the total outflows:

$$(9.2) \quad \Delta E = E_t - E_{t-1} = UE_t + NE_t - EU_t - EN_t$$

Total unemployment at any point in time ( $U_t$ ) is given by the following expression:

$$(9.3) \quad U_t = U_{t-1} + EU_t + NU_t - UE_t - UN_t$$

Equation 9.3 is evaluated as (in millions):

$$(9.3a) \quad 8.292 = 7.541 + 2.105 + 2.099 - 1.594 - 1.859$$

Thus the change in unemployment in any period,  $\Delta U$  is the total inflows minus the total outflows:

$$(9.4) \quad \Delta U = U_t - U_{t-1} = EU_t + NU_t - UE_t - UN_t$$

We can use the data from Table 9.3 to calculate so-called transition probabilities, which are the probabilities that transitions (changes of state) occur. These are obtained by dividing the elements of a row of the 3\*3 flow matrix by the corresponding row sum.

**Table 9.5 Labour market state transition probabilities, US, December 2015 to January 2016**

	Status Current Period		
Status Last Period	Employed	Unemployed	Not in Labour Force
Employed	0.95	0.01	0.03
Unemployed	0.21	0.54	0.25
Not in Labour Force	0.05	0.02	0.93

The interpretation of say 0.03 in the first row is that there is a probability of 0.03 that an individual who was employed in December 2015 left the Labour Force over the following month. Allowing for rounding errors the rows each sum to unity.

Economists thus consider the labour market to be very dynamic and the extent of this dynamism is measured by the gross flows between the three labour market states, which is also revealed by the transition probabilities between the three states, as opposed to those probabilities on the main diagonal of the transition matrix which measures the probability of remaining in the same labour market state.

Further, these flows are highly cyclical. For example, in a recession the flow EU increases while the flow UE declines. Workers also drop out of the labour force in greater numbers during a recession, so that labour force participation drops, and more new labour market entrants end up unemployed than in employment.

## References

Australian Bureau of Statistics (ABS), Labour Statistics: Concepts, Sources and Methods found at: [www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/973128B0B872F469CA2572C100244952](http://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/973128B0B872F469CA2572C100244952)

International Labour Organization (ILO) Labour Statistics Convention (No. 160) found at: [http://www.ilo.org/dyn/normlex/en/f?p=1000:12100:0::NO::P12100\\_INSTRUMENT\\_ID:312305](http://www.ilo.org/dyn/normlex/en/f?p=1000:12100:0::NO::P12100_INSTRUMENT_ID:312305)

International Labour Organization (ILO) (1988) Current International Recommendations on Labour Statistics, 1988 Edition, ILO, Geneva, page 47.

US Bureau of Labor Statistics (2016) 'Labor force status flows by sex, January 2016', Labor Force Statistics from the Current Population Survey, [http://www.bls.gov/web/empsit/cps\\_flows\\_current.htm](http://www.bls.gov/web/empsit/cps_flows_current.htm) (checked 10 February 2016).



# Chapter 10: Money and Banking

## Chapter Outline

10.1 Introduction

10.2 Some Definitions

- Money supply

10.3 Financial Assets

10.4 What do Banks Do?

- The orthodox view: the money multiplier

- MMT representation of the credit creation process

- Summary

- A bank's credit creation: A balance sheet analysis

Appendix: Advanced Material

- Theories of the yield curve and its dynamics

## Learning Objectives

1. Gain an understanding of definitions of the money supply and financial assets.
2. Recognise the sharp distinction between the MMT and orthodox representations of the process of credit creation by banks.
3. Be able to interpret a bank balance sheet and incorporate changes via flows of new transactions.

## 10.1 Introduction

In this Chapter we have a number of objectives. First we will introduce students to commonly used definitions of the **money supply**. Frequent reference has been made in earlier chapters to the purchase or sale of financial assets by both the government through the central bank and treasury, as well as by banks. Here we will provide students with a clear understanding of the generic characteristics of **financial assets**. We then devote space to the development of an understanding as to how banks behave in a modern monetary economy. In the process we will expose some long-standing myths about the role that banks play in the operation of the financial system. This analysis is complemented by the use of balance sheets in the following sub-section of the Chapter, which adds further clarity.

## 10.2 Some Definitions

### Money supply

Economists and commentators try to draw inferences about the economy from trends over time in particular monetary aggregates. There are a number of measures of the so-called money supply, which have been devised over the years, but there is some variation across countries. For example, the Australian Bureau of Statistics (2012) defines the following:

The **monetary base** (HPM) comprises holdings of notes and coin by the private sector; deposits of banks with the Reserve Bank (that is, reserves), and other Central Bank liabilities to the private sector. In Australia and the UK, but not the USA, HPM is defined as ‘narrow money’ or ‘M0’ and is the most liquid measure of the money supply. On the other hand, the US definition of M0 excludes bank reserves.

M1 is defined as currency (that is, notes and coins) plus current bank deposits held by the private non-bank sector. This measure is used by economists trying to quantify the amount of money in circulation. The M1 is a very liquid measure of the money supply, as it contains cash and assets (bank deposits) that can quickly be converted to currency. Typically, central banks issue notes as currency while treasuries issue coins as currency - however, it was common in the past for treasuries to also issue notes.

M3 is defined as M1 plus all other Authorised Deposit-taking Institution (ADI) deposits from the private non-bank sector plus certificates of deposit issued by banks less ADI deposits held with each other. ADIs include banks, building societies and credit unions.

Broad money is the widest definition of money published by the Reserve Bank of Australia, the central bank. Broad money is defined as M3 plus borrowings from the private sector by non-bank financial intermediaries (including cash management trusts) less their holdings of currency and bank deposits. This measure is generally used to estimate the entire supply of money within an economy.

The US Federal Reserve defines M2 as M1 + most savings accounts, money market accounts, retail money market mutual funds, and small denomination time deposits (certificates of deposit of under \$100,000). M2 has been typically used to forecast inflation.

### 10.3 Financial Assets

If a household engages in saving (a flow per period of time), over a number of months or years, then it will accumulate a growing stock of wealth over time. The household needs to make a decision about whether to continue to add its saving to its existing deposits at its bank or put together a portfolio of financial assets, which have different degrees of risk - for example, stocks (shares) or bonds (see below), which are also denominated in the money of account.

Treasuries in modern economies issue bonds (debt), which are financial assets bought and sold by the central bank, banks and the private sector. In Australia these bonds are also known as Commonwealth Government Securities (CGS).

These bonds acknowledge that the **issuer** is indebted (owes money) to the **bondholder** who buys the new bonds in the **primary** market. The bond is negotiable because ownership of the certificate can be transferred (sold) to another owner in the **secondary** market. Clearly secondary market trading has no impact at all on the volume of financial assets in the system since it just shuffles this wealth between wealth-holders.

The bond issuer must pay interest to the bondholder normally semi-annually and repay the principal later, when the bond matures. Bonds represent wealth for bondholders.

Thus a **bond** is a **formal contract** to repay a loan (IOU) with interest at fixed intervals. The bondholder is the lender (creditor). The borrower (debtor) issues the bond and the **coupon** is the interest rate, which is paid on the **face value** of the bond, which is printed on the bond. This means that semi-annual interest payments, say  $\$R$ , are constant.

The **issue price** is what investors pay for the bond when it is first issued, and is about equal to the face value on the bond. Later on, bonds may be traded: at a premium (above par, if good quality, so that there is minimal default risk by the issuer), or at a discount (price below par).

A **consol** is a perpetuity, so there is no maturity date. Interest is paid on this asset forever. Assume  $R$  is the annual interest payment, and  $r$  is the market rate of interest per annum, then the sale price,  $P$ , in the absence of risk, can be shown to be  $P=R/r$ . This is the present value of the stream of interest payments,  $R$ , forever.

**A high sale price  $P$  means a low rate of return or yield  $r=R/P$  (with  $R$  fixed) and conversely. So there is an inverse relationship between market (sale) price of a consol and its yield (rate of interest).**

This inverse relationship applies to bonds of shorter maturity, as well, but the algebra is a little bit more complicated. This is a very important relationship.

A relatively safe bond, with a low risk of default by the borrower, will attract a high auction price and a low yield (interest rate) since **coupon  $R$  is fixed** and the implied rate of interest,  $r \sim R/P$ .

A risky bond is less attractive to investors who would pay a lower price for it. They earn a higher yield, which incorporates a risk premium.

Treasury bonds (debt) are sold by auction, where banks and other institutions bid for them.

Ten-year treasury bonds issued in countries, such as Australia, USA, UK and Japan are very safe, and have attracted a relatively low yield (rate of interest) of under 5.5 per cent since 2008.

These countries issue their own currency. On the other hand, some of the Eurozone countries, including Greece, Portugal and Ireland, experienced ten-year bond rates in excess of 10 per cent over this period. These countries use a foreign currency, the Euro.

Bonds are also issued and sold in primary markets by state or provincial governments, multinational and local companies, credit institutions and other public bodies. Companies raise finance for new capital investment by: (i) issuing bonds; (ii) using retained profits, and (iii) undertaking a new share issue.

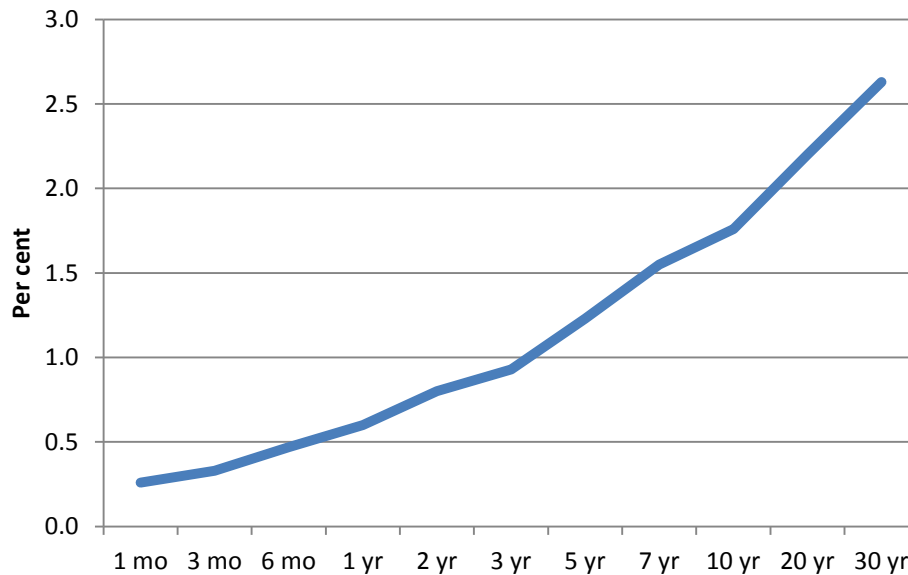
Treasuries and other institutions issue bond with different times to maturity. For example the US Department of the Treasury issues bonds of 1 month, 3 months, 6 months, 1 year, 2 year, 3 year, 5 year, 7 year, 10 year, 20 year and 30 year duration. We can use the term maturity and term interchangeably. So a 10-year treasury bond matures in 10 years.

There are two ways we can use data on yields for bonds of different maturities but of similar risk (see Mitchell, 2011b). First, we would see what is happening to the demand for bonds from investors. Rising yields signal falling demand. This could be a reflection of a strengthening economy with investors being prepared to acquire more risky assets and less very safe ones. This is also usually when the central bank pushes up the target interbank rate and bond yields more or less follow (see Chapter 15). Second, we will get an indication of what is happening to inflationary expectations and risk-assessments. We could graph a time series of yields for selected treasury bonds of different maturities say for the period of 1 January 2015 to 31 December 2015.

The second way of looking at the yields is to consider the **yield curve**. The yield curve is a graphical depiction of the **term-structure of risk-free interest rates** and plots the maturity of the government bond on the horizontal axis against the respective yields (rates of return) on the vertical axis. There are broadly three shapes that the curve will take:

- **Normal** – Under normal circumstances, short-term bond rates are lower than long-term rates. The central bank attempts to keep short rates down to keep levels of activity as high as possible and bond investors desire premiums in longer-term maturities to protect them against inflation. Thus, the yield curve is upward sloping.
- **Inverted** – Sometimes, short-term rates are higher than long-term rates and the yield curve is said to be inverted. Usually the economy starts to overheat and expectations of rising inflation lead to higher bond yields being demanded. The central bank responds to building inflationary pressures by raising short-term interest rates sharply. Although bond yields rise, the significant tightening of monetary policy causes short-term interest rates to rise faster, resulting in an inversion of the yield curve. The higher interest rates may then lead to slower economic growth.
- **Flat** – A flat yield curve is seen most frequently in the transition from positive to inverted, or vice versa. As the yield curve flattens the yield spreads drop considerably. A yield spread is the difference between, say, the yield on a one year and a 10-year bond. What does this signal about the future performance of the economy? A flat yield curve can reflect a tightening monetary policy (short-term rates rise). Alternatively, it might depict a monetary easing after a recession (easing short-term rates) so the inverted yield curve will flatten out (Mitchell, 2011b).

**Figure 10.1 US Treasury yield curve (February 3, 2016)**



Data Source: Daily Treasury Yield Curve Rates, Resource Center, US Department of the Treasury.

In Figure 10.1 we show the US Treasury Yield Curve for 3 February 2016, which conforms to the normal upward sloping form. This follows the Federal Open Market Committee's decision to raise the Federal Funds rate 25 basis points to a range of 25 to 50 basis points (0.25 percent to 0.50 percent) in December 2015. We will examine the impact of this decision on the yield curve in Chapter 15 Monetary Policy in Sovereign Nations.

We explore some factors that influence the shape of the Yield Curve in Advanced Material in the Appendix.

## 10.4 What do Banks Do?

### The orthodox view: the money multiplier

In most textbooks, banks are presented as financial intermediaries that take in deposits, hold a small fraction of these in the form of reserves, and then lend out the remainder. If each bank follows these principles in making loans, aggregate lending expands through the 'deposit or money multiplier'. For the moment assume that all banks are required to hold reserves to deposit ratio of 10 per cent. This is designed to enable them to readily respond to a loss of reserves resulting from spending by customers on say goods and services, whose sellers bank elsewhere and also those customers who seek to hold additional cash. We now outline the operation of the money multiplier (see also Mitchell, 2011a):

- i) Assume that a customer deposits say \$100 in Bank A;
- ii) Bank A retains \$10 of reserves to conform to the required reserves to deposit ratio of 0.1. To expand its loan portfolio and increase profits, the remaining \$90 is loaned to a customer whose deposits rise by \$90;
- iii) The customer spends these deposits and the recipient of the funds (seller) deposits \$90 in their bank, which for generality we will assume is Bank B;

- iv) Bank B then lends 0.9 times \$90 = \$81 (keeping 0.10 that is, \$9 as additional reserves as required) to a customer to finance their expenditure and so on.

In each stage the amount lent and then spent diminishes. It can be readily shown that \$900 of additional loans are created, which, with the initial new deposit, means that deposits have risen by a total of \$1000, which are 'backed' by \$100 of reserves, thereby conforming to the required 10 per cent ratio.

This example is what the mainstream textbooks call a fractional-reserve banking system and it purports to explain how banks create money, which increases the money supply, such as M1, due to the increase in current deposits. The multiplier, in terms of the initial deposit of \$100, is 10, which is the inverse of the required reserves to deposit ratio of 0.1. A smaller money multiplier results if the non-Government sector chooses to hold more cash when credit is created.

The standard example is typically assumed as a 10 per cent ratio, so that students could readily calculate a money multiplier equal to 10! On 12 April 1992, the US Fed, for the first time, set the required reserve ratio on demand deposits at the magical 10 per cent, making theory appear to coincide with reality.

By way of summary, banks are considered to be financial intermediaries that maximise profits. They take in deposits to build up reserves so that they can then on-lend the deposits at a higher interest rate. However prudential regulations require that they maintain a minimum reserve to deposit ratio. The fractional reserve requirements mean that the resulting credit creation process is finite.

In addition, many economists still believe that High Powered Money (HPM), which consists of bank reserves and cash held by the non-government sector, is under the control of the central bank. Thus by controlling the size of the stock of HPM and setting the required reserve ratio, the central bank controls the size of the money supply or quantity of money.

Money is said to be 'exogenous' in the control sense, determined by the central bank. This has been called the 'verticalist' approach, because in most textbooks the money supply is presented as 'vertical' (perfectly inelastic with respect to interest rates), where the supply of money (horizontal axis) is plotted against the interest rate (vertical axis). As we will see in Chapter 11, the quantity of money is alleged to determine the rate of inflation, under the Quantity Theory of Money.

The implication of the operation of the money multiplier is that a bank would forego profitable loan opportunities, if it did not have sufficient reserves to enable additional credit creation. Some allowance is made for discretion: the deposit multiplier is claimed to be a function of interest rates and interest rate differentials, bank preferences regarding their holdings of excess reserves, and also public preferences regarding their holdings of cash, as noted, and time deposit and demand deposit ratios. However, as Brunner (1968) 'demonstrated', these factors are of only minor importance.

### **MMT representation of the credit creation process**

We shall now argue that this characterisation of the credit creation process, which is driven by fractional reserve requirements, is not an accurate depiction of the way banks operate in a modern monetary economy characterised by a fiat currency and a flexible exchange rate.

In reality, the business of banking is complicated but is, in some respects, similar to that of other profit-seeking firms. Like non-bank firms, banks are seeking to earn profits and thereby generate returns for shareholders. Making loans secures profits as long as the banks are paying a lower rate of interest on funds that they borrow than they receive from their customers who take out loans.

First, a necessary condition for credit creation is that there are non-bank firms and/or households who are seeking loans to finance their planned spending on goods, services or assets. Second, some of these entities must be considered creditworthy by the banks, so that there is a high probability that the loan will be repaid in full. What constitutes credit-worthiness varies over the business cycle and lending standards tend to become more lax at boom times as banks chase market share. Third, the banks must assume that there is profit to be made by making these loans, as described above.

Banks make loans independently of their reserve positions (that is, their holdings of reserves, relative to their liabilities) and then borrow additional reserves if required. Bank managers generally neither know nor care, about the aggregate level of reserves in the banking system. Certainly, no loan officer ever checks the individual bank's reserve position before approving a loan. Bank lending decisions are affected by the price of reserves and expected returns, not by reserve positions. If the spread between the rate of return on an asset (loan) and the interbank rate is wide enough, even a bank that is already deficient in reserves will purchase the asset and cover the reserves needed by purchasing (borrowing) reserves in the interbank market.

The important point is that when a bank originates a loan to a firm or a household it is not lending reserves. Bank lending is not easier if there are more reserves, just as it is not harder if there are less. Bank reserves do not fund money creation in the way that is claimed in the money multiplier and fractional-reserve deposit story (Mitchell, 2011a).

The main difference between banks and other types of firms involves the nature of the liabilities. Banks 'make loans' by purchasing IOUs of 'borrowers'. This results in a bank liability, usually a demand deposit, at least initially, that shows up as an asset ('money') of the borrower. Thus a customer of a bank who secures a loan is simultaneously a 'creditor' of the bank, due to holding a demand deposit, but is also a 'debtor' to the bank. These creditors will almost immediately exercise their right to use the newly created demand deposits as a medium of exchange for purchases of goods and services. Bank liabilities (bank deposits) are the money used by households and non-bank firms for transactions in the form of cheques or transfers or are first redeemed at par (\$ for \$) against fiat money (which is guaranteed by the government) to enable cash to be used. The government will also accept some bank liabilities in payment of taxes.

In turn, bank reserves are the 'money' used as means of payment (or interbank settlement) among banks and for payments made to the central bank. Thus, when bank 'creditors' draw down their demand deposits, by either spending or choosing to hold more cash, this causes a corresponding loss of reserves for the individual bank. The bank may then either sell an asset, or increase its liabilities by borrowing additional reserves, in order to cover the loss of reserves. In the aggregate, however, such activities only shift reserves from bank to bank.

The interbank market (say the federal funds market in the US) functions to shuffle the reserve balances that the member (private) banks keep with the central bank to ensure that each of these banks can meet their reserve targets, which might be simply zero balances at the end of a period of time, which, for simplicity, we could assume is a day.

The bank expands its balance sheet by lending. Loans create deposits, which are then backed by reserves after the fact. The process of extending loans (credit), which creates new bank liabilities is unrelated to the reserve position of the bank. The major insight is that any balance sheet expansion, which leaves a bank short of the required reserves may affect the return it can expect on the loan as a consequence of the ‘penalty’ rate the central bank might exact through the discount window. But it will never impede the bank’s capacity to effect the loan in the first place. So it is quite wrong to assume that the central bank can influence the capacity of banks to expand credit by adding more reserves into the system. We will address this proposition in more detail in Chapter 15.

Aggregate excesses or shortages of reserves across the banking system have to be rectified by the central bank. Ultimately then, the size of the stock of reserves is not discretionary in the short run. The central bank can determine the price of reserves, admittedly within some constraints, but then must provide reserves more or less on demand to hit its ‘price’ target, that is the target interbank rate (for example, the Fed funds rate in the USA and the Bank Rate in the UK), which we will describe in detail in Chapter 13.

The approach outlined in this section has been called the ‘endogenous money’ approach, in the sense that the supply of bank money is determined ‘endogenously’ by the demand for bank loans, plus the willingness of banks to lend, (which gives rise to the creation of deposits) rather than ‘exogenously’ (Moore, 1988). According to those who adopt this approach, any impact of monetary policy on the quantity of money is very indirect and operates primarily through interest rate effects.

The demand for loans, in turn, is determined by spending decisions of private economic agents (including decisions regarding asset purchases). These can be affected, but only very indirectly, by the loan rate of interest. The supply of loans is then never independent of the demand; banks supply loans only because someone is willing to ‘borrow’ bank money by issuing an IOU to banks. This means that the interest rate cannot be determined by the ‘supply and demand’ of loans since supply and demand are not independent. Rather, banks are price-setters in short-term retail loan markets. They then meet the demand for loans with some quantity rationing, at that price. Thus some requests for loans are refused, even though these aspiring borrowers claim to be willing (and able) to pay the going interest rate.

There can be several reasons for such quantity rationing of large segments of the population. Banks might worry about the default risk of borrowers, but might not be able to raise interest rates sufficiently to cover default risk, so that quantity rationing is superior to price rationing.

Often, banks probably have better information than do borrowers about such risks. For example, the borrower who wishes to open a new restaurant might not have good access to information about bankruptcy rates in the industry or might simply be overly optimistic. On the other hand, banks can never know the future, so must operate on the basis of rules of thumb (for example, informal rules that restrict loan size). Some quantity rationing can even be irrational, perhaps discriminatory, because banks have traditionally forgone certain kinds of loans or are reluctant to lend to certain groups in the community. The key point is that the supply of loans does not simply adjust to the demand for loans at some interest rate.

The short-term retail interest rates can be taken as a mark-up over short-term wholesale interest rates. Exactly what determines the mark-up (and whether it is variable) is controversial, but not important to our analysis here (see Moore, 1988). Wholesale interest rates, finally, are under the



influence of central bank policy. Individual banks use wholesale markets to rectify a mismatch between retail loans and deposits. Most banks will not be able to match exactly their retail loans and deposits. Some banks will be able to make more retail loans than they can retain in deposits and thus suffer a loss of reserves, while others will find fewer loan customers than depositors, so they will have a surplus of reserves. Banks then use wholesale markets to either ‘purchase’ reserves by issuing wholesale liabilities (for example, negotiable, large denomination certificate of deposits (CDs) or by borrowing central bank funds), while surplus banks will sell their excess reserves.

As discussed above, the central bank sets the overnight interbank rate. This rate then determines other short-term wholesale rates (mainly as a mark-up, but also as a mark-down) through arbitrage. In conclusion, the supply of money is determined endogenously while the price of money short-term interest rate is determined exogenously as a result of central bank policy.

### Summary

The orthodox position is that banks leverage (create credit) when provided with new deposits, but are constrained by fractional reserve requirements. Since the central bank is claimed to be able to control the stock of HPM or monetary base, the central bank is able to control the supply of money according to this deterministic theory of credit creation.

MMT would first deny that the central bank can control the stock of HPM, because monetary policy is conducted by the central bank setting a target interbank rate and providing the right level of reserves to the banking system so that banks lend to and borrow from each other at this target rate (for more details, see Chapters 13 and 15). Second, a bank is not constrained by its reserve position in deciding whether to make a loan to a particular customer. As long as the customer appears creditworthy and the loan is profitable to the bank, it will make the loan and then acquire sufficient reserves by borrowing from other banks or the central bank. Thus in contrast to the orthodox position of deposits driving loans, MMT argues that loans drive deposits. Further the experience of the 1980s cast considerable doubt on the relation between total reserves (HPM) and monetary aggregates, because the ‘money multiplier’ became unstable.

### A bank’s credit creation: A balance sheet analysis

The balance sheet of a typical bank looks like this:

**Figure 10.2 A typical bank balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Advances (Loans)	Checking accounts
Securities	Savings accounts
Reserves	Other liabilities
Other Assets	Net Worth

The money entries are the cheque and savings accounts on the balance sheet. Note that they are the IOUs of banks, and hence appear as liabilities on the balance sheet. The bank promises to convert deposits in a cheque account (and deposits in most savings accounts) into cash on

demand. Banks hold financial assets in the form of loans to customers and securities, that is Treasury debt and other assets.

Firms in general and banks in particular, should have positive Net Worth that is the difference between Assets and Liabilities. Total assets in the left hand column will balance with the items in the Liability column, because the latter includes Net Worth.

The following simplified series of balance sheets will clarify the process of credit creation by Bank A. Let us assume that Bank A starts with the following very simple balance sheet.

**Figure 10.3 Bank A balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Building = \$200	Net Worth = \$200

Its owners have raised capital and bought the building. The owner's equity or net worth is equal to the value of the building they have purchased. Bank A has not engaged in any banking activity yet. Now a customer comes into the bank and says that they would like to borrow \$200 to finance the purchase of a car. The bank checks their creditworthiness (asks for income tax returns, proof of assets, credit history, etc.). If the customer is approved then the following occurs on the bank's balance sheet.

**Figure 10.4 Bank A balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Loan to Customer = \$200	Cheque Account of Customer = \$200
Building = \$200	Net Worth = \$200

The bank just created \$200 of money entries (deposits in the cheque account of the customer in return for the customer's IOU, or promise to pay \$200). The bank's total assets, which equal liabilities plus Net Worth, are now \$400.

Before we move on to the customer's spending of their deposit, let us examine this balance sheet carefully.

Where did the bank get the money entry it created?

- A cheque account was created ex-nihilo, that is, from nothing, by entering a number (200) in a computer. In the past banks could also issue their own banknotes - but generally only central banks can do that now.
- The bank did not need any prior deposits, or any cash in its vault. In fact the bank did not have any cash in its vault, nor any deposits in its account at the central bank in this simplified example.
- The bank is not lending anything it has, it just creates money entries, (bank deposits), at will.
- Those money deposits or entries are its liabilities/IOUs.
- By creating those bank IOUs, the bank promises to:

- Convert deposits into cash on demand;
- Accept any of those IOUs in payment of debts owed to the bank.

The cheque account is just a legal promise to convert to cash on demand, and to accept payment in the form of the bank's own IOUs. The bank does not have to have any cash now.

The success of the banking operation (lending by accepting an IOU, and the creation of a demand deposit) depends on:

- The capacity of the customer to repay that is creditworthiness. If they have problems in making timely payments on their debts, this affects the value of the bank's assets and its own income inflows and ultimately affects the net worth of the bank, the bank's capital ratio, and the shareholders' return on equity.
- The bank's capacity to acquire reserves at low cost if:
  - The customer wants to withdraw cash;
  - The bank needs to pay debts to other banks through an interbank settlement following the customer's spending (see below);
  - The bank needs to settle tax payments made by the customer to the government.

If these conditions are not satisfied the bank gets in trouble; it can become insolvent or illiquid. The first means its net worth falls to or below zero; the second means it cannot meet cash withdrawals or clearing. Thus, even though banks can create unlimited amounts of money deposits, they have no incentive to do so because they may be unprofitable.

So what happens if now the customer pays \$200 to a car dealer who happens to have a bank account at another bank called Bank B? The balance sheet of Bank A looks like this:

**Figure 10.5 Bank A balance sheet**

<b>Change in Assets</b>	<b>Change in Liabilities</b>
	Cheque Account of the customer = -\$200
	Reserves due to Bank B = +\$200

**Figure 10.6 Bank B balance sheet**

<b>Change in Assets</b>	<b>Change in Liabilities</b>
Claim on Bank A Reserves: +\$200	Cheque Account of Car Dealer +\$200

Bank A's liabilities in the form of the customer's cheque account have dropped by \$200 through the purchase of a car, but the transaction is not confined to the reduced balances in the customer's account at Bank A and the increased balances of the Car Dealer at Bank B. Bank A now owes Bank B \$200 and needs reserves to settle this debt, but does not have reserves. Where does it get the reserves?

The banks are required to keep reserve accounts at the central bank. These reserves are liabilities of the central bank and assets of the banks. These reserves function to ensure the payments (or

settlements) system functions smoothly. That system relates to the millions of transactions that occur daily between banks as cheques are tendered by citizens and firms and more. Without a coherent system of reserves, Bank A could easily find itself unable to fund Bank B's demands based on the cheque drawn on the customer's account and presented at Bank B by the Car Dealer.

Bank A will get the reserves from the source that is the least costly. It may sell assets, but, in our example, Bank A only has a building so it would be very costly to get reserves that way. Bank A could sell bonds, if it had any, or it may borrow reserves from other banks (domestic or foreign) or the central bank. A common way to get the reserves is to borrow from the central bank, which is the monopoly supplier of reserves. Figure 10.7 documents the latest change to Bank A's balance sheet, associated with obtaining these reserves:

**Figure 10.7 Bank A balance sheet**

<b>Change in Assets</b>	<b>Change in Liabilities</b>
Reserve: +\$200	Debt to Central Bank: +\$200

**Figure 10.8 Central bank balance sheet**

<b>Change in Assets</b>	<b>Change in Liabilities</b>
Reserve Loan to Bank A: +\$200	Reserve: +\$200

Now that Bank A has the reserves it can settle its debt with Bank B.

**Figure 10.9 Bank A balance sheet**

<b>Change in Assets</b>	<b>Change in Liabilities</b>
Reserves: -\$200	Reserves due to Bank B = -\$200

**Figure 10.10 Bank B balance sheet**

<b>Change in Assets</b>	<b>Change in Assets Liabilities</b>
Claim on Bank A: -\$200	
Reserves: +\$200	

The debt between the two banks has been settled. The final balance sheets of Bank A, Bank B and the central bank look like this:

**Figure 10.11 Bank A balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Funds Advance to Customer= \$200	Debt to Central Bank = \$200
Building = 200	Net Worth = 200

Bank A makes money as long as the interest it receives on the loan to the customer is higher than the interest it pays to the central bank on the reserves. The balance sheet of Bank B is shown below, where we assume Bank B had reserves prior to the checking account of the Car Dealer being increased by the sale of the car to the customer.

**Figure 10.12 Bank B balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Reserves + \$200	Checking Account of Car Dealer + \$200

Finally the balance sheet of the central bank is shown in Figure 10.13.

**Figure 10.13 Central bank balance sheet**

<b>Assets</b>	<b>Liabilities</b>
Reserve Loan to Bank A= \$200	Reserves: +\$200

Note that none of these operations involved any transfer of physical cash. It was all bookkeeping entries through keystrokes to computers. Also note we only show the assets and liabilities directly related to our examples. Of course, private banks and the central bank have many other assets and liabilities, as well as net worth on their balance sheets.

In practice, the central bank will usually not advance reserves to the bank directly in the form of an unsecured advance; instead it will ask for collateral (usually a treasury security) in exchange and will provide funds for less than the value of the collateral. So, if Bank A has a \$300 bond, it surrenders it to the central bank in exchange for reserves. The Fed will give only say \$285 if the discount rate is 5 per cent.

## **Appendix: Advanced Material**

### **Theories of the yield curve and its dynamics**

Mitchell (2011b) notes that there are various theories about the yield curve and its dynamics. All share some common notions – in particular that the higher is expected inflation the steeper the yield curve will be other things equal.

The basic principle linking the shape of the yield curve to the economy's prospects is explained as follows. The short end of the yield curve reflects the interest rate set by the central bank. The steepness of the yield curve then depends on the yield of the longer-term bonds, which are set by the market. But the short end of the curve is the primary determinant of its slope. In other words, the curve steepens mainly because the central bank is lowering the official cash rate, and it flattens mainly because the central bank is raising the official cash rate.

Bond traders link the dynamics of the yield curve to their expectations of the future economic prospects. When the yield curve flattens it is usually accompanied by deflation or steady and low inflation and vice versa.

One of the risks in holding a fixed coupon bond with a fixed redemption value is purchasing power risk. Economists believe that most people would prefer to consume now rather than later if there was to be a trade-off. To encourage foregone consumption now, a yield on savings must be provided by markets. The yield is intended to allow a person to consume more in the future than has been sacrificed now. But if the prices of real goods and services increase in the meantime, then inflation could completely wipe out any gain in real consumption, so that the real interest rate is zero.

Consider a person who invests in a one-year \$1,000 coupon treasury bond with a single coupon payment expected of \$100. The individual will expect to get \$1,100 on the redemption date. Assume that over the holding period, prices rise by 10 per cent. At the end of the year, a basket of goods that previously cost \$1,000 would now cost \$1,100. In other words, the investor is no better off at the end of the year as a result of the investment. The nominal yield has been offset by the price inflation.

Purchasing power risk is more threatening the longer is the maturity. This is one reason why longer maturity rates will be higher. The market yield is equal to the real rate of return required plus compensation for the expected rate of inflation. If the inflation rate is expected to rise, then market rates will rise to compensate. In this case, we would expect the yield curve to steepen, given that this effect will impact more significantly on longer maturity bonds than at the short end of the yield curve.

## References

ABS (2012) Year Book Australia, 1301.0, available at: <http://www.abs.gov.au/ausstats/abs@.nsf/mf/1301.0>

Brunner, K. (1968) "The Role of Money and Monetary Policy", Federal Reserve Bank of St Louis Review, 50, pp.8-24.

Mitchell, W.F. (2011a) "The role of bank deposits in Modern Monetary Theory", blog posted on Thursday, May 26, 2011, available at: <http://bilbo.economicoutlook.net/blog/?p=14620>.

Mitchell, W.F. (2011b) "When might that be?" blog posted on Thursday, July 21, 2011, available at: <http://bilbo.economicoutlook.net/blog/?p=15348>.

Moore, B. (1988) Horizontalists and Verticalists: The Macroeconomics of Credit Money, Cambridge University Press, Cambridge.

# Chapter 11: Unemployment and Inflation

## Chapter Outline

- 11.1 Introduction
  - 11.2 What is Inflation?
  - 11.3 The Quantity Theory of Money
  - 11.4 The Phillips Curve
    - The instability of the Phillips curve
  - 11.5 The Accelerationist Hypothesis and the Expectations-Augmented Phillips Curve
    - Introduction
    - Expectations of inflation
  - 11.6 Hysteresis and the Phillips Curve Trade-Off
  - 11.7 Underemployment and the Phillips Curve
  - 11.8 Demand-Pull and Cost-Push Inflation
  - 11.9 The Conflict Theory of Inflation
  - 11.10 Raw Material Price Rises
  - 11.11 Cost-Push and Demand-Pull Inflation Summary
  - 11.12 Incomes Policies
- Appendix: Advanced Material
- The Phillips curve algebra
  - Econometric mis-specification
  - The expectations-augmented Phillips curve
  - Inflationary expectations
  - Hysteresis and the Phillips Curve
  - Conflict theory and inflationary biases
  - The Scandinavian Model (SM) of inflation

## **Learning Objectives**

1. Understand how inflation is defined and the sufficient condition for its persistence.
2. Identify and differentiate between competing theories of inflation.
3. Recognise that the EAPC has had a profound influence on the conduct of macroeconomic policy in developed economies for over 40 years.



## 11.1 Introduction

In this Chapter, we will review the concept of inflation and discuss various approaches that seek to explain it. We will differentiate between **demand-pull** and **cost-push** inflation.

The first type of inflation has been termed demand-pull because excess nominal demand (relative to real output capacity) initially pushes up the price level. The second type of inflation is termed cost-push inflation because it originates from the costs of production increasing and pushing up the price level.

Both forms of inflation can be understood within a general framework whereby different claimants on real GDP and national income struggle to assert their aspirations. In this sense, we cast inflation within the general distributional struggle or conflict that is characteristic of capitalist economies, between workers seeking to maintain or achieve a higher real wage and firms seeking to maintain or raise their profit rate.

In the pre-Keynesian era the concept of full employment only allowed for voluntary unemployment that is, workers choosing to be unemployed. However, in the immediate post World War II Keynesian era, the concept of full employment was recast and the emphasis became one of providing enough jobs to match the work preferences of the available labour force. The notion of involuntary unemployment was at the heart of this conception of full employment. That is, full employment coincided with zero involuntary unemployment.

This post World War II consensus was steadily eroded away over the next 40 odd years. By the early to mid-1970s, mainstream macroeconomics reverted back to the pre-Keynesian notions of voluntary unemployment and effectively abandoned the concept of true full employment.

However, the process of abandoning true full employment began in the 1950s when the discussion turned to inflation and the trade-off between the twin evils of unemployment and inflation. This was the era in which the **Phillips curve** literature emerged, based on what was then considered to be a statistically reliable, inverse relationship between unemployment and inflation – the so-called **trade-off between unemployment and inflation**.

However, later Monetarist and New Classical reinterpretations of the trade-off appeared to be conclusive. Classical (pre-Keynesian) notions of a natural unemployment rate (understood to be equivalent to full employment) were revived which led to the rejection of demand management policies, which were aimed to limit unemployment to its frictional component.

In this Chapter, we will carefully analyse the Phillips curve and how the idea that there might be a trade-off between the twin-evils of unemployment and inflation has changed with the augmentation of the Phillips curve with inflationary expectations and the so-called **natural rate of unemployment**. Since the early 1970s, ideological dominance in this debate has been assumed by those who eschew the intervention of government and consider that the unfettered operation of the market will generate full employment. The persistence of mass unemployment around the world is testament to the error of their thinking.

Finally, drawing on empirical evidence, we develop a model of inflation, which exploits the concept of hysteresis to justify the restoration of a trade-off between the inflation rate with a broader measure of labour underutilisation. The model supports the MMT and post-Keynesian view more generally, that there is an important role for government to play in maintaining low levels of unemployment.

## 11.2 What is Inflation?

There are misconceptions as to what inflation actually is. An increase in wages or prices is a necessary but not sufficient condition for an inflationary process to unfold. Thus a negotiated pay increase for workers or companies increasing their prices to try to increase profit or a rise in local prices of imported goods following a depreciation of the exchange rate may or may not initiate an inflationary process.

**Inflation is the continuous rise in the price level, so, the price level has to be rising for a number of time periods. A once-off price rise is not an inflationary episode.**

If the price level rises by 10 per cent every month, for example, then we would be observing an inflationary episode. In this case, the inflation rate would be considered stable with the price level rising at a constant rate per period.

If the price level was rising by 10 per cent in month one, then 11 per cent in month two, then 12 per cent in month three and so on, then we would be observing an accelerating inflation rate. Extreme cases of **accelerating** inflation are referred to as hyperinflation. There have been few instances of this problem in recorded history, but the Weimar Republic in the 1920s and Zimbabwe at the beginning of the 21st century are notable examples. They were both marked by a dramatic contraction of the supply potential of the respective economies prior to the hyperinflation.

Alternatively, if the price level was rising by 10 per cent in month one, 9 per cent in month two and so on, then the rate of inflation is falling or **decelerating**. If the price level starts to fall, then the growth of the price level is negative and this would be a **deflationary** episode. Students may wish to refresh their understanding of the measurement of the Consumer Price Index and the computation of the inflation rate in Chapter 4, Section 4.7.

We can define a normal price level as being the prices firms are willing to charge when they are operating at normal capacity and earning a profit rate that satisfies their strategic aspirations. Please refer to the discussion of mark-up pricing in Chapter 8.

However, the economic (business) cycle fluctuates around these normal rates of capacity utilisation and firms not only adjust to the flux and uncertainty of aggregate demand by adjusting output, but in some cases, will vary prices. This is particularly the case during a recession.

When there are very depressed levels of activity, firms might offer discounts in order to increase sales and hence capacity utilisation. They thus temporarily suppress their profit margins in order to try to raise their respective market shares, when overall demand is falling. As demand conditions become more favourable, firms start withdrawing the discounts and prices return to those levels that offer the desired rate of return to firms at normal rates of capacity utilisation.

We do not consider these cyclical adjustments in prices to constitute inflation.

### 11.3 The Quantity Theory of Money

The Classical theory of employment is based on the view that the real variables in the economy – output, productivity, real wages, and employment – were determined by the equilibrium outcome in the labour market.

The Classicists assumed that the real wage is determined in the labour market, that is, exclusively by labour demand and labour supply. Labour demand was inversely related to the real wage because they asserted that marginal productivity was subject to diminishing returns and the supply of labour was positively related to the real wage because workers would prefer to work more hours as the price of leisure (the real wage) rose.

The real wage is construed in this theory as being the price of leisure in the sense that it represents the goods and services foregone (via lost income) of an hour of non-work (leisure). The real wage is thus a relative price – of leisure relative to other goods and services.

The important Classical result is that the interaction between the labour demand and supply functions determines the real level of economic activity at any point in time. The aggregate supply of goods and services is determined by the level of employment and the prevailing technology, which maps how much output is forthcoming for a given level of employment. The more productive is labour the higher will the output supply be at each level of employment.

**Say's Law**, which follows from the loanable funds doctrine (see Chapter 14), is then invoked to assume away any problems in matching aggregate demand with this supply of goods and services. The loanable funds doctrine posits that saving and investment will always be brought into balance by movements in the interest rate, which is construed as being the price of today's consumption relative to future consumption.

The theory thus assumes that two relative prices – the real wage in the labour market and the interest rate in the loans market – ensure that full employment occurs (with zero involuntary unemployment). Knowledge of the general price level was thus irrelevant to explaining the real side of the economy.

This separation between the explanation for the determination of the real economic outcomes and the theory of the general price level is referred to as the **classical dichotomy**, for obvious reasons. The later Classical economists believed that if the supply of money was, for example, doubled, that there would be no impact on the real performance of the economy. All that would happen is that the price level would double.

The classical dichotomy that emerged in the C19th stands in contradistinction to the earlier ideas developed by economists such as David Hume that there was a trade-off between unemployment and inflation that could be manipulated (in policy terms) by the central bank varying the money supply (Hume, 1752).

It is of no surprise that the Classical employment model relied, in part, on the notion of a classical dichotomy for its conclusions. Its origins were based on a barter model where there is an absence of money and owner producers trade real products. Clearly, this conception of an economy has no application to the monetary economy we live in.

The development of Classical monetary theory was only intended to explain the level and change in the general price level. The main attention of the Classical economists was in trying to

understand the supply of output and the accumulation of productive capital (and hence economic growth).

The theory of the general price level that emerged from the classical dichotomy was called the **Quantity Theory of Money**. The theory had its origins in the work of French economists in the sixteenth century, in particular, Jean Bodin.

Why would we be interested in something a French economist conceived in the sixteenth century? The answer is that in the same way that the main ideas of Classical employment theory still resonate in the public debate (for example, the denial that mass unemployment is the result of a deficiency of aggregate demand), the theory of inflation that arises from the Quantity Theory of Money is still influential and forms the core of what became known as **Monetarism** in the 1970s.

As we have learned already in reading this textbook, economics is a contested discipline and different schools of thought advance conflicting policy frameworks. Monetarism and its more modern expressions form one such school of thought in macroeconomics and rely on the Quantity Theory of Money for its inflation theory.

We will also see that the crude theory of inflation that emerges from the Quantity Theory of Money has intuitive appeal and is not very different to what we might expect the average lay person would believe, namely that growth in the money supply causes its value of money to decline (that is, causes inflation).

The Quantity Theory of Money was very influential in the nineteenth century. The theory begins with what was known as the **equation of exchange**, which is an accounting identity.

We write the equation as:

$$(11.1) \quad M_s V \equiv PY$$

$PY$  is the nominal value of total output (which is simply the definition of nominal GDP in the national accounts) given  $P$  is the price level and  $Y$  is real output. Consistent with that definition you will understand that  $PQ$  is a flow of output and expenditure.

$M_s$  is the quantity of money in circulation (the money supply, say M2 which was defined in Chapter 10) and  $V$  is called the velocity of circulation, which is the average circulation of the money stock.  $V$  is thus the turnover of the quantity of money per period in making transactions.

To understand velocity, we can consider the following example. Assume the total stock of money is \$100, which is held by the two people that make up this economy. In the current period (say a year), Person A buys goods and services from Person B for \$100. In turn, Person B buys goods and services from Person A for \$100.

The total transactions equal \$200 yet there is only \$100 in the economy. Thus each dollar must be used twice over the course of the year. So the velocity in this economy is 2.

The money supply is a stock (so many dollars at a point in time). Any given stock of money might turnover several times in any given period in the course of all the myriad of transactions that are made using money. As we learned in Chapter 5 when we considered stocks and flows, flows add to or subtract from related stocks.

The velocity of circulation converts the stock of money into a flow of money and renders the left-hand side of Equation (11.1) commensurate with the right-hand side.

As it stands, Equation (11.1) is a self-evident truth because it is an accounting statement. It is obvious that the total value of spending ( $M_sV$ ) will have to equal to the total nominal value of output ( $PY$ ). In other words, there is no theoretical content in the relationship as it stands. In other words, the equation is always satisfied due to the way we have defined the variables.

We thus need to introduce some behavioural elements into Equation (11.1) in order to use it as a theory of the general price level.

In this regard, it is important to see the Quantity Theory of Money and Say's Law as being mutually reinforcing planks of the Classical theory. The latter was proposed to justify the presumption that full employment output would be continuously supplied and sold, which meant that the former would ensure that changes in the stock of money would only impact on the price level.

As Keynes observed, price level changes did not necessarily correlate with changes in the money supply, which led to his rejection of the Quantity Theory of Money.

In turn, his understanding of how the price level could change without a change in the money supply was informed by his rejection of Say's Law – that is, his recognition that total employment was determined by effective demand and the capitalist monetary economy could easily fall into a state where effective demand was deficient.

But the Classical theorists considered that a flexible real wage would ensure that full employment was attained – at least as a normal state where competition prevailed and there were no artificial real wage rigidities imposed.

As a result, they considered  $Y$  to be fixed at the **full employment output level**.

Additionally, they considered  $V$  to be constant given that it was determined by customs and payment habits. For example, people are paid on a weekly or fortnightly basis and shop, say, once a week for their needs.

Equation (11.2) depicts the resulting causality that defines the Quantity Theory of Money as an explanation of the general price level. The horizontal bars above the  $V$  and  $Y$  indicate they are assumed to be constant. It follows that changes in  $M$  will directly and only impact on  $P$ .

$$(11.2) \quad M\bar{V} \equiv P\bar{Y}$$

$$\therefore M \rightarrow P$$

To understand this theory more deeply it is important to note that the Classical economists considered the role of money to be confined to acting as a medium of exchange to free people from the tyranny of a double coincidence of wants in barter. That is, to overcome the problem that a farmer who had carrots to offer but wanted some plumbing done could not find a plumber desiring any carrots.

Money was seen as lubricating the process of real exchange of goods and services and there was no other reason why a person would wish to hold it.

The underlying view was that if individuals found they had more money than in the past then they would try to spend it. Logically, it followed that they considered a rising stock of money to be associated with higher growth in aggregate demand (spending).

As Equation (11.2) shows, monetary growth (and the assumed extra spending) would directly lead to price rises because the economy was already assumed to be producing at its maximum productive capacity and the habits underpinning velocity were stable.

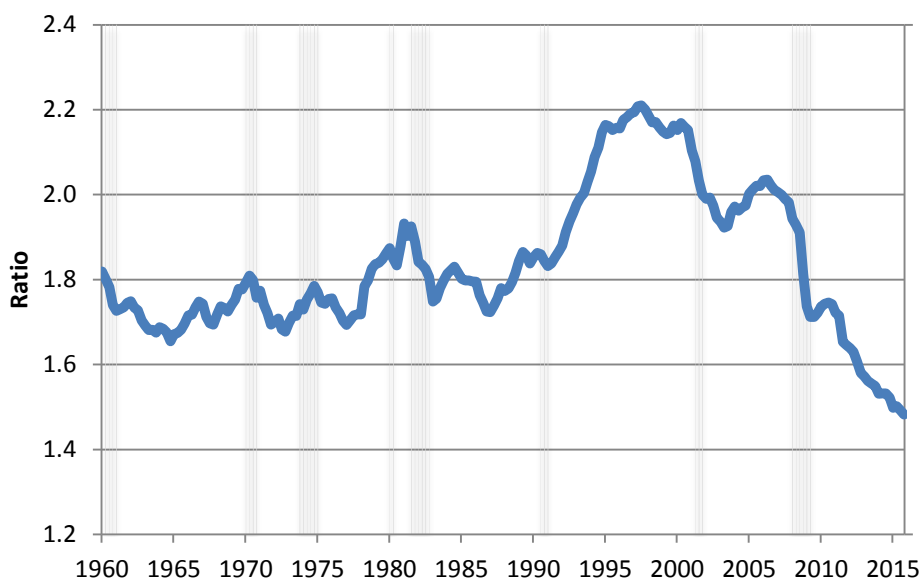
For now students should note two empirical facts. First, capitalist economies are rarely at full employment. Since economies typically operate with spare productive capacity and often with high rates of unemployment, it is hard to maintain the view that there is no scope for firms to expand real output when there is an increase in nominal aggregate demand.

Thus, if there was an increase in availability of credit and borrowers used the deposits that were created by the loans to purchase goods and services, firms with excess capacity are likely to respond by raising real output to maintain market share rather than raising prices.

Second, the empirical behaviour of the velocity of circulation demonstrates that the assumption that it is constant is implausible. Figure 11.1 uses data provided by the US Federal Reserve Bank of St. Louis and shows the velocity of circulation, constructed as the ratio of nominal GDP to the M2 measure of the money supply.

The US Federal Reserve Bank of St. Louis defines this measure: "...as the rate of turnover in the money supply—that is, the number of times one dollar is used to purchase final goods and services included in GDP." (2016)

**Figure 11.1 Velocity of M2 money stock (M2V), US, 1950-2012**



Source: US Federal Reserve Bank of St. Louis. Shaded areas indicate US recessions.

The evidence does not support the claims of the Quantity Theory of Money. There is not a simple proportionate relationship between increases in the money supply and rises in the general price level.

## 11.4 The Phillips Curve

In Chapter 8 we derived what we defined as the General Aggregate Supply Function (Figure 8.5), which was reverse L shaped. The horizontal segment was explained by the price mark-up rule and the assumption of constant unit costs over a range of output. In other words, firms in aggregate are assumed to supply as much real output (goods and services) as is demanded at the current price level, up to a limit defined by the available capacity.

The Aggregate Supply Function becomes vertical at full employment because beyond that point the economy exhausts its capacity to expand short-run output due to shortages of labour and capital equipment. At that point, firms will be trying to outbid each other for the already fully employed labour resources and in doing so would drive money wages up. Under normal circumstances, the economy will rarely approach the output level ( $Y^*$ ), which means that the economy usually faces constant costs.

We acknowledged in Chapter 8, however, that rising costs **might** be encountered given that all firms are unlikely to hit full capacity simultaneously, so that cost pressures would begin to mount before the overall full capacity output is reached. Thus the reverse L-shaped Aggregate Supply curve represented an analytical simplification.

In the pre-Keynesian era, unemployment was considered to be a voluntary state and full employment was thus defined in terms of the employment level determined by the intersection of labour demand and labour supply. So by construction, full employment reflected the optimal outcome of maximising, rational and voluntary decision making by workers and firms. At the so-called full employment real wage, every worker wanting work could find an employer willing to offer the desired hours of employment and every employer could secure the services of willing employees.

In 1958, New Zealand economist **Bill Phillips** published a statistical study, which showed the relationship between the unemployment rate and the rate of change (growth rate) of money-wage rates for the United Kingdom, for the period 1861 to 1957.

Phillips believed that since money wage costs represented a high proportion of total costs, movements in money wage rates would drive movements in the general price level. Later, economists constructed the relationship as being between the rate of price inflation and the unemployment rate.

In the Phillips curve framework, the level of economic activity is represented by the unemployment rate, so it linked the level of economic activity to changes in the price level. Therefore, when the unemployment rate rises above some irreducible minimum, economic activity is declining and as the unemployment rate moves towards that irreducible minimum, the economy moves closer to full capacity and full employment. Okun's Law, which documents the relationship between changes in the unemployment rate and output gaps (the difference between potential output and actual output), provides the extra justification for the link between unemployment and output. Figure 11.2 shows a stylised Phillips curve.

Phillips (1958, p.283) explained this empirical relationship by reference to the supply and demand for labour:

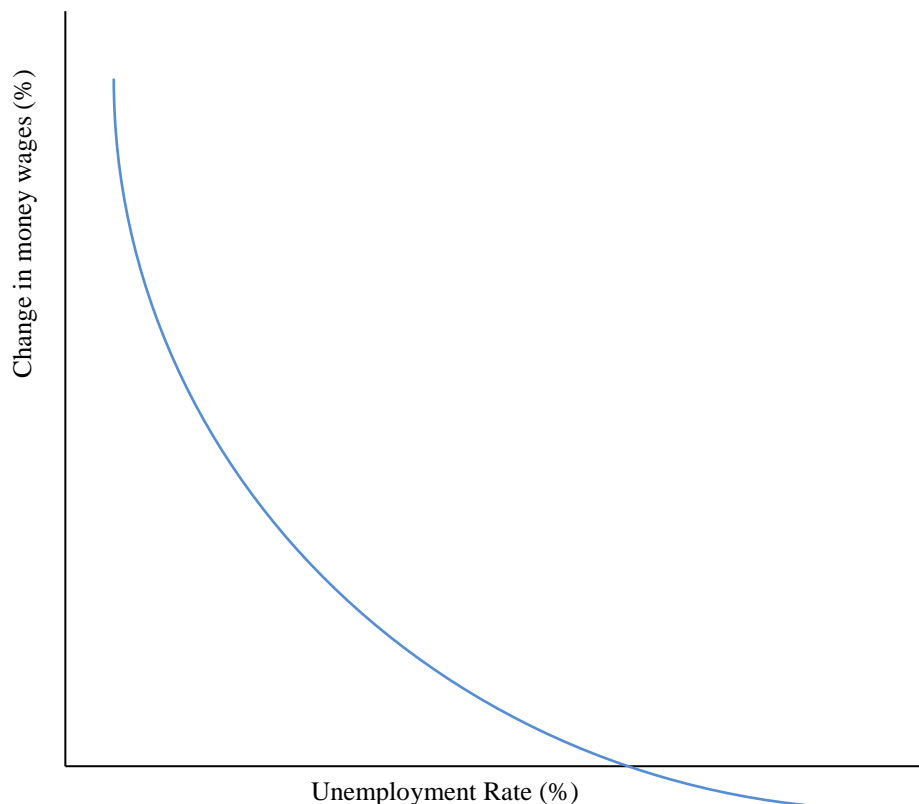
When the demand for labour is high and there are very few unemployed we should expect employers to bid wage rates up quite rapidly ... On the other hand it appears that workers

are reluctant to offer their services at less than the prevailing rates when the demand for labour is low and unemployment is high so that wage rates fall only very slowly. The relation between unemployment and the rate of change of wage rates is therefore likely to be highly non-linear.

He also recognised that in reference to wage bargaining, the direction of change in the economy was a factor that had to be considered, quite apart from the level of economic activity. Thus, at a particular rate of unemployment, employers would bid more vigorously for the services of labour, when demand was increasing as opposed to falling.

Note that Figure 11.2 assumes that at very high levels of unemployment, the growth in the money wage is negative. In normal economic conditions there tends to be downward money wage rigidity, and so the Phillips Curve would always be associated with positive or (close to) zero money wage growth. However during the GFC a number of Eurozone economies experienced falling wages.

**Figure 11.2 The basic Phillips curve**



In addition, he claimed that a rising cost of living would also contribute to the growth in money wage rates, but would be of less importance unless there was a “very rapid rise in import prices” (p.284).

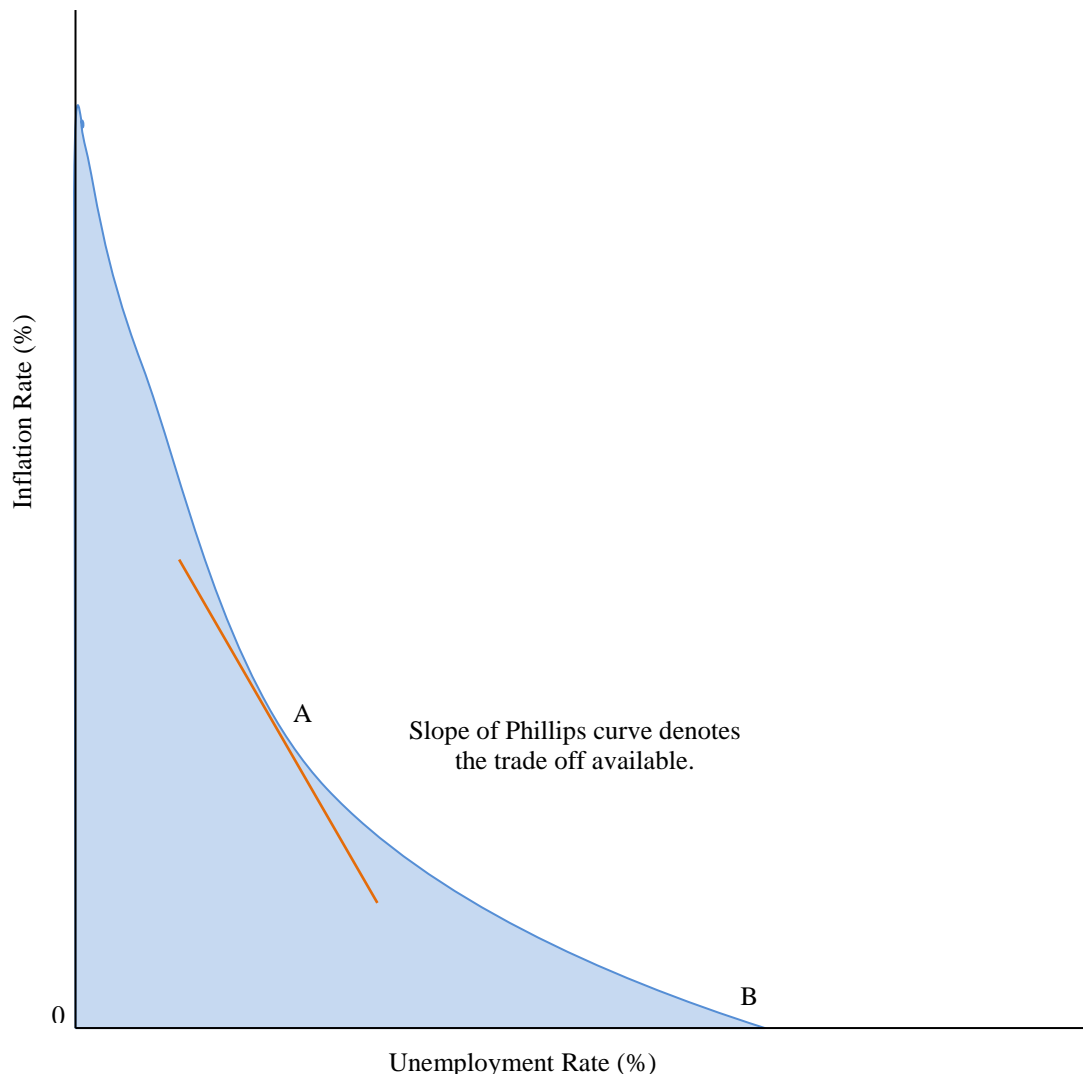
By acknowledging that the movement in retail prices is intrinsic to understanding the movement in real wages, Phillips reinforced the arguments made by Keynes that the real wage was not



determined in the labour market and thus workers could not directly manipulate the prevailing real wage.

In later work, Samuelson and Solow (1960) defined the Phillips Curve as a policy tool, which the government could use to lessen the burden of unemployment. Reference was made to a **menu of choices**, which implied that policy makers could choose a preferred combination of inflation and unemployment, that is a particular point on their estimated Phillips curve. This introduced the idea of a **policy trade-off** between unemployment and inflation. If the government wanted to sustain lower unemployment rates then the cost of that policy decision would be higher inflation.

**Figure 11.3 The unemployment-inflation choice set**



What would be the best policy choice for government? This was a political decision and governments tried to assess what the socially acceptable combination of the twin evils – inflation and unemployment – might be. Depending on the ideological preferences of the voters, some nations might choose Point B on Figure 11.3 while other nations would prefer Point A with higher inflation accompanied by lower unemployment.

There would be a political problem for government if the socially acceptable combinations of inflation and unemployment were unattainable because they lay within the shaded area, rather than on the Phillips Curve trade-off.

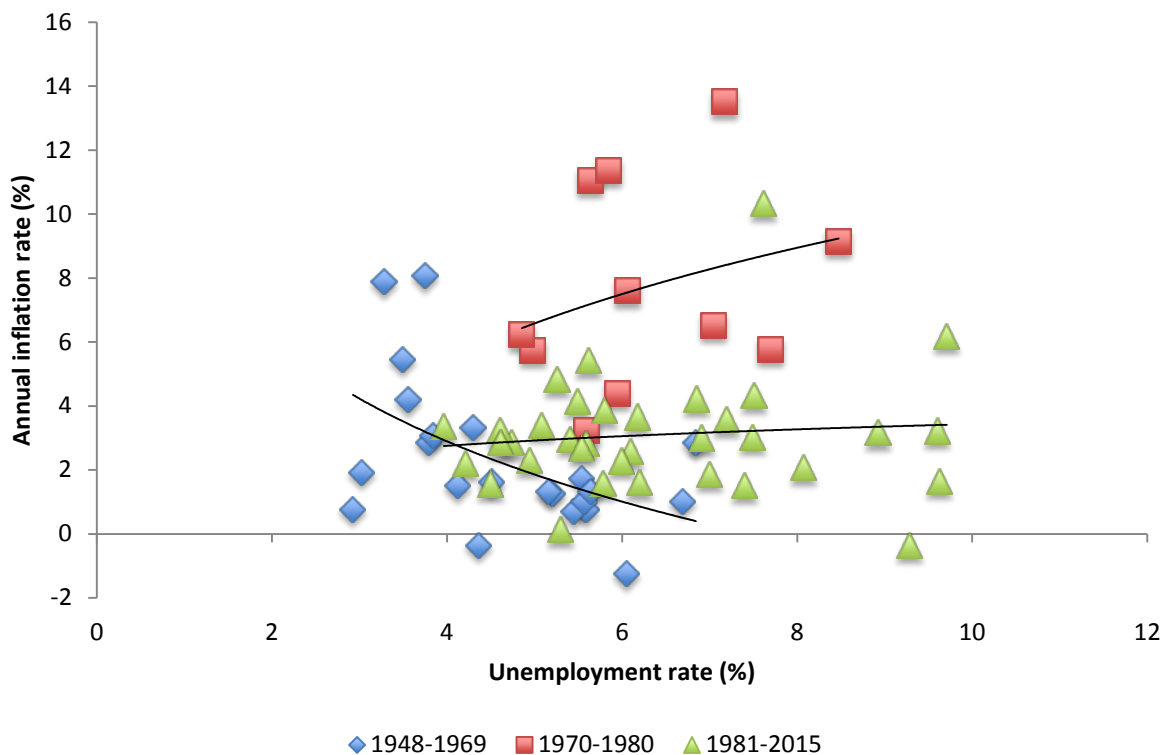
### The instability of the Phillips curve

Consider Figure 11.4, which shows the combinations of the unemployment rate and the annual price inflation rate for the US from 1948 to 2012, subdivided into three periods: 1948-1969; 1970-1980; and 1981-2012. The black lines depict the logarithmic regression between the inflation rate and the unemployment rate for the three sub-periods.

The blue diamonds show observations for the period 1948 to 1969. The curve, which fits the data quite well, shows a simple Phillips curve trade-off of the type depicted in Figure 11.2.

However, consider the observations for 1970 to 1980, which are shown by the red squares. Those observations are clearly inconsistent with a stable Phillips curve and seem to imply a positively sloping relationship. This apparent **shift** in the Phillips curve was considered to be a collapse in the relationship and led to accusations that the underlying conceptualisation of the Keynesian Phillips curve was flawed.

**Figure 11.4 The shifting US Phillips curve - 1948-2012**



Source: US Bureau of Labor Statistics. The inflation rate is the annual rise in the CPI.

By the 1980s, when inflation moderated, it became hard to determine any relationship between inflation and unemployment in the US economy.

From an empirical perspective then, the belief that the Phillips curve was a stable relationship, which could be exploited in a predictable manner by policy makers according to their preferences between inflation and unemployment, became highly questionable.

We always have to be very careful when we visualise data in this way. First the choice of dates for the different periods is important. Second, the observations between 1970 and 1980 may in fact signify a shifting Phillips curve relationship and the regression line is just picking up the shifting function. However for the trade-off to be empirically sustained it is necessary for consecutive annual observations to define a trade-off, via econometric estimation and a rationale needs to be provided as to whether the trade-off is shifting.

## 11.5 The Accelerationist Hypothesis and the Expectations-Augmented Phillips Curve

### Introduction

The legacy of the earlier Keynes and Classics debate persisted through the 1950s and 1960s. The neo-classical school was unwilling to accept the basic insights provided by Keynes that effective demand drove output and national income and that the capitalist monetary system was susceptible to crises of over-production and unemployment.

In the late 1960s, this on-going debate about the effectiveness of fiscal and monetary policy in stabilising the economic cycle was rehearsed within the Phillips curve framework.

A group of economists, centred at the University of Chicago were opposed to government attempts to maintain full employment. Their argument largely reflected their belief that a self-regulating free market would generate optimal outcomes. Thus, they were adherents of the neo-classical model that considered most government intervention to be problematic (see the discussion of the orthodox, neoclassical approach in Chapter 1).

This debate at the microeconomic level was manifested in demands for widespread deregulation in product, labour and financial markets and a major reduction in the size of government.

At the macroeconomic level, the Phillips curve was the 'battleground'. During the 'stable' Phillips curve policy era, policy makers assumed they could target a low unemployment rate and incur a modest inflation rate as a consequence. The extent to which inflation rose was determined by the slope of the Phillips curve, which was considered to be relatively flat.

The emerging Monetarists, who eschewed government intervention, challenged that view and asserted that there was no permanent (long-run) trade-off between inflation and unemployment. They claimed that, ultimately the market would ensure the unemployment rate was stable around its so-called **natural rate** and attempts by government to push the unemployment below its natural rate would lead to accelerating inflation.

Chicago economist Milton Friedman was the most vocal Monetarist and in a famous article in 1968 outlined what became known as the **accelerationist hypothesis**.

### Expectations of inflation

The empirical instability in the relationship between unemployment and inflation opened the way for the so-called Monetarist paradigm in macroeconomics to gain ascendancy.

The Monetarists reinterpreted the inflation, unemployment trade-off by adding the role of inflationary expectations, and in doing so, revived the Classical (pre-Keynesian) notion of a natural unemployment rate (defined as equivalent to full employment). The devastating consequence of this assertion was the rejection of a role for demand management policies to limit unemployment to its frictional component.

The Keynesians had adopted the Phillips curve, a macroeconomic relationship, yet they had developed very little microeconomic theory to underpin it. They had justified the Phillips curve as a competitive adjustment process, such that if there was growing demand for labour, money wages rose as unemployment fell.

But Monetarists claimed that workers cared about real wages rather than nominal wages, since the real wage represented a worker's capacity to buy goods and services. Thus the basic Phillips curve was defective, because it only focused on the relationship between percentage change in money wages and the unemployment rate and ignored the influence of a changing price level, in terms of expectations about inflation.

The **accelerationist hypothesis** was advanced in 1968 by Milton Friedman before the empirical breakdown of the relationship between inflation and unemployment emerged in the early 1970s.

So while the Phillips curve presented the Monetarists with the opportunity to debate the failings of the mainstream Keynesian analysis, it was the subsequent empirical havoc created by the 1970s oil price shocks, which added weight to their theoretical arguments, even though they had been shown to be deficient.

The mis-specification of the Phillips curve, which had ignored inflationary expectations, was not significant while inflation was negligible. Once inflation rates soared throughout the world with the oil price rises of the early 1970s, all the Phillips curve relationships broke down. As a result the Monetarist concept of a natural rate of unemployment appeared to be validated, along with the rejection of aggregate demand management through fiscal policy.

There were two basic propositions that Friedman asserted in his attack on the Phillips curve.

First, he claimed that there is a natural rate of unemployment, which is determined by the underlying structure of the labour market and the rate of capital formation and productivity growth. He believed that the economy always tends back to that level of unemployment even if the government attempts to use fiscal and monetary policy expansion to reduce unemployment.

He later noted that the **natural rate of unemployment** is not "immutable and unchangeable" but is insensitive to monetary (aggregate demand) forces (Friedman, 1968: 9). That is, he considered increasing nominal aggregate demand would not reduce the natural rate.

He also considered that the natural rate was, in part, "man-made and policy-made" (p.9). For example, Monetarists argued that imposing minimum wages and providing unemployment benefits would increase the natural rate.

The concept of the natural rate of unemployment that Friedman developed follows straightforwardly from the Classical labour market. We briefly summarise the arguments here. The Monetarist approach was an attempt to promote ideas that Keynes demolished in the 1930s. Friedman asserted that real wages were the relevant object of concern from the perspective of firms and workers rather than money wages. Workers supplied labour based on the opportunity cost of leisure, which is the income given up by say an extra hour of leisure, which is the real

wage. On the other hand, firms employed labour to maximise profit, so the demand for labour is also a function of the real wage.

It was argued that real wages would adjust to ensure that the labour market cleared at the 'natural rate' level of unemployment (i.e. the demand for labour equalled the supply of labour). While there might be temporary deviations around that rate, for reasons we will explore next, over time, the economy would always be tending back to the natural rate of unemployment.

The natural rate was thus conceived as being the level of unemployment that arose as a result of natural frictions in the labour market and had no cyclical component. In other words, it did not arise as a result of a deficiency in aggregate demand. Friedman considered that these frictions could include the distortions arising from policy decisions, which were noted above.

Second, the Phillips curve is, at best, a short-run relationship that can only be exploited through aggregate demand expansion as long as workers suffer from money illusion by confusing money wage rises with real wage rises. In other words, any given short-run Phillips curve is dependent on workers assuming that the prevailing rate of price inflation is stable.

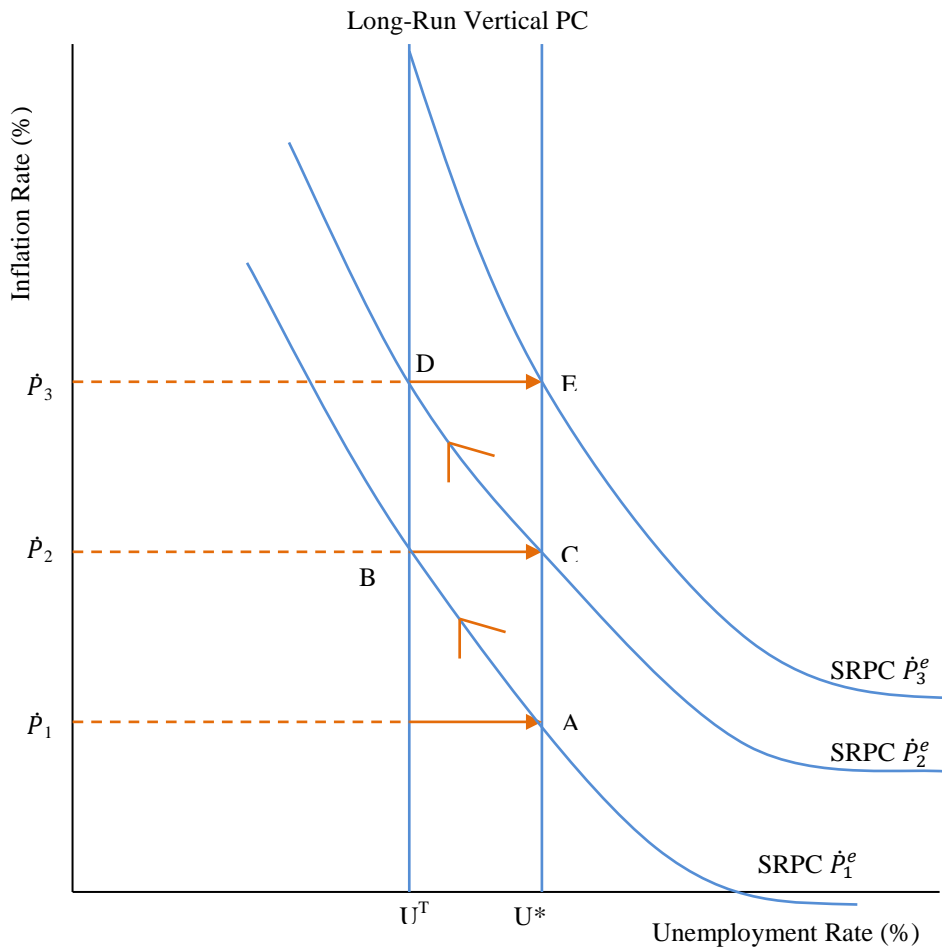
However, Friedman and others argued that eventually workers would realise that their real wage was being eroded as price inflation outstripped money wages growth in the light of the expansion of aggregate demand. In doing so, they would start to form expectations of higher inflation.

As a consequence, workers would build these inflationary expectations into their future outlook and pursue money wage increases, which reflected not only the state of the labour market (relative strength of demand and supply) but also how much they expected prices to rise in the period governed by the money wage bargain.

The Monetarists argued that if the government attempted to reduce unemployment below the natural rate, then as the inflation rate rose, workers would demand even higher money wages growth to achieve their desired real wage levels. Ultimately, this would result was a rising rate of inflation

Figure 11.5 captures the accelerationist hypothesis. The short-run Phillips curves are shown conditional on a specific expectation of inflation held by the workers. The superscript  $e$  denotes **expected inflation**. We use the terminology, 'expectations are realised' to denote a state where the expectations are equal to the actual inflation outcome.

**Figure 11.5 The expectations-augmented long-run Phillips curve**



We start at Point A, where the inflation rate is  $\dot{P}_1$  and the unemployment rate is at its so-called ‘natural rate’ ( $U^*$ ). At this point, the expectations about the rate of inflation held by workers ( $\dot{P}_1^e$ ) are consistent with the actual inflation rate,  $\dot{P}_1$ . According to Friedman, the labour market would be operating at the natural rate of unemployment, whenever inflationary expectations are realised.

To see how the accelerationist hypothesis plays out, we assume that the government is under political pressure and forms the view that the unemployment rate,  $U^*$ , is too high. It believes it can use expansionary fiscal and monetary policy to target a lower rate of unemployment,  $U^T$ . They also think they can exploit the Phillips Curve trade-off and move the economy to Point B, with a higher inflation rate ( $\dot{P}_2$ ) as the cost of the lower unemployment rate.

Consequently, the government stimulates nominal aggregate demand to push the economy to point B. The increased demand for labour pushes up the inflation rate (to  $\dot{P}_2$ ) and money wage rates also rise in the labour market. The accelerationist hypothesis assumes that the price level accelerates more quickly than money wages and as a consequence the real wage falls.

The Monetarists resurrected the Classical labour market and placed it at the centre of their attack on Keynesian macroeconomics. Accordingly, firms will offer more employment because the real wage has now fallen and the demand for labour is an inverse function of the real wage.

Why would workers supply more labour if the real wage was falling? In the Classical labour market it is assumed that labour supply is a positive function of the real wage so workers will withdraw labour if the real wage falls.

The Monetarist approach overcomes that apparent problem by imposing asymmetric (different) expectations on the workers and firms. Firms are assumed to have complete price and wage information at all times so they know the level of the actual real wage at any point in time.

However, the workers were assumed to gather information about the inflation rate in a lagged or adaptive fashion and thus could be fooled into believing that the real wage was rising when, in fact, it was falling.

Thus, workers are assumed to be initially oblivious to the higher inflation – that is, their inflationary expectations do not adjust to the actual inflation rate immediately. As a consequence they mistake the rising nominal wages for an increasing real wage and willingly supply more labour even though the real wage has actually fallen.

The central proposition of the Classical labour market is that workers care about real wages not money wages. The accelerationist hypothesis added the idea that workers form **adaptive expectations of inflation**, which means that it takes some time for them to differentiate between movements in money wages and movements in real wages.

Monetarists asserted that Point B is unstable and can only persist as long as workers are fooled into believing the money wage increases they received were equivalent to real wage increases.

But inflationary expectations adapt to the actual higher inflation rate after a time. Once workers increase their inflationary expectations to  $\dot{P}_2$  then the SRPC shifts out, because their expectations of the underlying rate of price inflation have risen. The labour market will then settle at Point C, which is consistent with the new expected inflation rate.

The path the labour market takes as inflationary expectations adjust to the actual inflation rate and the **short-run Phillips curve** shifts (that is, from Point B to Point C or Point D to Point E) is an empirical matter. But for Monetarists, once inflationary expectations have fully adjusted to the current inflation rate (at Points C and E, for example), the economy will return to the natural rate of unemployment ( $U^*$ ), irrespective of government attempts to target the lower unemployment rate.

For Friedman, the short-run dynamics of the labour market were driven by the capacity of the government to ‘fool’ workers into believing the inflation rate was lower than the actual inflation rate. As long as some of the inflation rate is unanticipated by workers, the government can maintain the unemployment rate below the natural rate but at a cost of rising inflation.

This narrative seeks to explain mass unemployment in the same way. The Friedman story is that mass unemployment occurs when workers refuse to accept money wage offers that they think generate a real wage below the actual real wage consistent with these offers. Workers think the real wage implied by the wage offer is too low because they wrongly believe that inflation is higher than it is. That is, their inflationary expectations exceed the actual inflation rate.

As a consequence, they start quitting their jobs and/or refuse to take new job offers thinking it is better to search for positions which offer higher real wages. Once they realise they have mistakenly thought inflation was higher than it actually is, they start to accept the job offers at the current money wage levels and increase their labour supply.

Friedman was thus forced to explain changes in unemployment in terms of swings in the supply of labour, driven by misconceptions of what the actual inflation rate was. At the empirical level this theory predicts that quits will fall as employment rises.

If the quit rate was, indeed **counter-cyclical**, the resulting changes in labour supply would be consistent with Friedman's theory. The empirical evidence is that quit rates are pro-cyclical, which means they rise when the labour market is strong and workers feel confident about their chances of securing a new job (after quitting their current jobs) and fall when the labour market is weak and workers fear ongoing unemployment. This is exactly the opposite to what would be required to substantiate Friedman's natural rate theory.

American economist, Lester Thurow summarised this issue succinctly:

... why do quits rise in booms and fall in recessions? If recessions are due to informational mistakes, quits should rise in recessions and fall in booms, just the reverse of what happens in the real world. (Thurow 1983: 185)

The introduction of the role of inflationary expectations in the Phillips curve focused attention on how such expectations were formed. What behavioural models could be invoked to capture expectations? There were two main theories advanced by economists: (a) adaptive expectations, and later (b) rational expectations.

Both theories considered the formation of expectations to be endogenous to the economic system. That is, developments within the system conditioned the way in which workers (and firms) formed views about the future course of inflation. We consider the implications of these two theories in the Advanced Material – Inflationary expectations in the Appendix.

## 11.6 Hysteresis and the Phillips Curve Trade-Off

While the focus of economists was on trying to estimate how fast individual expectations responded to rising inflation and the role that inflationary expectations played in wage and price formation, a new strand of literature emerged which challenged the Monetarist contention that there was no long-run trade-off between inflation and the unemployment rate.

At the empirical level it was noted that the estimates of the unobserved natural rate (sometimes called the **Non-Accelerating Inflation Rate of Unemployment (NAIRU)**) derived from econometric models seemed to track the actual unemployment rate with a lag.

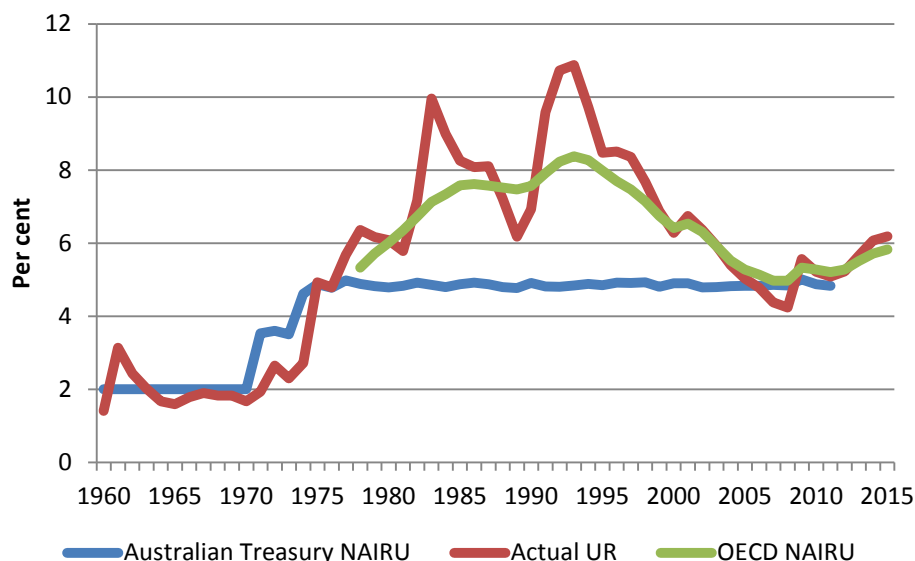
At the time that Monetarism became influential in economic policy making, unemployment rates were rising due to the policy response to the major oil price rises in the early and mid-1970s that caused accelerating inflation. We will examine this period separately later. The estimates of the natural rate seemed to rise too and without any consistent explanation.

In Figure 11.6, we show Australian Treasury and OECD estimates of the NAIRU and the corresponding unemployment rate for Australia from 1960 to 2015. The Treasury estimates end in 2011 while the OECD estimates begin in 1978. First, why did the Treasury estimates of the NAIRU, which are meant to reflect 'structural' factors jump so violently in 1974, around the same time the actual unemployment rate rose sharply? This period was highly turbulent (OPEC oil crisis) and marked the end of the post-war full employment era when unemployment rates were usually below 2 per cent. You can clearly see that the estimated NAIRU tracks the actual unemployment rate upwards. No coherent explanation has ever been given to explain that jump.



Structural factors tend to impact slowly and gradually. Second, why are the Treasury estimates flat after that and quite different to the OECD estimates, which tend to track the actual unemployment rate? Third, with the exception of 2001, the OECD estimates of the NAIRU steadily declined from 1993 to 2007. Likewise, there was a steady decline in OECD estimates of the NAIRU for the UK and USA for more than 10 years from the early 1990s.

**Figure 11.6 Annual Australian unemployment rate, Treasury and OECD NAIRU estimates, 1960-2015**



These observations led economists to question the idea that there was a **cyclically-invariant** natural rate of unemployment. It appeared that the best estimates of the unemployment rate that was consistent with stable inflation at any point in time were highly cyclical, since they followed the actual unemployment rate, which reflects the business cycle.

While there are various explanations that have been offered to rationalise the way the estimated natural rates of unemployment fell over the 1990s (for example, demographic changes in the labour market with youth falling in proportion), one plausible explanation is that there is no separate informational content in these estimates and they just reflect in some lagged fashion the dynamics of the unemployment rate – that is, the **hysteresis hypothesis**.

This was a new theory that emerged to explain the apparent cyclical relationship between the equilibrium unemployment rate and the actual unemployment rate (see also Chapter 9).

The early work showed that the increasing NAIRU estimates (based on econometric models) merely reflected the decade or more of high actual unemployment rates and restrictive fiscal and monetary policies. Thus these estimates were not indicative of increasing structural impediments in the labour market, due to, for example, demographic changes, which could result in an influx of young unskilled workers into the labour market or rising minimum wages and/ or welfare distortions, such as more generous unemployment benefits.

The **hysteresis effect** describes the interaction between the actual and equilibrium unemployment rates. The significance of hysteresis is that the unemployment rate associated

with stable prices, at any point in time should not be conceived of as a rigid non-inflationary constraint on expansionary macro policy. The equilibrium rate itself can be reduced by policies, which reduce the actual unemployment rate.

The importance of hysteresis is that a long-run inflation-unemployment rate trade-off can still be exploited by the government and one of the major planks of Monetarism would be invalid.

One way to explain this phenomenon is to focus on the way in which the labour market adjusts to cyclical changes in economic activity.

Recessions cause unemployment to rise and due to their prolonged nature, the short-term joblessness becomes entrenched in long-term unemployment. The unemployment rate behaves asymmetrically with respect to the business cycle, which means that it jumps up quickly but takes a long time to fall again.

The non-wage labour market adjustment that accompanies a low-pressure economy, which could lead to hysteresis, is well documented. Training opportunities are provided with entry-level jobs and so the (average) skill of the labour force declines as vacancies fall.

New entrants are denied relevant skills (and socialisation associated with stable work patterns) and redundant workers face skill obsolescence.

Both groups need jobs in order to update and/or acquire relevant skills. Skill (experience) upgrading also occurs through mobility, which is restricted during a downturn.

The idea is that structural imbalance increases in a recession due to the cyclical labour market adjustments commonly observed in downturns, and decreases at higher levels of demand as the adjustments are reversed. Structural imbalance refers to the inability of the actual unemployed to present themselves as an effective excess supply.

## **11.7 Underemployment and the Phillips Curve**

As we saw in Chapter 9, underemployment has become an increasingly significant component of labour underutilisation in many nations over the last two decades. In some nations, such as Australia, the rise in underemployment has outstripped the fall in official unemployment. National statistical agencies have responded to these trends by publishing more regular updates of underemployment. They have also constructed new data series to provide broader measures of labour wastage (for example, the Australia Bureau of Statistics Broad Labour Underutilisation series, which is published on a quarterly basis).

After the major recession that beset many nations in the early 1990s, unemployment fell as growth gathered pace. At the same time, inflation also moderated and this led economists to increasingly question the practical utility of the concept of a cyclically invariant natural rate (or NAIRU) for policy purposes, quite apart from the conceptual disagreements.

This scepticism was reinforced because various agencies produced estimates of the natural rate of unemployment that declined steadily throughout the 1990s as the unemployment rate fell (see Figure 11.6). As the unemployment rate went below an existing natural rate estimate (and inflation continued to fall) new estimates of the natural rate were produced, which showed it had fallen further. This reinforces our conclusion that NAIRU estimates have no predictive capacity in relation to the relationship between movements in the unemployment rate and the inflation rate.

The question then arises as to why the unemployment rate and the inflation rate both fell in many nations during the 1990s. What does this mean for the Phillips curve?

To understand this more fully, economists started to focus on the concept of the excess supply of labour, which is a key variable constraining wage and price changes in the Phillips curve framework.

The standard Phillips curve approach predicts a statistically significant, negative coefficient on the official unemployment rate (a proxy for excess demand). However, the hysteresis model suggests that state dependence is positively related to unemployment duration and at some point the long-term unemployed cease to exert any threat to those currently employed.

Consequently, they do not discipline the wage demands of those in work and do not influence inflation. The hidden unemployed are even more distant from the wage setting process. So we might expect that the short-term unemployment is a better excess demand proxy in the inflation adjustment function.

While the short-term unemployed may be proximate enough to the wage setting process to influence price movements, there is another significant and even more proximate source of surplus labour available to employees to condition wage bargaining – the underemployed.

The underemployed represent an untapped pool of additional working hours that could be clearly redistributed among a smaller pool of persons in a relatively costless fashion if employers so desired.

It is thus reasonable to hypothesise that the underemployed pose a viable threat to those in full-time work who might be better placed to set the wage norms in the economy.

This argument is consistent with research in the institutionalist literature that shows that wage determination is dominated by insiders (the employed) who set up barriers to isolate themselves from the threat of unemployment. Phillips curve studies have found that within-firm excess demand for labour variables (like the rate of capacity utilisation or rate of overtime) to be more significant in disciplining the wage determination process than external excess demand proxies, such as the unemployment rate.

It is plausible that while the short-term unemployed may still pose a more latent threat than the long-term unemployed, the underemployed are also likely to be considered an effective surplus labour pool. In that case we might expect downward pressure on price inflation to emerge from both sources of excess labour.

Figure 11.7 shows the relationship between the unemployment rate and inflation in Australia between 1978 and 2011. The sample is split into three sub-samples. The first from March 1978 to September 1983 is defined by the starting point of the most recent consistent Labour Force data (February 1978) and the peak unemployment rate from the 1982 recession (September 1983).

The second period December 1983 to September 1993 depicts the recovery phase in the 1980s and then the period to the unemployment peak that followed the 1991 recession. The final period goes from December 1993 to September 2015.

The solid lines are simple linear trend regressions. The dotted black line with the arrow head links the September 1995 observation (top) with the September 1997 observation (at arrow head).

**Figure 11.7 Inflation and unemployment, Australia, quarterly data, 1978-2015**



Source: Australian Bureau of Statistics, Consumer Price Index and Labour Force.

The relationship between the annual inflation rate and the unemployment rate clearly shifted after the 1991 recession. Focusing on the dotted arrow line (joining September 1995 and September 1997), this was a period when the Phillips curve began to flatten and move inwards. Over these years, the unemployment rate was stuck due to a lack of aggregate demand growth but the inflation rate was falling.

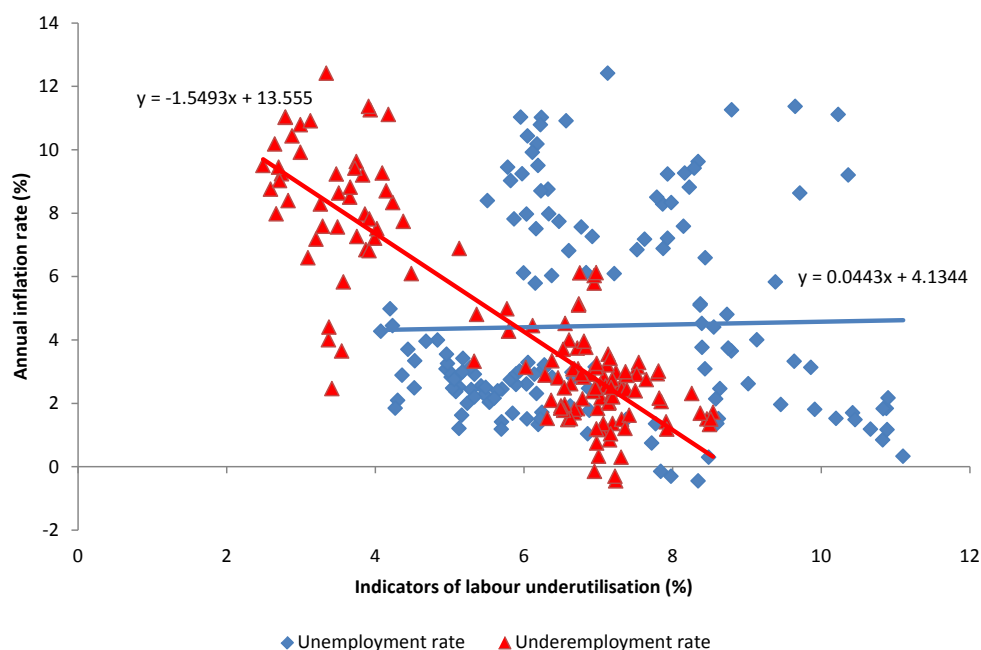
This has been explained, in part, by the fall in inflationary expectations. The 1991 recession was particularly severe and led to a sharp drop in the annual inflation rate and with it a decline in survey-based inflationary expectations.

The other major labour market development that arose during the 1991 recession was the sharp increase and then persistence of high underemployment as firms shed full-time jobs, and as the recovery got underway, began to replace the full-time jobs that were shed with part-time opportunities. Even though employment growth gathered pace in the late 1990s, a majority of those jobs in Australia were part-time. Further, the part-time jobs were increasingly of a casual nature.

Figure 11.8 shows the relationship between unemployment and inflation in Australia from 1978 to 2013. It also shows the relationship between the underemployment estimates provided by the Australian Bureau of Statistics and annual inflation for the same period.

The equations shown are the simple regressions, depicted graphically by the solid lines. The graph suggests the negative relationship between inflation and underemployment is stronger than the relationship between inflation and unemployment. More detailed econometric analysis confirms this to be the case.

**Figure 11.8 Inflation and unemployment and underemployment, Australia, quarterly data, 1978-2013**



Source: Australian Bureau of Statistics, Consumer Price Index and Labour Force.

The inclusion of underemployment in the Phillips curve specification helps explain why low rates of unemployment have not been inflationary in the period leading up to the Global Financial Crisis. It suggests that changes in the way the labour market operates – with more casualised work and underemployment – have been significant in explaining the impact of the labour market on wage inflation and general price level inflation.

Monetary policy has been the main arm of macroeconomic policy since the 1970s, with the primary objective of controlling inflation. In Chapter 15, we shall discuss the performance of the Reserve Bank of Australia in meeting its inflation target of 2 – 3 percent annual inflation and achieving full employment.

## 11.8 Demand-Pull and Cost-Push Inflation

Economists distinguish between cost-push and demand-pull inflation although, as we will see, the demarcation between the two types of inflation is not as clear-cut as one might think.

Demand-pull inflation refers to the situation where prices start accelerating continuously because nominal aggregate demand growth outstrips the capacity of the economy to respond by expanding real output.

Gross Domestic Product (GDP) is the market value of final goods and services produced in some period. We represent that as the product of total real output ( $Y$ ) and the general price level ( $P$ ), that is,  $GDP = PY$ .

We have learned from the National Accounts, that aggregate demand is always equal to GDP or  $PY$ . It is clear that if there is growth in nominal spending that cannot be met by an increase in real output ( $Y$ ) then the general price level ( $P$ ) has to rise.

The standard price Phillips curve model that evolved in the 1960s was supported by the dominant view of inflation at the time based on Keynes' notion of an inflation gap.

In his 1940 pamphlet, *How to pay for the war: a radical plan for the chancellor of the exchequer*, John Maynard Keynes outlined his concept of the inflationary gap.

He wanted to apply the notion of effective demand that he developed in the *General Theory*, to help understand how an equilibrium corresponding to less than full employment could arise in a monetary economy, but there would be a transition to a fully employed economy during wartime.

With the onset of World War II, large-scale spending programs were implemented. Keynes argued that, as employment rose, rising household incomes, would drive up consumer spending, which would cause inflation to accelerate, even if money wage rates were constant.

While Keynes' plan was devised in the context of war-time spending when faced by tight supply constraints (that is, a restricted ability to expand real output), the concept of the inflationary gap has been generalised to describe situations of excess demand, where aggregate demand is growing faster than the aggregate supply capacity can absorb it.

He defined the inflationary gap as an excess of planned expenditure over the available output at pre-inflation or base prices. The pre-inflation benchmark output was that corresponding to the full utilisation of capacity. Thus, if an economy could meet the growth in nominal expected demand by rapidly expanding the capacity to produce goods and services, an inflationary gap would not open.

This idea was distilled into the **demand-pull theory of inflation**. Once full employment was reached then nominal demand growth beyond that level would be inflationary.

The Phillips curve is clearly consistent with this view of inflation. The theory claimed that as nominal demand growth pushed the unemployment rate towards its irreducible minimum (frictional unemployment), wage and price inflation would start to rise. In other words, an inflationary gap would be created by the emergence of excess aggregate demand.

There are several factors present in the real world that attenuate these demand effects on the inflation rate.

First, there are also extensive costs incurred by firms when they change prices, which leads to a 'catalogue' approach where firms will forecast their expected costs over some future period and set prices according to their desired return. They then signal those prices in their catalogues and advertising to consumers and stand ready to supply whatever is demanded at that price (up to exhaustion of capacity). In other words, they do not frequently alter their prices to reflect changing demand conditions. Only periodically will firms typically revise their price catalogues.

Second, trust and reliability are important in economic transactions. Firms for example, seek to build relationships with their customers that will ensure product loyalty. In this context, firms will not seek to vary prices once they are notified to consumers.

Third, firms also resist cutting prices when demand falls because they want to avoid so-called adverse selection problems, where they gain a reputation only as a bargain price supplier. Firms value 'repeat sales' and thus want to foster consumer goodwill.

Circumstances change somewhat when the economy approaches full productive capacity. Then the mix between real output growth and price rises becomes more likely to be biased toward price rises (depending on bottlenecks in specific areas of productive activity). At full capacity, GDP can only grow via inflation (that is, nominal values increase only). At this point the inflationary gap is breached.

Keynes also suggested that inflation could arise due to **cost-push factors** (also called sellers' inflation). There had been a long line of authors who had identified inflation emerging as a result of **distributional struggle** over the available real income being produced, including **Michal Kalecki** who wrote from a Marxist perspective. We will consider the underlying forces that might drive these cost-push factors in the next section.

For now, we recognise that, in product markets, firms have price setting power and set prices by applying some form of profit mark-up to costs. Firms seek to achieve target profit rates that satisfy their shareholders or owners, which is expressed in the size of the mark-up on their unit costs. Unit costs are driven largely by wage costs, productivity movements and raw material prices.

Shifts in any of these determinants can generate cost pressures, which price-setting firms may respond to by passing on as increased prices.

For example, reductions in the **reserve army of unemployed** as the economy approaches full employment gives workers more bargaining power. Trade unions are more likely to demand higher money wages. Firms may fear prolonged strikes, which will damage them at a time when profits are high. To protect their market share they are more likely, under these circumstances, to concede to the workers' demands, knowing that they can, in turn, use their price-setting power to defend their real profits by increasing prices (that is, restore the previous mark-up).

Another example of a cost-push pressure might come from an increase in a significant imported raw material price, such as oil. Such a price rise will impact on the profit margin and firms will be motivated to pass the cost rise on in the form of higher prices. Workers may then respond to the real wage cuts that arise from the rise in prices by pressing for money wage increases. We will examine this dynamic in the next section.

For now, it is sufficient to recognise that the Phillips curve could also subsume the cost-push theories of inflation, since they are also consistent with a negative relationship between unemployment and inflation, given that workers tend to be more aggressive with respect to bargaining at lower levels of unemployment. The idea of stagflation that Lerner advanced in the early 1950s would also help understand the empirical instability in the Phillips curve that began to manifest in the late 1960s, which led to a major shift in macroeconomic thinking.

When the US government prosecuted the Vietnam War effort in the 1960s, the inflation rate began to rise. In the late 1960s and early 1970s, the demand-pull pressures of the spending associated with the war effort combined with sharp rises in oil prices following the formation of the Organisation of Petroleum Exporting Countries cartel (OPEC). OPEC oil prices quadrupled in 1973 and generated huge cost shocks to oil-dependent economies such as the US and Japan.

## 11.9 The Conflict Theory of Inflation

The **Conflict Theory of Inflation** situates the problem as being intrinsic to the power relations between workers and capital (class conflict), which are mediated by government within a capitalist system. It brings together social, political and economic considerations into a generalised view of the inflation cycle. This mediation by government varies over the course of history and in more recent times has been biased towards protecting the interests of capital, particularly financial capital, at the expense of workers' real wage aspirations.

The conflict theory derives directly from cost-push theories referred to above. Within the Keynesian tradition, in Chapter 14 of Abba Lerner's 1951 book - *Economics of Employment* there is a coherent discussion of how distributional struggle may lead to a wage-price spiral and generalised inflation.

This dynamic could easily lead to a series of wage and price rises as each party seeks to defend their real stake in production. As a result, inflation becomes the product of distributional struggle over real income shares, reflecting the relative bargaining strengths of workers and employers.

Lerner showed that the dynamic for this wage-price spiral could also result from capital seeking to expand its real share of income by pushing up the mark-up on unit costs. Such a strategy could only be successful if workers conceded the real wage cut implied by the rising prices. Firms would be more likely to attempt this strategy when they perceived the bargaining power of workers to be weak - that is, when the unemployment rate was higher. In this way, Lerner recognised that high inflation and high unemployment could co-exist - that is, he foreshadowed the possibility of **stagflation**.

Conflict theory recognises that the money supply is endogenous, which is in contradistinction to the Monetarist's Quantity Theory of Money that erroneously considers that the money supply is exogenously controlled by the central bank (see Chapter 10).

From the workers' perspective, real wages growth (increasing the capacity of the nominal or money wage to command real goods and services) is a primary aim of wage bargaining.

Firms have an incentive to resist real wages growth that is not underpinned by productivity growth because they would have to 'pay' for it by either reducing their real margins or raising prices. They also try to capture as much productivity growth in the form of increased profits as they can.

The conflict theory assumes that firms and trade unions have some degree of market power (that is, they can influence prices and wage outcomes) without much correspondence to the state of the economy. They are assumed to both desire some targeted **real output share** and use their capacity to influence **nominal prices and wages** to extract that real share.

In each period, the economy produces a given real output (real GDP), which is shared between the groups with distributional claims in the form of wages, profits, taxes etc. In the following discussion, we assume away the other claimants and concentrate only on the split between wages and profits. Later, we will introduce a change in an exogenous claimant in the form of a rise in the price of a significant raw material.

If the desired real output shares of the workers and firms is consistent with the available real output produced, then there is no incompatibility and there will be no inflationary pressures. The



available real output would be distributed each period in the form of wages and profits, which satisfy the respective claimants.

The capacity of workers to realise nominal wage gains is considered to be pro-cyclical – that is, when the economy is operating at **high pressure** (high levels of capacity utilisation) workers are more able to succeed in securing money wage gains. This is especially the case if they are organised into coherent trade unions, which function as a countervailing force to offset the power of the employer.

When employers are dealing with workers individually they have more power than when they are dealing with one bargaining unit (trade union), which represents all workers in their workplace.

The pro-cyclical nature of the bargaining power held by workers arises because unemployment is seen as disciplining the capacity of workers to gain wages growth – in line with Marx's reserve army of labour.

In this context, a so-called **battle of the mark-ups** can arise where workers try to get a higher share of real output for themselves by pushing for higher money wages and firms then resist the squeeze on their profits by passing on the rising costs – that is, increasing prices to restore the mark-up.

At that point there is no inflation – just a once-off rise in prices and no change to the distribution of national income in real terms.

The problem arises when the sum of the distributional claims (expressed in nominal terms – money wage demands and mark-ups) are greater than the real output available measured at current prices and neither bargaining party desires to concede to the other in real terms. In those circumstances, inflation can occur via the **wage-price or price-wage spiral** mechanisms.

It is here that the concept of real wage and/or real profit margin resistance becomes relevant. A wage-price spiral begins with workers pushing for higher real wages whereas a price-wage spiral refers to a dynamic where firms initiate the bargaining war by trying to push up their real profit margin.

In the latter case, if the economy is operating at high pressure, workers may resist the attempt by firms to increase their real profit margin. They may seek to maintain their previous real wage and will thus respond to the increasing price level by imposing further nominal wage demands. If their bargaining power is strong (which from the firm's perspective is usually in terms of how much damage the workers can inflict via industrial action on output and hence profits) then they are likely to be successful. If not, they may have to accept the real wage cut imposed on them by the increasing price level.

If firms are not willing to absorb the squeeze on their real output claims then they will raise prices again and the beginnings of a price-wage spiral begins. If this process continues then a cost-push inflation is the result.

The dynamic that drives a cost-push inflation is seen to arise from the underlying social relations in the economy. It is here that we can consider a general theory of inflation, which recognises that the two sides of the labour market are likely to have conflicting aims and seek to fulfil those aims by imposing real costs on the other party.

The wage-price spiral might also become a **wage-wage-price spiral** as one section of the workforce seeks to restore relativities after another group of workers succeeds in their nominal wage demands.

The role of government is also implicated. While it is the distributional conflict over available real output which initiates the inflationary spiral, government policy has to be compliant for the nascent inflation to persist.

Business firms will typically access credit (for example, overdrafts) to **finance** their working capital needs in advance of realisation of revenue via sales. In an inflationary spiral, as workers seek higher nominal wages, firms will judge whether the costs of industrial action in the form of lost output and sales are higher than the costs of accessing credit to fund the higher wages bill. Typically, the latter option will be cheaper.

If credit conditions become tighter and thus loans become more expensive, then firms will be less able to pay the higher money wages demanded by workers. The impact of the higher interest rates may thus lead to a squeeze on real wages with the consequent negative impact on consumption spending. Firms will also be less willing to invest in new projects given that the cost of funds is higher.

As a consequence, if monetary policy becomes tighter there will be some point where real production growth declines and the workers who are in weaker bargaining positions are laid off. The rising unemployment, in turn, eventually discourages the workers from pursuing their on-going demand for wage increases and in time the inflationary process is choked off.

The Conflict Theory of Inflation thus hypothesises a trade-off between inflation and unemployment.

The alternative policy stance is for the central bank to accommodate the inflationary struggle by leaving its monetary policy settings (interest rates) unchanged. This accommodation would also likely see the fiscal authorities maintaining existing tax rates and spending growth.

The commercial banks would continue to extend loans and in the process, create deposits in the accounts of its business clients. The central bank would then ensure that there were sufficient reserves in the banking system to maintain stability in the payments system. The nominal wage-price spiral would thus fuel the demand for more loans with little constraint.

There are also strong alignments between the conflict theory of inflation and Hyman Minsky's financial instability notion. Both theories consider that behavioural dynamics vary across the business cycle. When economic activity is strong, the risk-averseness of banks declines and they become more willing to extend credit to marginal borrowers. Equally, firms will be more willing to pass on nominal wage demands because it becomes harder to find labour and the costs of an industrial dispute in terms of lost sales and profits are high. Workers also have more bargaining power due to the buoyant conditions.

At low levels of economic activity, the falling sales and rising unemployment militates against profit push and wage demands. It also is associated with higher loan delinquency rates and banks become more conservative in their lending practices.

## 11.10 Raw Material Price Increases

Up until now we have been concentrating on workers pursuing nominal wage increases in order to gain higher real wages and/or firms pushing profit margins up to gain a greater profit share of real income as the main drivers of an inflationary process.

However, **raw material price shocks** can also trigger cost-push inflation. These cost shocks may be imported (for example, an oil-dependent nation might face higher energy prices if world oil prices rise) or domestically sourced (for example, a nation may experience a drought which increases the costs of food which impacts on all food processing industries).

Take for example the situation where there is a price rise for an essential imported resource. The imported resource price shock amounts to a loss of real income for the nation in question. That is, there is less real income to distribute to domestic claimants.

The question then is who will bear this loss? With less real income being available for distribution domestically, the reactions of the claimants are crucial to the way in which the economy responds to the impost.

The loss has to be shared or borne by one of the claimants or another. If local firms pass the raw material cost increases on in the form of high prices, then workers would endure a cut in their real wages.

If workers resist this erosion of their real wages and push for higher nominal wages growth then firms can either accept the squeeze on their profit margin or resist.

You can see that the dynamics of the Conflict Theory of Inflation are triggered by the raw material price rise.

The government can employ a number of strategies when faced with this dynamic. It can maintain the existing nominal demand growth, which would be very likely to reinforce the spiral.

Alternatively, it can use a combination of strategies to discipline the inflation process including the tightening of fiscal and monetary policy to create unemployment (the NAIRU strategy); the development of consensual incomes policies and/or the imposition of wage-price guidelines (without consensus) (see below).

Ultimately, if the claimants of real income continue to try to pass the raw material price rise onto each other, then it is likely that contractionary government policy will be introduced and unemployment will rise.

## 11.11 Cost-Push and Demand-Pull Inflation Summary

A cost-push inflation requires certain aggregate demand conditions for it to be sustained. In this regard, it is hard to differentiate between an inflationary process which was initiated from supply-side pressure from one that was initiated by demand pressures.

For example, an imported raw material shock means that a nation's real income that is available for distribution to domestic claimants is lower. This will not be inflationary unless it triggers an on-going distributional conflict as domestic claimants (workers and capital) try to pass the real loss onto each other.

However, that conflict needs ‘oxygen’ in the form of on-going economic activity in sectors where the spiral is robust. In that sense, the conditions that will lead to an accelerating inflation – high levels of economic activity – will also sustain an inflationary spiral emanating from the demand-side.

## 11.12 Incomes Policies

Governments facing a wage-price spiral and who are reluctant to introduce a sharp contraction in the economy, which might otherwise discipline the combatants in the distribution struggle, have from time to time, considered the use of so-called **incomes policies**.

These policies have been introduced, in various forms, in many countries as a way of reducing supply-side cost pressures and allowing employment to stay at a higher level.

Incomes policies in general, are measures that are aimed to control the rate at which wages and prices rise, typically as the economy moves towards, or is at full employment.

In the context of the Phillips curve, incomes policies were seen as a way of flattening the Phillips curve and reducing the inflationary impact of a reduction in unemployment.

Many countries have at various times introduced these types of policies.

For example, in 1962 the US government introduced wage-price guideposts, which allowed for an average rate of nominal wage increase equal to the average annual rate of productivity growth in the overall economy. Other nominal incomes, including profits, were to be tied to this rule.

Taken together, it was considered that this rule would stabilise the growth in nominal incomes and reduce any inflationary pressures associated with the maintenance of full employment.

The rule also sought to distribute productivity gains across all income earners and thus reduce the distributional conflict, which may instigate a wage-price spiral.

For a time, the guidelines seemed to work. But as US government expenditure grew as a result of its prosecution of the Vietnam War effort and unemployment fell below four per cent, wage increases began to exceed average productivity growth. By 1966, the guidelines provided no discipline on the growth of nominal incomes in the US.

It was clear that the US government was unable to compel employers to follow the guideposts in the wage bargaining process.

Despite the failure of the wage-price guideposts, the Republican administration under Nixon reintroduced an incomes policy in 1971. Initially, this was in the form of a 90-day freeze on wages and other nominal incomes. Later, compulsory growth guidelines were set for wages and prices growth and these were replaced with a voluntary mechanism.

Soon after (in 1973), the government introduced yet another freeze on prices, followed by sector-by-sector price rises in line with cost increases with a freeze on profit margins. The experiment ended in April 1974.

It is argued that the institutional structures that made economies more susceptible to distributional conflict in the late 1960s and early 1970s also made the operation of incomes policies difficult. Highly concentrated industries with large firms exercising significant price-setting power were interacting with strong trade unions.

These firms are in a strong position to pass on wage demands in the form of higher prices and governments are reluctant, or are unable constitutionally, to mandate strict wage-price controls in normal times.

The other problem with average productivity rules is that they undercompensate workers in above-average productivity growth sectors and overcompensate workers in below-average sectors.

However, incomes policies have worked more effectively in some European nations, for example, Austria and in Scandinavian countries. These nations have long records of collective bargaining and are more attuned to tri-partite negotiations than the English-speaking nations.

A good example of a successful income policy approach, where wages and prices growth was driven by productivity growth in certain sectors, is the so-called Scandinavian Model (SM) of inflation, which is outlined in the Appendix.

By the 1970s, with the rising dominance of Monetarism, which eschewed institutional solutions to distributional conflict in favour of market-based approaches, incomes policies lost favour in most countries.

The Monetarist approach combined the use of persistently high unemployment and increased policy attacks on trade unions in many advanced nations to reduce the bargaining power of workers. This reduced the inflationary tendency because workers were unable to pursue real wages growth and as a result productivity growth outstripped real wages growth. This led to a substantial redistribution of real income towards profits during this period.

The rise of Thatcherism in the UK exemplified this increasing dominance of the Monetarist view in the 1980s.

In the next Chapter we will introduce the concept of buffer stocks in a macroeconomy (employment and unemployment) and analyse how they can be manipulated by policy to maintain price stability.

## Appendix: Advanced Material

### Phillips curve algebra

We have discussed the original Phillips curve (Phillips, 1958), which was a relationship between the growth in money wages and the unemployment rate.

The simplest non-linear form of the original Phillips curve relationship is given as:

$$(11.3) \quad \dot{W}_t = \alpha_0 + \beta U_t^{-1} \quad \beta > 0$$

where  $\dot{W}_t$  is the rate of money wage inflation (the dot above the  $W$  symbolising a rate of change),  $\alpha_0$  is a constant,  $\beta$  is a coefficient which tells us how the rate of wage inflation responds to the excess demand variable,  $U_t^{-1}$  and the subscript  $t$  denotes the current period.

The term  $\beta U_t^{-1}$  measures the impact of the state of the labour market on money wage growth, so that as the rate of unemployment rises, the rate of money wage inflation falls.  $U_t^{-1}$  is the inverse of the unemployment rate to capture the non-linear shape of the Phillips curve as hypothesised by Phillips himself and later Samuelson and Solow. The rate of money wage inflation asymptotes to a value of  $\alpha_0$  as the rate of unemployment increases.

You should recall from Chapter 7 (The Real Expenditure Model) that the convention is to impose positive values on coefficients, such as the marginal propensity to import,  $m$ , and to impose appropriate signs on the corresponding terms within the equation, so imports appear in the aggregate expenditure expression with a negative sign. On the other hand, a constant (intercept) term appears with a positive sign, but may be positive or negative. In this non-linear specification the  $U_t^{-1}$  term appears with a positive sign and parameter  $\beta$  is positive.

We noted in the discussion above, that when money wages grow in line with labour productivity there would be no inflationary pressures coming from the labour market. In other words, price inflation (holding other cost factors constant) will be equal to the growth in money wages minus labour productivity growth, say  $\dot{\gamma}$ . Then the price inflation equation can be written as:

$$(11.4) \quad \dot{P}_t = \dot{W}_t - \dot{\gamma} = \alpha_0 + \beta U_t^{-1} - \dot{\gamma} = \alpha + \beta U_t^{-1}$$

where  $\alpha = \alpha_0 - \dot{\gamma}$

A simpler derivation is in linear terms so the money wage Phillips curve is written as:

$$(11.5) \quad \dot{W}_t = \alpha_0 - \beta U_t \quad \beta > 0$$

where  $\beta$  is now the sensitivity of the growth of money wages to changes in the unemployment rate.

Then the expression for the price Phillips curve can be written as:

$$(11.6) \quad \dot{P}_t = \alpha - \beta U_t$$

where again  $\alpha = \alpha_0 - \dot{\gamma}$

This also tells us that the rate of general price inflation will be higher the lower is the unemployment rate and the lower is the productivity growth. If you compare Equations (11.4) and (11.6) you can see that each has an excess demand term relating to the rate of unemployment and a constant term.

## Econometric mis-specification

It was known that the Phillips curve became unstable (moved around) in the late 1960s and was particularly susceptible to sudden and/or large increases in inflation. The econometrically-estimated consumption functions in the large macroeconomic policy models, which were popular in the 1960s, also became unstable in the 1970s. Some economists successfully showed that the failure of the large-scale econometric models to forecast variables such as savings and consumption in the early 1970s could be traced to the mis-specification of the structural consumption function. Most of these models ignored the possibility that rising inflation would influence consumption (for example, if consumers expect prices to rise quickly in the future they may bring forward consumption decisions).

The breakdown of the Phillips curve in the late 1960s was another ‘econometric’ function that was mis-specified because it also ignored the possibility that rising inflation might become self-fulfilling as workers and firms sought to protect their real wages and real profit margins. This means that an **inflationary expectations** term should be included in the equation.

Another consideration as to why the discussions about instability were largely ignored is that the ‘textbook’ model of the Phillips curve was very attractive in its simplicity. Textbooks typically stylise discussions and eschew complicated stories for the sake of pedagogy. You should be aware that we have minimised this tendency in this text. We consider a rich treatment of institutions and history to be an important part of the learning process in macroeconomics.

It is a fact though, that the mainstream Keynesian consensus in the 1960s abstracted from the potential instability that was rooted in the institutional nature of wage and price setting. Instead, policy makers pursued the attractive notion that they could permanently maintain low unemployment rates as long as they ensured effective demand was sufficient relative to the non-government sector’s saving plans and any demand leakages from net exports.

## The expectations-augmented Phillips curve

Here we present a more analytical version of the Friedman natural rate hypothesis. The original Phillips curve related the growth of money wages to the unemployment rate. Friedman claimed that the simple version of the Phillips curve, whether specified in its original form or in the price inflation form, overlooked the fact that workers would be concerned about the growth in real wages. In other words, the rate of money wages growth would be influenced by the expected inflation rate, independently of the state of the labour market.

This conjecture led Friedman to incorporate a term for the influence of inflationary expectations in the wage bargaining process in the wage Phillips curve:

$$(11.7) \quad \dot{W}_t = \alpha_0 - \beta U_t + \phi \dot{P}_t^e \quad 0 \leq \phi \leq 1$$

The additional term  $\dot{P}_t^e$  represents inflationary expectations that are formed by workers, which condition the wage bargaining process. We assume that the coefficient  $\phi$  lies between 0 and 1. If  $\phi = 0$ , then wage inflation will only depend on the state of the labour market captured by the excess demand term ( $\beta U$ ). If  $\phi = 1$ , then any change in inflationary expectations is passed on fully to wages growth.

The subscripts might be confusing here. We assume that workers form expectations of inflation in period  $t$ , in the prior period and then bargain for wages growth in the current period based on what they think the inflation rate will be.

Adding this term to the Phillips curve led to the development of the **Expectations-Augmented Phillips Curve (EAPC)**:

$$(11.8) \quad \dot{P}_t = \alpha - \beta U_t + \varphi \dot{P}_t^e$$

and again the constant term is  $\alpha = \alpha_0 - \dot{\gamma}$ .

In terms of Figure 11.9, the inflationary expectations term on the right hand side of Equation (11.8) shifts the **short-run Phillips curve**, denoted by the remaining right-hand side terms.

If workers' inflationary expectations increase, then the short-run Phillips curve shifts out and vice-versa.

After the EAPC replaced the simple Phillips curve as the main framework for considering the relationship between inflation and unemployment, economists began to focus on the value of  $\phi$ . Many econometric studies were conducted to estimate its value.

Why does the value of  $\phi$  matter? What would happen if  $\phi = 1$ ?

Friedman defined the long-run steady-state (stable) inflation rate to occur when the actual rate of inflation was equal to the expected rate of inflation. That is, workers' inflationary expectations are consistent with the actual inflation rate. At this point, he claimed the economy would be operating at the natural rate of unemployment.

In that case, the EAPC would collapse to what is referred to as the long-run steady-state Phillips curve:

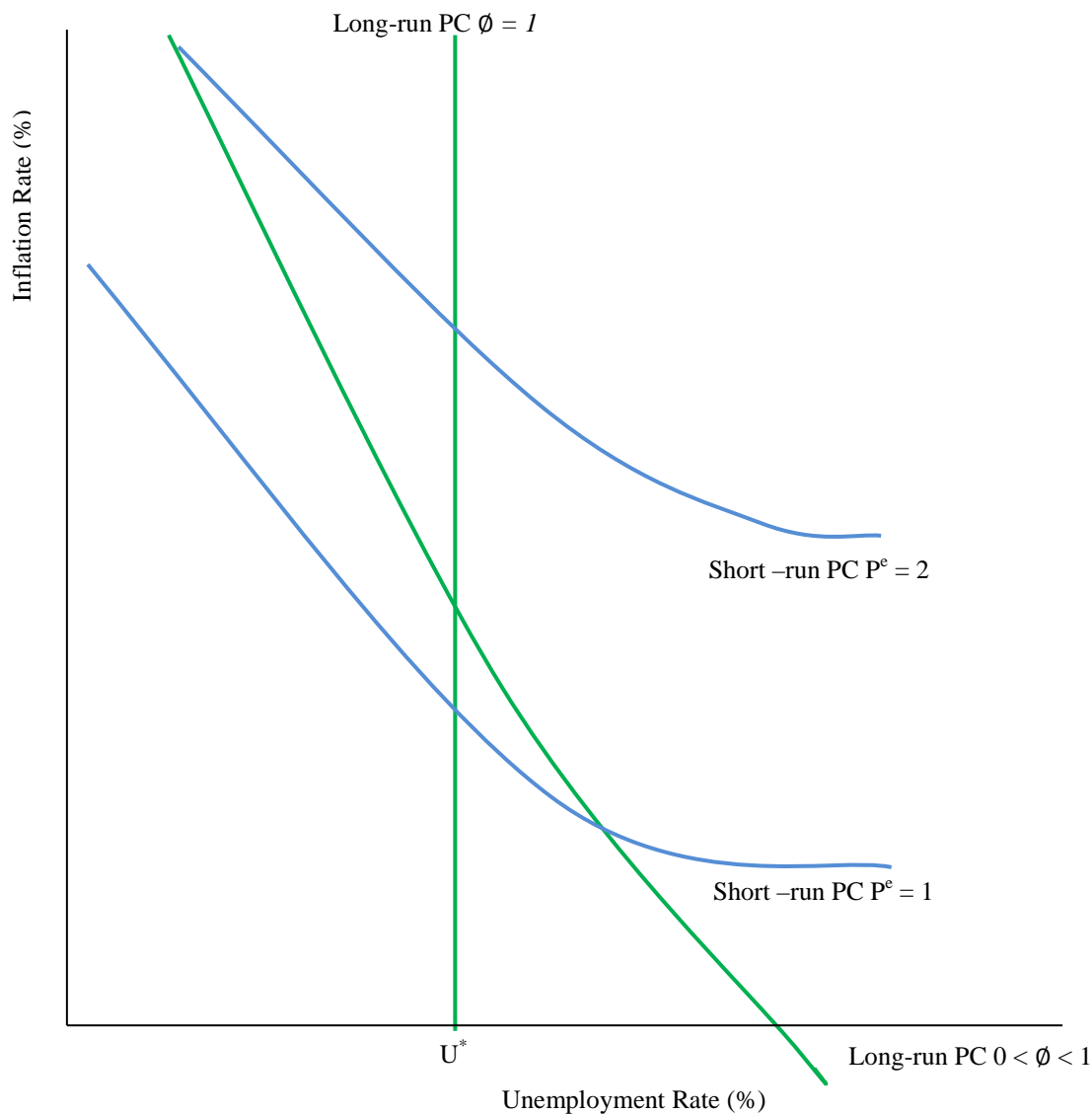
$$(11.9) \quad \begin{aligned} \dot{P}_t &= \alpha - \beta U_t + \varphi \dot{P}_t^e \\ \dot{P}_t(1 - \varphi) &= \alpha - \beta U_t \\ \dot{P}_t &= \alpha/(1 - \varphi) - \beta U_t/(1 - \varphi) \end{aligned}$$

Examine this relationship carefully because it looks similar to the short-run Phillips curve (Equation 11.6), except the coefficients are now divided by the term  $(1 - \phi)$ . The negative slope of this long-run Phillips curve,  $-\beta/((1-\phi))$  is steeper than the slope of the short-run Phillips curve,  $\beta$  and the closer  $\phi$  is to one, the steeper is the slope of the long-run Phillips curve. Once  $\phi$  equals one, the slope becomes vertical and there is no longer any relationship between inflation and the unemployment rate (see below). In other words, the trade-off vanishes.

Figure 11.9 depicts the two cases. There is a family of short-run Phillips curves (SRPC) (two are shown). The first long-run Phillips curve drawn on the assumption that  $0 < \phi < 1$ , is steeper than the short-run curves but non-vertical. It means that in the long run there is still a trade-off between the inflation rate and the unemployment rate, but it is a steeper trade-off than occurs in the short-run before inflationary expectations adjust upwards to the new inflation rate.



**Figure 11.9 Short- and long-run Phillips curves**



Note:  $P^e$  is the expected inflation rate. The higher the expected rate, the higher is the short-run Phillips curve.

The second long-run Phillips curve assumes that  $\phi = 1$  and is vertical as a consequence, which means there is no long-run trade off between inflation and the unemployment rate that can be exploited by the government. Under these assumptions, the economy always tends back to the natural rate of unemployment  $U^*$ , once inflationary expectations have adjusted to the actual inflation rate.

You can now see why economists who became captive of this framework were interested in the value of  $\phi$ . For Keynesians, a value of  $\phi$  less than one maintained their policy position that the government could use expansionary fiscal and monetary policy to reduce the unemployment rate should they consider the current rate to be too high.

For Monetarists, a value of  $\phi = 1$ , was consistent with their claims that the Keynesian aggregate demand management framework was flawed and would cause rising inflation should the

government try to push the unemployment rate below the natural rate, which is based on inflationary expectations being equal to the actual inflation rate.

Thus, at the time, the focus of the macroeconomic debate was on the value of  $\phi$ .

To see the way the natural rate of unemployment emerges out of this framework, we can solve Equation (11.9) for the long-run unemployment rate. After the relevant algebraic manipulation we get:

$$(11.10) \quad U^* = \alpha/\beta - \dot{P}_t(1 - \phi)/\beta$$

which shows there is still a trade-off in the long-run between unemployment and inflation as long as  $\phi \neq 1$ . Once,  $\phi = 1$ , the long-run unemployment rate becomes Friedman's natural rate and the equation representing that case is written as:

$$(11.11) \quad U^* = \alpha/\beta$$

This means that in the **Friedman natural rate hypothesis**, there are only two factors which influence the long run or natural rate of unemployment: (a) the rate of growth of productivity which is captured in the  $\alpha$  term; and (b) the short-run responsiveness of wage inflation to movements in the unemployment rate ( $\beta$ ). Note that given  $\beta$  is assumed to be positive, the term ( $\alpha/\beta$ ) is positive.

As a result, the higher is the growth in productivity, other things equal, the lower will be the natural rate. The Monetarists assumed that productivity growth was a structural phenomenon and invariant to aggregate demand policies.

It is clear, that in the Expectations-Augmented Phillips Curve framework, the government could only achieve temporary reductions in the unemployment rate below the natural rate as long as it could maintain a wedge between the expected inflation rate and the actual inflation rate, Once the workers' inflationary expectations adjusted, then the trade-off disappeared and the economy would return to the natural rate of unemployment, albeit with higher inflation.

Continued attempts at driving down the unemployment rate below the natural rate would, according to the Monetarists, just result in accelerating inflation.

### **Inflationary expectations**

#### *The Adaptive Expectations Hypothesis*

The assumption that workers formed their expectations of inflation in an adaptive manner allowed the Monetarists to conclude the government attempts to reduce the unemployment rate would only cause accelerating inflation and that the economy would always tend back to the natural rate of unemployment.

The only way the government could sustain an unemployment rate below the natural rate using aggregate demand stimulus would be if they continually drove the price level ahead of the money wage level and forced the workers to continually misperceive the true inflation rate.

The Adaptive Expectations hypothesis is expressed in terms of the past history of the inflation rate. The assertion is that the workers adapt their expectations of inflation as a result of learning from their past forecasting errors.

The following model expresses this idea:

$$(11.12) \quad \dot{P}_{t+1}^e = \dot{P}_t^e + \lambda(\dot{P}_t - \dot{P}_t^e) \quad 0 < \lambda < 1$$

The left-hand side of Equation (11.12) is the expected inflation rate in the next period ( $t + 1$ ) formed by workers in period  $t$ . Equation (11.12) has two components on the right-hand side.

First,  $\dot{P}_t^e$  is the expected inflation rate in the current period. Thus, workers use this inflation rate as a baseline to what they think the inflation rate in the next period will be.

Second, the term  $\lambda(\dot{P}_t - \dot{P}_t^e)$  captures the forecast error in the current period.  $\dot{P}_t^e$  was the expectation that workers formed in period  $t-1$  of the inflation rate in period  $t$ . The difference between that expectation and the actual rate that occurred is the size of their forecast error. The coefficient  $\lambda$  measures the strength of adaptation to error. The higher is  $\lambda$ , the more responsive workers' expectations will be to the actual rate of inflation. If  $\lambda = 1$ , then the expected inflation rate in period  $t+1$  is simply the actual rate of inflation in period  $t$ .

### *The Rational Expectations Hypothesis*

An extreme form of Monetarism, which became known as New Classical Economics posits that no policy intervention from government can be successful because so-called economic agents (for example, households and firms) form expectations in a rational manner.

This literature, which evolved in the late 1970s claimed that the stimulation of aggregate demand, say via fiscal policy, would be ineffective in real terms but highly inflationary.

The theory claimed that as economic agents formed their expectations rationally, they were able to anticipate any government policy action and its intended outcome and change their behaviour accordingly which would undermine the desired impact of the policy.

For example, individuals might anticipate a rise in government spending and predict that taxes would rise in the future to pay back the deficit. As a result the private individuals would reduce their own spending to save for the higher taxes and that action would thwart the expansionary impact of the public spending increase.

Recall that under adaptive expectations, economic agents are playing catch-up all the time. They adapt to past forecasting errors by revising their current expectations of inflation accordingly.

In this context, Monetarists like Milton Friedman claimed that the government could exploit a short-run Phillips curve for a time with expansionary policy by tricking workers into thinking their real wages had risen when in fact their money wage increases were lagging behind the inflation rate.

But Monetarists considered the unanticipated inflation would induce the workers to supply a higher quantity of labour than would be forthcoming at the so-called natural rate of output (defined in terms of a natural rate of unemployment).

Under adaptive expectations, the workers take some time to catch up with the actual inflation rate. Once they adjust to the actual inflation rate and realise that their real wage had actually fallen, they would withdraw their labour back to the natural level.

The use of adaptive expectations to represent the way workers adjusted to changing circumstances was criticised because it implied an implausible irrationality. In a period of continually rising prices, workers would never catch up. Why wouldn't they realise after a few

periods of errors that they were systematically under-forecasting and seek to compensate by overshooting the next period?

The theory of rational expectations was developed, in part, to meet these objections. When forming their expectations, economic agents were considered to act in a rational manner consistent with the assumptions in mainstream microeconomics pertaining to *Homo Economicus*.

This required that economic agents used all the information that was available and relevant at the time when forming their views of the future.

What information do they possess? The rational expectations (RATEX) hypothesis claims that individuals essentially know the true economic model that is driving economic outcomes and can make accurate predictions of these outcomes. Any forecasting errors are random. The proponents of RATEX said that predictions derived from rational expectations are on average accurate.

These proponents assumed that all people understood the economic model that policy makers use to formulate their policy interventions. The most uneducated person is assumed to have highly sophisticated knowledge of the structural specification of the economy that treasury and central banks deploy in their policy-making processes.

Further, people are assumed to be able to perfectly predict how policy makers will respond (in both direction and quantum) to past policy forecast errors. According to the RATEX hypothesis, people are able to anticipate both policy changes and their impacts.

As a result, any 'pre-announced' policy expansions or contractions will have no effect on the real economy. For example, if the government announces it will be expanding the deficit and adding new high powered money, we will also assume immediately that it will be inflationary and will not alter our real demands or supply (so real outcomes remain fixed). Our response will be to simply increase the value of all nominal contracts and thus generate the inflation that we predict via our expectations.

The government can thus never trick the private sector. The introduction of rational expectations into the debate, thus, went a step further than the Monetarists who conceded that governments could shift the economy from the 'natural level' by introducing unanticipated policy changes.

The New Classical Economics denied that governments could alter the course of the real economy at all. In other words, there was not even the possibility of a short-run trade-off between inflation and the unemployment rate. Workers would always know the future inflation rate and build it fully into each round of money wage bargaining.

The economy would thus always stay on the long-run Phillips curve.

While there are some very sophisticated theoretical critiques of the RATEX hypothesis (for example, the Sonnenschein-Mantel-Debreu theorem), which extend the notion of the fallacy of composition, some simple reflection suggests that the informational requirements necessary for the hypothesis to be valid are beyond the scope of individuals.

A relatively new field of study called behavioural economics has attempted to examine how people make decisions and form views about the future. The starting point is that individuals have what are known as cognitive biases, which constrain their capacity to make rational decisions.

RATEX-based models have failed to account for even the most elemental macroeconomic outcomes over the last several decades. They categorically fail to predict movements in financial, currency and commodity markets.

The 2008-09 Global Financial Crisis was not the first time that models employing rational expectations categorically failed to predict major events.

Rational expectations impose a mechanical forecasting rule onto individual decision-making when, in fact, these individuals exist in an environment of endemic uncertainty in which the future is unknowable.

As we will see in later Chapters, endemic uncertainty is a major problem facing decision-makers at all levels and of all types in a capitalist monetary economy. Uncertainty about economic events, such as movements in asset prices or job security, may encourage individuals to hold money, which is the most liquid of all assets and is a store of value.

In the real world, people have imperfect knowledge of what information is necessary for forecasting and even less knowledge of how this choice of information will impact on future outcomes. We also do not know how we will react to changing circumstances until we are confronted with them. The nature of endemic uncertainty is that we cannot know the full range of options that might be presented to us at some time in the future.

### **Hysteresis and the Phillips curve**

Here we will learn that if there is hysteresis present in the labour market, then a long-run trade-off between inflation and the unemployment rate is possible even if the coefficient on the augmented term in the Phillips curve (the coefficient on the inflationary expectations term) is equal to unity. This result was shown in Mitchell (1987).

At any point in time there might be an equilibrium unemployment rate, which is associated with price stability, in that it temporarily constrains the wage demands of the employed and balances the competing distributional claims on output. We might call this unemployment rate the **Macroequilibrium unemployment rate (MRU)**.

The interaction between the actual and equilibrium unemployment rates has been termed the hysteresis effect. The significance of hysteresis, if it exists, is that the unemployment rate associated with stable prices, at any point in time should not be conceived of as a rigid non-inflationary constraint on expansionary macro policy.

The equilibrium rate itself can be reduced by policies, which reduce the actual unemployment rate. Thus, we use the term MRU, as the non-inflationary unemployment rate, as distinct from the Monetarist concept of the NAIRU, to highlight the hysteresis mechanism, which is driven by the business cycle.

The idea is that structural imbalance increases in a recession due to the cyclical labour market adjustments commonly observed in downturns, and decreases at higher levels of demand as the adjustments are reversed. Structural imbalance refers to the inability of the actual unemployed to present themselves as an effective excess supply.

To see how hysteresis alters the Phillips curve, we start with a standard wage inflation equation such as:

$$(11.13) \quad \dot{W}_t = \alpha - \beta(U_t - U_t^*) + \theta \dot{P}_t^e$$

so that the rate of growth in money wages in time  $t$ ,  $\dot{W}_t$  is equal to some constant ( $\alpha$ ) which captures productivity growth and other influences, less the deviation of the unemployment rate from its steady-state value. The gap  $(U_t - U_t^*)$  is just a different way of capturing the excess demand in the labour market. If the gap is positive then the actual unemployment is above the MRU and there should be downward pressure on money wage demands, other things equal.

If the gap is negative then the actual unemployment is below the MRU and there should be upward pressure on money wage demands, other things equal.

The additional term captures inflationary expectations as explained in our derivation of the EAPC.

The **hysteresis effect**, that is, the tracking of the actual unemployment rate by the equilibrium rate of unemployment could be represented in a number of ways. In this example, we follow Mitchell (1987) who represented  $U^*$  as a weighted average of the actual unemployment rate and the equilibrium rate in the last period.

The following model shows that the MRU adjusts to the actual unemployment with a lag:

$$(11.14) \quad (U_t^* - U_{t-1}^*) = \mu (U_{t-1} - U_{t-1}^*)$$

This says that the **current MRU**  $U_t^*$  is equal to its value last period  $U_{t-1}^*$  plus some fraction of the gap between the actual and MRU last period.

The value of  $\mu$  measures the sensitivity of MRU to the current state of activity. The higher is  $\mu$ , other things equal, the greater the capacity of aggregate policy to permanently reduce unemployment without ever-accelerating inflation.

### Conflict theory of inflation and inflationary biases

There was a series of articles in the journal *Marxism Today* in 1974, which advanced the notion of inflation being the result of a distributional conflict between workers and capital.

You can view a limited archive of *Marxism Today* since 1977 - [http://www.amielandmelburn.org.uk/collections/mt/index\\_frame.htm](http://www.amielandmelburn.org.uk/collections/mt/index_frame.htm) - which is a very valuable resource.

The article by Devine introduced the notion that inflation was a structural construct. He said:

The phenomenon in need of explanation is not inflation in the abstract but inflation in the world of state monopoly capitalism in the period since the second world war. Two questions arise. First, throughout this period inflation has been chronic in all the major capitalist countries; why has it replaced depression as the principle 'economic' problem confronting the capitalist system as a whole? Second, within this overall framework, why has the rate of inflation varied between countries and at different times? (Devine, 1974: 80).

He argued that the increased bargaining power of workers (that accompanied the long period of full employment in the post Second World War period) and the declining productivity in the

early 1970s imparted a structural bias towards inflation, which manifested in the inflation breakout in the mid-1970s which he says “ended the golden age” (p.80).

In this context, the “relatively full employment” has meant that:

... money wages (earnings) have risen continuously, although at varying rates, for a prolonged period unprecedented in the history of capitalism. This has had far reaching effects on the functioning of the capitalist system. Faced with rising money wages, capitalists have sought to contain the increase in real wages and fend off pressure on profits by increasing prices. (Devine, 1974: 80).

On the industry side, large, oligopolistic firms with price-setting power competed against each other in non-price forms (for example, product quality, etc). The firms, however, were interdependent because market share was sensitive to their pricing strategies. When a firm was faced with nominal wage demands, the management knew that its rivals would be similarly pressured and their competitive positions would not depend on the absolute price level. Rather, a firm could lose market share if they increased prices, while other firms maintained lower prices.

But in an environment where the firms considered that the government would continue to ensure that effective demand was sufficient to maintain full employment there was no reason to assume that rising prices would damage their sales.

As a result, firms had little incentive to resist the wage demands of their workers and strong incentives to protect their market share and profits by passing on the demands in the form of higher prices.

This structural depiction of inflation – embedded in the class dynamics of capital and labour, both of which had increased capacity to set prices and defend their real shares of production – implicates Keynesian-style approaches to full employment.

There was also an international component to the structural theory. It was argued that the Bretton Woods system imparted deflationary forces on economies that were experiencing strong domestic demand growth. As national income rose and imports increased, central banks were obliged to tighten monetary policy to maintain the agreed exchange rate parity and the constraints on monetary growth acted to choke off incompatible nominal claims on the available real income.

However, when the Bretton Woods system of convertible currencies and fixed exchange rates collapsed in 1971 the structural biases towards inflation came to the fore.

Devine (1974: 86) argued that:

... floating exchange rates have been used as an additional weapon available to the state. Given domestic inflation, floating rates provide a degree of flexibility in dealing with the resultant pressure on the external payments position. However, if a float is to be effective in stabilising a payments imbalance it is likely to involve lower real incomes at home. If a reduction in real wages (or their rate of growth) is not acquiesced in there will then be additional pressure for higher money wages and if this cannot be contained the rate of inflation will increase and there will be further depreciation.

The structuralist view also noted that the mid-1970s crisis – which marked the end of the Keynesian period – was not only marked by rising inflation but also by an on-going profit squeeze due to declining productivity and increasing external competition for market share. The

profit squeeze led to firms reducing their rate of investment (which reduced aggregate demand growth), which combined with harsh contractions in monetary and fiscal policy created the stagflation that bedevilled the world in the second half of the 1970s.

The resolution to the **structural bias** proposed by economists depended on their ideological persuasion. On the one hand, those who identified themselves as Keynesians proposed incomes policies as a way of mediating the distributional struggle and rendering nominal income claims compatible with real output.

On the other hand, the emerging Monetarists considered the problem to be an abuse of market power by the trade unions and this motivated demands for policy makers to legislate to reduce the bargaining power of workers. The rising unemployment was also not opposed by capital because it was seen as a vehicle for undermining the capacity of the trade unions to make wage demands.

From the mid-1970s, the combined weight of persistently high unemployment and increased policy attacks on trade unions in many advanced nations reduced the inflation spiral as workers were unable to pursue real wages growth and productivity growth outstripped real wages growth. As a result, there was a substantial redistribution of real income towards profits during this period.

The rise of Thatcherism in the UK exemplified this increasing dominance of the Monetarist view in the 1980s.

### **The Scandinavian Model (SM) of inflation**

This model, originally developed for fixed exchange rates, dichotomises the economy into a competitive sector (C-sector) and a sheltered sector (S-sector). The C-sector produces products, which are traded on world markets, and its prices follow the general movements in world prices. The C-sector serves as the leader in wage settlements. The S-sector does not trade its goods externally.

Under fixed exchange rates, the C-sector maintains price competitiveness if the growth in money wages in its sector is equal to the rate of change in its labour productivity (assumed to be superior to S-sector productivity) plus the growth in prices of foreign goods. Price inflation in the C-sector is equal to the foreign inflation rate if the above rule is applied. The wage norm established in the C-sector spills over into wages growth throughout the economy.

The S-sector inflation rate thus equals the wage norm less its own productivity growth rate. Hence, aggregate price inflation is equal to the world inflation rate plus the difference between the productivity growth rates in the C- and S-sectors weighted by the S-sector share in total output. The domestic inflation rate can be higher than the rate of growth in foreign prices without damaging competitiveness, as long as the rate of C-sector inflation is less than or equal to the world inflation rate.

In equilibrium, nominal labour costs in the C-sector will grow at a rate equal to the norm (the sum of the growth in world prices and the C-sector productivity). Where non-wage costs are positive (taxes, social security and other benefits extracted from the employers), nominal wages would have to grow at a lower rate. The long-run tendency is for nominal wages to absorb the room provided. However in the short-run, labour costs can diverge from the permitted growth path. This disequilibrium must emanate from domestic factors.



The main features of the SM can be summarised as follows:

The domestic currency price of C-sector output is exogenously determined by world market prices and the exchange rate.

The surplus available for distribution between profits and wages in the C-sector is thus determined by the world inflation rate, the exchange rate and the productivity performance of industries in the C-sector.

The wage outcome in the C-sector is spread to the S-sector industries either by design (solidarity) or through competition.

The price of output in the S-sector is determined (usually by a mark-up) by the unit labour costs in that sector. The wage outcome in the C-sector and the productivity performance in the S-sector determine unit labour costs.

An incomes policy would establish wage guidelines, which would set national wages growth according to trends in world prices (adjusted for exchange rate changes) and productivity in the C-sector. This would help to maintain a stable level of profits in the C-sector.

Whether this was an equilibrium level depends on the distribution of factor shares prevailing at the time the guidelines were first applied.

Clearly, the outcomes could be different from those suggested by the model if a short-run adjustment in factor shares was required. Once a normal share of profits was achieved the guidelines could be enforced to maintain this distribution.

A major criticism of the SM as a general theory of inflation is that it ignores the demand side. Uncoordinated collective bargaining and/or significant growth in non-wage components of labour costs may push costs above the permitted path. Where domestic pressures create divergences from the equilibrium path of nominal wage and costs there is some rationale for pursuing a consensus based incomes policy.

An incomes policy, by minimising domestic cost fluctuations faced by the exposed sector, could reduce the possibility of a C-sector profit squeeze, help maintain C-sector competitiveness, and avoid employment losses. Significant contributions to the general cost level and hence prices can originate from the actions by government. Payroll taxation, various government charges and the like may in fact be more detrimental to the exposed sector than increased wage demands from the labour market.

Although the SM was originally developed for fixed exchange rates, it can accommodate flexible exchange rates. Exchange rate movements can compensate for world price changes and local price rises. The domestic price level can be completely insulated from the world inflation rate if the exchange rate continuously appreciates (at a rate equal to the sum of the world inflation rate and C-sector productivity growth).

Similarly, if local price rises occur, a stable domestic inflation rate can still be maintained if a corresponding decrease in C-sector prices occurs. An appreciating exchange rate discounts the foreign price in domestic currency terms.

What about terms of trade changes? Terms of trade changes, which in the SM justify wage rises, also (in practice) stimulate sympathetic exchange rate changes. This combination locks the economy into an uncompetitive bind because of the relative fixity of nominal wages. Unless the

exchange rate depreciates far enough to offset both the price fall and the wage rise, profitability in the C-sector will be squeezed.

It was considered appropriate to ameliorate this problem through an incomes policy. Such a policy could be designed to prevent the destabilising wage movements, which respond to terms of trade improvements. In other words, wage bargaining, consistent with the mechanisms defined by the SM may be detrimental to both the domestic inflation target and the competitiveness of the C-sector, and may need to be supplemented by a formal incomes policy to restore or retain consistency.

## References

- Devine, P. (1974) 'Inflation and Marxist Theory', *Marxism Today*, March, 79–92.
- Friedman, M. (1968) 'The Role of Monetary Policy', *The American Economic Review*, 58(10), March, 1-17
- Hume, D. (1752) 'Of Money' in D. Hume (ed.) *Political Discourses*, Fleming, Edinburgh.
- Keynes, J.M. (1936) *The General Theory of Employment, Interest and Money*, Macmillan, London.
- Keynes, J.M. (1940) *How to Pay for the War: A radical plan for the Chancellor of the Exchequer*, London, Macmillan.
- Lerner, A. (1951) *Economics of Employment*, McGraw-Hill, New York.
- Mitchell, W.F. (1987) 'NAIRU, Structural Imbalance and the Macroequilibrium Unemployment Rate', *Australian Economic Papers*, 26(48), June, 101-118.
- Phillips, A.W. (1958) 'The Relation between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957', *Economica*, 25(100), 283-299.
- Samuelson, P.A. and Solow, R.M. (1960) 'Analytical Aspects of Anti-Inflation Policy', *American Economic Review*, 50(2), Papers and Proceedings of the Seventy-second Annual Meeting of the American Economic Association, May, 177-194.
- Thurow, L.C. (1983) *Dangerous Currents: The state of economics*, Random House, New York.
- US Federal Reserve Bank of St. Louis (2016) "Money Velocity". Accessed on 20 February 2016 from <https://research.stlouisfed.org/fred2/categories/32242>

# Chapter 12: Full Employment Policy

## Chapter Outline

- 12.1 Introduction
  - 12.2 Full Employment as a Policy Goal
  - 12.3 Unemployment Buffer Stocks and Price Stability
  - 12.4 Employment Buffer Stocks and Price Stability
    - The JG wage
    - The JG as an automatic stabiliser
    - Inflation control and the JG
    - Open economy impacts
    - Employment buffer stocks and responsible fiscal design
    - Conclusion
  - 12.5 Alternative Policies for Promotion of Employment
    - Behaviouralist, structuralist, and Keynesian approaches
    - Private sector incentives
    - Direct job creation by government
- Appendix: Advanced Material
- Measuring the costs of unemployment buffer stocks
  - Buffer stocks in agriculture
  - A plausible adjustment path
  - Would the NAIBER be higher than the NAIRU?

## **Learning Objectives**

1. Recognise that full employment should be the primary macroeconomic policy goal in a civilized society.
2. Acknowledge that current policy settings in most developed economies are based on the inflation rate as the primary policy target, which is addressed by an unemployment buffer stock.
3. Capacity to analyse the economic consequences of implementing an employment buffer stock (Job Guarantee).
4. Assess other policies which promote employment.

## 12.1 Introduction

In Chapter 11, we discussed how distributional conflict between the claimants of real income could trigger inflation if the competing nominal claims (wages, profits) exceeded the actual amount of nominal income produced in each period.

We saw how this conflict could be triggered by rising real wage aspirations from workers, rising profit rate aspirations from price setters (firms), and exogenous squeezes on available national income arising from, for example, an imported raw material price rise.

The underlying dynamics of the capitalist system are driven by the target rates of profit determined by firms. In this context, workers may create unemployment by seeking real shares of real national income that undermine the capacity of firms to achieve the target rate of profit. Unemployment would rise from a reduction in effective demand that follows firms' withdrawal of investment spending in response to a squeeze on the rate of profit.

An inflationary spiral arising from demand-pull forces or cost-push forces requires certain aggregate demand conditions to be maintained if that spiral is to continue.

As we saw in Chapter 11, this observation means that the concept of supply-side inflation blurs with the concept of demand-pull inflation, although their originating forces might be quite different.

In this Chapter, we compare the two broad ways in which price stability may be achieved. We construct the discussion in terms of a comparison between two types of buffer stocks both of which are created by government policy aimed at reducing aggregate demand pressures that are fuelling the inflationary spiral.

The **two buffer stocks** that we will compare and contrast are:

- **Unemployment Buffer Stocks:** Under a Natural Rate of Unemployment (NRU) also referred to a Non-Accelerating Inflation Rate of Unemployment (NAIRU) regime, inflation is controlled using tight monetary and fiscal policy, which leads to a buffer stock of unemployment. This is a very costly and unreliable target for policy makers to pursue as a means for inflation proofing that is the achievement of low and stable inflation.
- **Employment Buffer Stocks:** The national government exploits the fiscal power embodied in a fiat-currency issuing system to introduce full employment based on an employment buffer stock approach. The Job Guarantee (JG) model is an example of an employment buffer stock policy approach.

The two buffer stock approaches to inflation control both introduce so-called **inflation anchors**. In the NAIRU case, the anchor is unemployment, which serves to discipline the labour market and prevent inflationary wage demands from being pursued. Under a Job Guarantee, the inflation anchor is provided in the form of a fixed wage employment guarantee provided by the government.

The NAIRU approach to price stabilisation is based on government spending being subject to a **quantity rule**. This means that the government plans for a quantity of dollars to be spent at prevailing market prices to prosecute its socio-economic program. Spending over-runs are usually met with cut backs in an attempt to meet the fiscal estimates.

Conversely, the employment buffer stock approach represents a shift from spending based on a quantity rule to spending being underpinned by a **price rule**. Accordingly, the government offers a fixed wage (that is a price) to anyone willing and able to work, and lets market forces determine the total quantity of government spending that would be required to satisfy the demand for public sector jobs under the Job Guarantee.

We will explain how spending on a price rule provides the government with a superior inflation control mechanism. We will see that when the private sector is inflating, a tightening of fiscal and/or monetary policy can shift workers into a fixed-wage Job Guarantee sector to achieve inflation stability without causing costly unemployment. This program both anchors the general price level to the price of employed labour of this (currently unemployed) buffer, and can produce useful output with positive supply side effects.

In this Chapter, we first elaborate on why full employment should be the key macroeconomic policy goal. Second we outline and contrast the two buffer stock schemes, which are designed to control inflation. Finally the Chapter briefly considers other employment generation schemes.

## 12.2 Full Employment as a Policy Goal

In our discussion of Public Purpose in Chapter 1, we noted that in a modern capitalist economy access to employment is required for full participation in society. Employment, especially in formal sector jobs, not only integrates individuals into networks linked to the workplace, but also into the social and political environment more generally.

On the other hand, it has been well documented that sustained unemployment imposes significant economic, personal and social costs that include:

- Loss of current national output and income;
- Social exclusion and the loss of freedom;
- Skill loss;
- Psychological harm;
- Ill health and reduced life expectancy;
- Loss of motivation;
- The undermining of human relations and family life;
- Racial and gender inequality; and
- Loss of social values and responsibility.

Thus, macroeconomic policy that uses unemployment to promote macro stability not only forces those who are already disadvantaged to bear most of the costs, but it also impedes the development of social cohesion. Joblessness is usually concentrated among groups that suffer other disadvantages: racial and ethnic minorities, immigrants, younger and older individuals, women (especially female heads of households with children), people with disabilities, and those with lower educational attainment. Lack of employment is highly correlated with poverty and with a higher degree of social isolation.

The **United Nations Universal Declaration of Human Rights** includes the right to work, not only because it is important in its own right, but also because many of the other economic and

social entitlements proclaimed to be human rights cannot be secured without paying jobs. Also **Full, productive and decent employment** is one of the **Millennium Development goals**. Amartya Sen (1997) supports the right to work because the economic and social costs of unemployment are staggering with far-reaching consequences beyond the single dimension of a loss of income.

Markets are not necessarily good at securing the economic and social entitlements proclaimed to be human rights in the Universal Declaration. This is why extra market policy has been used to safeguard a variety of human rights. Unemployment and poverty are generally seen as the necessary cost of maintaining macroeconomic stability, especially price and exchange rate stability. This raises important questions: should a nation fight inflation by keeping a portion of its population unemployed and impoverished? Are there other tools available to achieve these ends? In particular, should policymakers accept some inflation and currency depreciation in order to eliminate unemployment and poverty?

There are strong ethical arguments against using poverty and unemployment as the primary policy tools to achieve price and exchange rate stability - especially given that costs of poverty and unemployment are not shared equally. And even if price and currency stability are highly desired, it is doubtful that a case can be made for their status as a human right on par with the right to work.

Hence, at the very least, safeguards are required to protect the minority which suffers large concentrated costs in the form of unemployment, as a consequence of a policy that leads to the benefits of lower inflation accruing to society as a whole.

Unemployment as a stabilisation tool fails on several accounts: it violates various human rights including the right to employment, and sacrifices economic performance by generating redundant human resources. Indeed, with some notable exceptions, those countries with the highest rates of underutilised labour resources tend to be nations with high poverty rates.

Only government can guarantee the right to a job because markets have not, and cannot, operate at anything approaching true, full employment on a consistent basis without direct job creation on a large scale. Finally, only the government can offer an infinitely elastic demand for labour (offering to hire all who cannot otherwise find employment) because it does not need to take account of narrow market efficiency concerns.

Private firms only hire the quantity of labour needed to produce the level of output that is expected to be sold at a profitable price. Government can take a broader view to include promotion of the public interest, including the right to work. For these reasons, government should and must play a role in providing jobs to achieve social justice. A JG program can secure the right to work, but with minimal undesired impacts on wages, prices, government fiscal policy, and the value of the currency.

Forstater (2006) has argued that it is difficult to conceive of a policy that secures a greater range of social and economic rights than one directed to the achievement of full employment. His “fundamental welfare theorem of political economics” says “there is no single policy that carries with it more potential benefits than true full employment, or a guaranteed job for everyone ready and willing to work” (Forstater 2006: Slide 2).

In addition to income, employment also provides useful production as well as recognition for doing something worthwhile. While economists usually focus only on the economic multiplier,

there are also **social multipliers** associated with job creation - the benefits that include decreased crime and drug use; enhanced family and community cohesion; strengthened security, education, and healthcare; protection for the disadvantaged; environmental protection; improved local and state government budgets; greater equality of distribution of consumption, income, wealth, and power; induced investment in poor communities; and promotion of social and political stability. Only the introduction of a safety net provided by the introduction of a Job Guarantee can ensure protection of the right to employment.

While economic growth and development are desirable, they do not ensure either **full or decent employment**. Alternative strategies for promoting full and decent employment will be explored in Section 12.5. Certainly it is necessary to attack problems of unemployment, underemployment, and insufficient pay using a variety of programs and policies. These should include both private and public initiatives. However, it will be argued that the private sector alone will not be able to provide for full, productive, and decent employment for all, even with substantial support by government for job creation in the private sector.

Hence, a **Job Guarantee** (JG) will be required - neither as an emergency policy nor as a substitute for private employment, but as a permanent complement to private sector employment. A direct job creation program can provide employment at a basic wage for those who cannot otherwise find work. No other program can guarantee access to jobs at decent wages.

### 12.3 Unemployment Buffer Stocks and Price Stability

There have been two striking developments in economics over the last forty years. First, a major theoretical revolution has occurred in macroeconomics (from Keynesianism to Monetarism and beyond) since the mid-1970s. Second, unemployment rates have persisted at the highest known levels in the post-World War II period and during the GFC rose even higher.

Prior to Keynes's General Theory, unemployment at the aggregate level was seen by many orthodox economists as a temporary deviation from equilibrium which was due to labour market frictions or other market disruptions. Keynes changed the discourse to one that blamed aggregate unemployment on insufficient aggregate demand. This led to the belief that Keynesian demand management policy was the proper response in the post-war period. Unemployment rates were usually below 2 per cent throughout this period.

However, in the early 1970s, the Phillips Curve trade-off appeared to break down with countries experiencing stagflation. In time, most of the mainstream rejected demand management and returned to the older pre-Keynesian belief that some level of aggregate unemployment is: (a) temporary and due to shocks; (b) optimal because it is voluntary; and/or (c) the necessary cost of promoting stability. Thus the concept of full employment as a genuine policy goal was thus abandoned with the introduction of the **natural rate of unemployment hypothesis**, which has become a central plank of current mainstream thinking.

It asserts that there is only one unemployment rate consistent with stable inflation. Under this hypothesis, there is no discretionary role for aggregate demand management and only microeconomic changes can reduce the natural rate of unemployment. Thus, the policy debate became increasingly concentrated on deregulation, privatisation, and reductions in the provisions of the Welfare State within an environment of tight monetary and fiscal regimes.



The almost exclusive central bank focus on maintaining price stability on the back of an overwhelming faith in the NAIRU ideology has marked the final stages of the abandonment of earlier full employment policies.

Under inflation targeting (or inflation-first) monetary regimes, central banks shifted their policy emphasis. They now conduct monetary policy to meet an inflation target and arguably, have abandoned any obligations they have to support a policy environment which achieves and maintains full employment.

Unemployment since the mid-1970s has mostly persisted at high levels although in some economies, low quality, casualised work has emerged in the face of persistently deficient demand for labour hours by employers. In this case, underemployment has replaced some unemployment.

As we saw in Chapter 11, underemployment acts in a similar way to unemployment as a disciplining force on workers' wage aspirations and demands. It weakens the capacity of workers to secure nominal wages growth.

Thus, labour underutilisation (i.e. short-term unemployment and underemployment) temporarily balances the conflicting demands of labour and capital by disciplining the aspirations of labour so that they are compatible with the profitability requirements of capital.

Similarly, low product market demand, the analogue of high unemployment as workers' incomes fall, suppresses the ability of firms to increase prices to protect raise real margins.

Thus by inducing labour slack into the economy, inflation targeting supported by passive fiscal policy leaning towards austerity, has created what Karl Marx called a **reserve army of the unemployed** and this reduces the chances of an inflationary spiral emerging from the wage bargaining process.

We have seen significant shifts in the distribution of national income towards profits since the mid-1980s as real wages growth has lagged behind productivity growth. This redistribution of national income has overridden the previous outcomes when strong trade unions met on more equal terms with employer groups to determine a distribution of national income that would be acceptable to both sides of the bargaining process.

But with trade unions weaker as a result of shifting industry composition towards services, smaller public sectors and anti-union legislation, the likelihood of explosive wage-price spirals has been significantly reduced in developed economies, including the USA, UK and Australia. As a consequence, the use of unemployment as a tool to suppress price pressures has, based on the OECD experience since the 1990s, been successful.

The empirical evidence is clear that most OECD economies have not provided enough jobs since the mid-1970s and the conduct of monetary policy has contributed to the malaise. Central banks around the world have forced the unemployed to engage in an involuntary fight against inflation and the fiscal authorities in many cases have further worsened the situation with complementary austerity measures.

These costs are very large and have long-term consequences. In terms of the goals of macroeconomic policy they also present a major conflict. As we have learned, a central idea in economics whether it be microeconomics or macroeconomics is efficiency – getting the best out

of what is available. We have discussed the difficulties that economists have in defining such a concept and its ideological dimensions.

But economists could put aside their difference and agree that at the macroeconomic level, the efficiency frontier should be defined in terms of full employment. The major debate, which we covered in Chapter 11, concerned how we might define full employment. But it is a fact that full employment should be a central focus of macroeconomic theory.

Certainly mass unemployment involving hundreds of thousands or millions of workers not producing any output or national income would violate our notion of macroeconomic efficiency under any reasonable definition of that term.

Further, persistently high unemployment not only undermines the current welfare of those affected and slows down the growth rate in the economy below its potential, but also reduces the medium- to longer-term capacity of the economy. The erosion of skills and lack of investment in new capacity means that future productivity growth is likely to be lower than if the economy was maintained at higher rates of activity.

The key question to ask advocates of the unemployment buffer stock approach to inflation control is whether the economy, once deflated by restrictive aggregate demand management, can be restarted without inflation.

If the underlying causes of the inflation are not addressed, a demand expansion will merely reignite the tensions over the distribution of income between wage earners and profit recipients and a wage-price outbreak is likely. As a basis for policy the NAIRU approach has major limitations, because it addresses the symptoms and not the causes of inflation, and as a consequence, provides no firm basis for sustained full employment and price stability. In short, its success as an inflation anchor requires a chronic pool of high unemployment.

The disciplining power of unemployment requires that the unemployed constitute a threat to those still in work, so that they will moderate their wage demands. However, over time, the threat from this unemployment pool starts to wane as the unemployed endure skill losses, while suffering lengthening periods of unemployment, and firms introduce new technologies and processes. This is referred to hysteresis (see Chapter 11).

In this case, the so-called NAIRU has to be pushed higher and higher by contractionary fiscal and monetary policy for the same degree of threat to be maintained.

On any reasonable grounds, this approach to price stability is very costly and ultimately, unworkable in a modern economy. High and sustained levels of unemployment undermine the social and political stability of a nation, which creates unintended costs that go far beyond those that are itemised above.

Research has been undertaken to calculate the sacrifice ratio associated with the implementation of the unemployment buffer stock policy. The ratio is defined as the ratio of the accumulated loss of output during a disinflation episode expressed as a percentage of initial output divided by the overall reduction in the inflation rate. Output loss can be temporary with the potential (long term) output level being restored.

On the other hand, the concept of **persistence** means that actual output remains below its potential level after the disinflation period has finished. The longer this output gap exists, the longer is the persistence.

In this context, **hysteresis** refers to the permanent losses of potential output that arise as a consequence of the disinflation policy. Thus the growth of potential output is permanently lowered, which in turn limits the long-term growth of actual output.

A more detailed treatment of the sacrifice ratio is provided in the Appendix of this Chapter.

## 12.4 Employment Buffer Stocks and Price Stability

Given the importance placed on access to employment for those of working age who want a job, a better alternative to an unemployment buffer stock to achieve price stability would be to utilise an employment buffer stock, as long as price stability was not compromised.

In this section, we outline an employment program for these unemployed persons as an activity floor in the real sector, which both anchors the general price level to the price of this (currently unemployed) buffer stock of employed labour but also can produce useful output with positive supply side effects.

Recall that the MMT approach argues that the imposition of taxes by the currency-issuing government generates a demand for the currency. The currency's value is determined by what must be done to obtain it. The currency will be worth the amount of labour it can buy on the margin, which is the wage paid in the JG program. The wage and benefit package in the JG program sets a standard for what must be done to earn currency-denominated income. To be sure, the anchor is not tight. Some people earn a level of income, either from the private sector or from the government, which is disproportionate to their level of endeavour. If everyone could get income while doing nothing (**if money grew on trees!**) then the currency's value would approach zero. However, in the real world, the government's currency does not grow on trees and most people have to do something to get it. For that reason, at the margin, currency is valuable.

Between 1945 and the mid-1970s, Western governments realised that with deficit spending supplementing private demand, they could ensure that all workers who wanted to work could find jobs. Although private employment growth was relatively strong during this period, governments were important employers in their own right, and also maintained a buffer of jobs for the least skilled workers; for example, in the major utilities, the railways, local public services and major infrastructure functions of government. By absorbing workers who lost jobs when private investment declined, governments acted as an economic safety valve.

British economist Paul Ormerod (1994: 203) noted that the economies that avoided high unemployment in the 1970s maintained a:

... sector of the economy which effectively functions as an employer of last resort, which absorbs the shocks which occur from time to time, and more generally makes employment available to the less skilled, the less qualified.

He concluded that societies with a high degree of social cohesion (such as Austria, Japan and Norway) were willing to ensure that everyone had access to paid employment opportunities.

The **employment buffer stock** approach – which is referred to in the literature as the **Job Guarantee (JG)** – defines a policy framework where the government operates a buffer stock of jobs to absorb workers who are unable to find employment in the private sector.

Analogous to the central bank's function of lender of the last resort, the JG functions as a buffer which employs all job seekers who have not obtained regular public or private sector jobs, at an acceptable minimum wage. In this sense, the government acts as an employer of the last resort. The jobs are available on demand.

While it is easy to characterise the JG as purely a public sector job creation strategy, it is important to appreciate that it is actually a macroeconomic policy framework designed to deliver full employment and price stability based on the principle of buffer stocks.

Under a JG, the government thus provides an unconditional, open-ended job offer at a given wage to anyone who desires to work. Instead of a person becoming unemployed when aggregate demand falls below the level required to maintain full employment, the person would enter the JG workforce.

The JG pool expands (declines) when private sector activity declines (expands). The JG thus fulfils an absorption function, which minimises the costs associated with the flux of economic activity when aggregate demand fluctuates.

In the event of a decline in aggregate demand, total demand for non-JG labour workers declines according to the employment requirements function we defined in Chapter 8. In this situation, the workers who were displaced from their jobs would have an option – accept a JG position or wait for conditions to improve in the non-JG economy.

It is clear that the choice facing workers will be influenced by several factors. First, the government may offer workers the choice between the JG wage and the unemployment benefit, the latter being lower. Second, some workers, especially those in higher-skilled positions, may receive redundancy payments and use these to support themselves through the spell of unemployment. Economists call this response – wait unemployment. Some workers may feel that accepting a low-skill JG job would disadvantage them professionally and thus wait for circumstances to improve.

We assume for the moment that the JG policy does not offer an unemployment benefit and that most displaced workers will prefer a JG position over wait unemployment. These assumptions serve to simplify the analysis and relaxing them does not alter the basic dynamics of the system.

When private economic activity picks up, workers would be bid out of the JG pool by employers and the buffer stock of jobs would contract.

Finally, the JG program helps to stabilise aggregate wages, since no worker's wage can fall below the JG wage. For most of the working population, the wage is usually the most important source of income. Stable aggregate wages in turn help to stabilise consumption. For that reason, this targeted approach to sustaining full employment helps to also stabilise aggregate demand, output, and prices.

## **The JG wage**

Why would workers accept these bids? The buffer stock employees would be paid a minimum wage, which would define the level of income necessary for a full-time worker to enjoy an adequate social and material existence.

The nation's workforce would always remain fully employed, with only the mix between private and public sector employment fluctuating as it responds to the spending decisions of the private sector. Since the JG wage is open to everyone, it will effectively be the national minimum wage.

While it is preferable to avoid disturbing the private sector wage structure when the JG is introduced, a case can be made to offer the JG wage at a level higher than the existing private minimum if it is thought that productivity is too low in the economy.

This is particularly relevant in developing economies where many market-based jobs pay wages that are below the poverty line and provide no incentives for employers to invest in more productive capital, or for workers to invest in human capital.

The minimum wage should not be determined by the capacity to pay of the private sector. It should be an expression of the aspiration of the society in terms of the lowest acceptable standard of living. Any private operators who cannot afford to pay the minimum should exit the economy.

The Government would supplement JG earnings with a wide range of social wage expenditures, including adequate levels of public education, health, child care, and access to legal aid.

Further, the JG policy does not replace the conventional use of fiscal policy to achieve social and economic outcomes. Typically, the JG would be accompanied by higher levels of public sector spending on public goods and infrastructure. These supplements would be in addition to the scheme but not essential for the scheme to function effectively.

## **The JG as an automatic stabiliser**

The JG wage thus defines the wage floor for the economy and serves as an automatic stabiliser, similar to the tax system.

Recall that automatic stabilisation refers to the components of the government fiscal outlays and receipts, which rise and fall as the economic cycle fluctuates without there being any explicit change in government spending or tax settings.

They operate to stabilise the economic cycle providing a floor following a fall in aggregate demand during an economic downturn and a ceiling for aggregate demand as the economy grows. At full employment, the automatic stabiliser component of aggregate demand is zero.

Thus, when the economy is in decline, tax revenue falls and welfare payments rise, which expands the fiscal deficit of the government automatically. The introduction of the JG would have the same counter-cyclical impact. When the economy was faltering, the spending associated with the JG would rise and vice-versa when times were good.

In this regard, the JG is a superior (more powerful) automatic stabiliser than a system of unemployment benefits (under the unemployment buffer stock option) because aggregate demand slumps less and therefore the positive impact on real output is greater than would be the case if the government merely paid unemployment benefits. Further the operation of the JG sustains full employment.

Automatic stabilisers have the desirable characteristic of providing immediate, counter-cyclical spending injections (or withdrawals) when private activity fluctuates. They avoid the so-called policy lags, which relate to the time delays in the government identifying that a significant shift in private demand has occurred, designing a policy response to that shift, providing appropriate legislation to support an intervention, and then executing the intervention.

In some cases, the time delays can result in the major part of a new policy intervention arriving too late and working to destabilise the cycle. For example, if by the time the government has designed and implemented a new discretionary spending injection, the private sector has already resumed normal spending growth, then the impulse of government spending might lead to the economy overheating. This economic destabilisation would not occur under a JG. Workers, who have become unemployed, following a fall in aggregate demand, can readily identify themselves to the appropriate government agency and secure a JG job.

The fixed wage offer that defines the JG policy also serves to stabilise the growth rate in money wages in the economy and thus provides a nominal anchor against inflation.

A JG program can be a complement to any of the other approaches examined below. By design, it is a complement to private sector employment and to any other active labour market policies, demand fine-tuning policies, and welfare or other social safety nets. A universal JG program, which employs anyone who is ready and willing to work, is the only type of program that can ensure that the human right of employment is continuously met. If the program wage is a living wage, it also helps to ensure that other human rights are met, by providing sufficient income. A properly designed program will not only produce socially useful goods and services, but it will also promote feelings of self-worth and accomplishment among program participants. Finally, JG generates full employment, and macroeconomic stability but with the least disruption to markets.

The notion of a JG has a long history, and there are many examples of such programs through history and across the globe - although usually on a small scale or temporary basis.

### **Inflation control and the JG**

While introducing a public sector job creation capacity to the economy, the JG is better thought of as a macroeconomic policy framework designed to ensure that full employment and price stability is maintained over the private sector economic cycle.

What are the mechanics of inflation control under a JG? In Chapter 11, we examined the way in which incompatible claims over the available real income could cause wage-price pressures to escalate into an inflationary episode as the claimants (labour and capital) attempted to defend their real income shares.

In an unemployment buffer stock system, unemployment is used to discipline wage demands by workers and to soften the product market to discourage a profit-margin push by firms as a means of curbing wage-price pressures and maintaining stable inflation.

We define the **Buffer Employment Ratio (BER)** as:

$$(12.1) \quad BER = JGE/E$$

where  $JGE$  is total employment in the Job Guarantee buffer stock and  $E$  is total employment in the economy. **The BER rises when the JG pool expands and falls when the JG pool contracts.**

The JG approach stands in contradistinction to the NAIRU approach because instead of manipulating the employment rate by creating unemployment when wage-price pressures develop, the government manipulates the BER.

When the level of private sector activity and the distributional conflict is such that wage-price pressures form as the precursor to an inflationary episode, the government manipulates fiscal and monetary policy settings (preferably fiscal policy) to reduce the level of private sector demand.

Labour is then transferred from the inflating private sector to the **fixed wage JG sector** and the BER rises. This will eventually ease the inflationary pressures arising from the wage-price conflict.

There can be no inflationary pressures arising directly from a policy where the government offers a fixed wage to any labour that is unwanted by other employers. The JG involves the government **buying labour off the bottom**, in the sense that employment at the minimum wages does not impose pressure on the market-sector wage structure. By definition, the unemployed have no market price because there is no market demand for their services.

By not competing with the private market, the JG would avoid the inflationary tendencies of past Keynesian policies, which attempted to maintain full capacity utilisation by ‘hiring off the top’ (that is, making purchases at market prices and competing for resources with all other sources of spending in the economy).

The BER conditions the overall rate of wage demands. When the BER is high, real wage demands will be correspondingly lower and the capacity of firms to push profit margins up is reduced, due to weaker product demand.

So instead of a buffer stock of unemployed being used to discipline the distributional struggle, the JG policy achieves this via compositional shifts in employment through transfers in and out of the JG pool.

Importantly, the JG can also deal with a supply-shock (such as a rise in the price of a key non-labour raw material) that generates incompatible claims on national income that ultimately cause inflation.

**The NAIRU defines the unemployment buffer stock associated with stable inflation. In a JG setting, we define the Non-Accelerating Inflation Buffer Employment Ratio (NAIBER) as the BER that achieves stable inflation following the redistribution of workers from the inflating private sector to the fixed price JG sector.**

**The NAIBER is a full employment steady state JG level, which is dependent on a range of factors including the historical path the economy has taken.**

An aim of government is to minimise the NAIBER so that higher levels of non-JG employment can be sustained with stable inflation. Initiatives that may reduce the value of the NAIBER include public education to stimulate skill development and engender high productivity growth; institutionalised wage setting processes where productivity growth is shared equitably across all income claimants, and restrictions on anti-competitive cartels that may add pressures for profit margin push.

However, while central banks and treasuries devote a lot of resources in trying to estimate the NAIRU, we consider it would not be worth trying to estimate or target a particular NAIBER. The point is that the aim of policy is to fully employ labour while maintaining price stability.

### Open economy impacts

The JG requires a flexible exchange rate to be effective. A once-off increase in import spending is likely to occur when the policy is introduced because the JG workers will have higher disposable incomes.

In most nations, the impact would be modest. We would expect any modest depreciation in the exchange rate to have low exchange rate pass through effects on the price level via higher import prices and provide a modest boost to net exports and local employment, as explained in Chapter 16.

### Employment buffer stocks and responsible fiscal design

In an open economy, the level of economic activity (output) determined by private domestic spending (consumption plus investment) and net external spending (exports minus imports) might not be sufficient to generate full employment. Further, if one or more of those components of spending declines, then activity will decline.

In Chapter 7, we learned that a **spending gap** is defined as the spending required to create demand sufficient to elicit an output level, which at current levels of productivity, will provide enough jobs (measured in working hours) for all the workers who desire to work.

A zero spending gap occurs when there is full employment. We assume that there is no capacity-constrained unemployment where the level of capital stock is unable to support enough jobs to satisfy the available labour supply at existing productivity levels.

The role of aggregate government policy interventions is to ensure there is no spending gap. If we assume that monetary policy changes are relatively ineffective as a counter-stabilisation policy tool, then if private spending declines from a given position of full employment, the only way that the spending gap can be filled is via a fiscal stimulus – directly through government spending and/or indirectly, via a tax cut, which will increase private disposable income and stimulate subsequent private spending.

To recapitulate: the essence of the income-expenditure framework developed earlier, the sources of expenditure, which sum to aggregate spending (demand), are:

- Household consumption ( $C$ )
- Private Investment ( $I$ )
- Government spending ( $G$ )
- Export revenue ( $X$ )

The income payments to resource owners involved in the production of output generated by these spending flows can be used in the following ways:

- Household consumption ( $C$ )
- Household saving ( $S$ )



- Taxation payments ( $T$ )
- Import spending ( $M$ )

Clearly, the sources of income have to equal the uses (as a convention of the National Accounts). As we learned in Chapter 5 in sectoral accounting, this allows us to write the two sides of income generation like this:

$$(12.2) \quad C + I + G + X = C + S + T + M$$

Given  $C$  cancels out we know that:

$$(12.3) \quad I + G + X = S + T + M$$

The left-hand side ( $I + G + X$ ) are called **injections** – because they inject new demand into the economy whereas the right-hand side ( $S + T + M$ ) are **leakages** because they drain aggregate demand.

The left-hand side of this equation is always brought into equality with the right-hand side via national income adjustments (that is, variations in the level of aggregate activity brought about by spending variations).

**The way national income adjustments impact on the injections and leakages in the income-expenditure system is one of the first principles of macroeconomics.**

So if for example, Private Investment increases (with  $G$  and  $X$  constant), aggregate demand rises and firms react by increasing output to meet the new orders.

In doing so they will increase employment and pay out more in wages overall. The increased income is then used by workers to consume more, but also to increase saving ( $S$ ), pay more tax ( $T$ ) at current tax rates, and increase imports ( $M$ ).

The economy will stop expanding in response to this stimulus once the change in investment is equal to the sum of the changes in  $S$ ,  $T$  and  $M$ . We identified this dynamic response and subsequent resolution with the **expenditure multiplier**, which results in the movement to a new expenditure-income equilibrium after an exogenous spending injection occurs (Chapter 7).

**The economy is thus in a steady-state (that is, at rest or in equilibrium) when the sum of the injections equals the sum of the leakages.** Whenever this relationship is disturbed (by a change in the level of injections, however sourced), national income adjusts and brings the income-sensitive leakages into line with the new level of injections. At that point the system is at rest.

Three points should be reiterated.

First, this position of rest does not necessarily and will rarely coincide with full employment. There is no automatic tendency in the capitalist monetary system for the economy to sustain or achieve full employment.

The system will adjust to dramatically lower levels of injections and come to rest even if there are high unemployment levels. We now appreciate that economies can settle at very high levels of unemployment and stay there unless total injections increase. Typically, if private spending is depressed then that intervention will have to come from a fiscal policy stimulus.

Second, when an economy is at rest and there is high unemployment, there must be a spending gap given that mass unemployment is the result of deficient demand.

Accordingly, if there is no dynamic which would lead to an increase in private (or non-government) spending, then the only way the economy will increase its level of activity is if there is increased net government spending.

This means that the increasing government spending ( $G$ ) has to more than offset the increased drain (leakage) coming from taxation revenue ( $T$ ). That is, a fiscal deficit is needed if there is a non-government spending gap.

Third, this doesn't mean that a fiscal deficit is always required. In some circumstances, a surplus might be the appropriate fiscal stance.

If the non-government decisions taken together (consumption and saving decisions by households, investment decisions by production firms and the outcomes of the external sector) indicate a desire to net save which might be written as:

$$(12.4) \quad I + X < S + M$$

then the only way the level of activity corresponding to these levels of leakages and injections can be maintained on an on-going basis (whatever the rate of unemployment) is if  $G > T$ . That is a fiscal deficit is required on a continuous basis to sustain a given level of activity.

In this case, **a fiscal deficit finances the desire by the non-government sector to save overall** by maintaining sufficient demand to produce a level of income which will generate that level of net saving.

Responsible fiscal policy thus requires the following two conditions to be fulfilled:

**First, the discretionary fiscal position (deficit or surplus) must fill the gap between the savings minus investment minus the gap between exports minus imports.**

In notation this is given as:

$$(12.5) \quad (G - T) = (S - I) - (X - M)$$

So for income to be stable, the fiscal deficit will equal the excess of saving over investment (which drains domestic demand) minus the excess of exports over imports (which adds to demand).

If the right-hand side of the equation:  $(S - I) - (X - M)$  – is in surplus overall – that is, the non-government sector is saving overall, then the only way the level of national income can remain stable is if the fiscal deficit offsets that surplus.

A surplus on the right-hand side can arise from  $(S - I) > (X - M)$  (that is, the private domestic sector net saving being more than the net export surplus) or it could be associated with a net exports deficit (draining demand and adding foreign savings) being greater than the private domestic sector deficit (investment greater than saving) which adds to demand.

**Second, most importantly, a stable level of national income doesn't necessarily define a state of full employment.**

We can define a full employment level of national income as that level which is generated when all resources are fully utilised according to the preferences of workers and owners of land and capital etc.

Given that  $S$ ,  $T$  and  $M$  are all positively related to the level of national income, there is a unique level of each of these flows that is defined at full employment. Changes in behaviour (for

example, an increased desire to save per dollar earned) will change that unique level, but for given behavioural preferences and parameters we can define levels of each.

We denote  $S(Y_f)$ ,  $M(Y_f)$  as the flows corresponding to full employment income ( $Y_f$ ). We also consider investment to be sensitive to national income (this is outlined in the so-called accelerator theory) such that higher levels of output require more capital equipment for a given technology. So  $I(Y_f)$  might be defined as the full employment flow of investment. We consider export spending to be determined by the level of world income.

Accordingly, a **full-employment fiscal deficit condition** for stable national income is written as:

$$(12.6) \quad (G - T) = S(Y_f) + M(Y_f) - I(Y_f) - X$$

The sum of the terms  $S(Y_f)$  and  $M(Y_f)$  represent drains on aggregate demand when the economy is at full employment and the sum of the terms  $I(Y_f)$  and  $X$  represents spending injections at full employment.

If the drains outweigh the injections then for national income to remain stable, there has to be a fiscal deficit ( $G - T$ ) sufficient to offset that gap in aggregate demand.

If the fiscal deficit is not sufficient, then national income will fall and full employment will not be achieved. If the government tries to expand the fiscal deficit beyond the full employment limit, ( $G - T(Y_f)$ ), then nominal spending will outstrip the capacity of the economy to respond by increasing real output, and while income will rise, it will be all due to price effects (that is, inflation would occur).

In this sense, MMT specifies a strict discipline on fiscal policy. If the goal is full employment and price stability then the full-employment fiscal deficit condition has to be met.

The question then arises: how do employment buffer stocks relate to this condition?

We used the term **loose full employment** in relation to the JG because the employment generated is at minimum wages. The government expands the JG pool by **purchasing off the bottom** of the labour market.

In that context, the automatic stabiliser response associated with the conduct of the JG represents the minimum fiscal shift that is required to maintain employment at its previous level in the face of a falling level of private demand.

The maintenance of the level of employment, however, is accomplished by increasing the BER. That is, more workers are working on minimum wage and less on market wages when the JG pool expands.

The government may decide that it has non-inflationary room to then expand non-JG employment via direct job creation in the career section of the public sector or by a general fiscal stimulus designed to increase private sector employment.

In this case, the actual deficit spending that will satisfy the full employment fiscal deficit condition varies according to the proportion of the deficit that is associated with JG employment.

## Conclusion

There are many microeconomic factors that are relevant to a full understanding of how a Job Guarantee would work in practice. Questions relating to the type of jobs, the levels of

government involved in funding and operations, the relationship with the existing income support system, the integration of training pathways into the policy, the role of trade unions, the choices available to workers for fractional employment, the capacity of the government to sack workers and more.

While these are important factors, which have been dealt with in the literature, they lie outside of our macroeconomic focus in this textbook. More information can be found in the references at the end of this Chapter.

## **12.5 Alternative Policies for Promotion of Employment**

### **Behaviouralist, structuralist, and Keynesian approaches**

There are a range of strategies which have been adopted to address the problem of joblessness, of which the most important are behaviourist (problems with the individuals who are unemployed), structuralist (for example, skills mismatch), and job shortage. Kaboub (2008) provides an historical overview of attitudes of economists toward the unemployed, from Petty (1662: 160) (the unemployed “ought neither to be starved, nor hanged, nor given away” as they represent a resource that could be used in public employment to enrich the nation) to Beveridge (1945:10) (who wanted full employment, defined as “having always more vacant jobs than unemployed”) and to lesser known advocates of Job Guarantee-type schemes such as Pierson and Wernette. He also surveys employment strategies adopted during the US New Deal, in the Swedish post-war model, in India, in Argentina, and recently by France in a pilot program to create jobs for laid off workers.

Public attitudes and policy have generally emphasised behavioural and structural problems. This leads to policies that try to motivate and train the unemployed, together with the promotion of greater flexibility (such as wage flexibility) that would reduce labour market frictions. However, if the problem is a job shortage, all that these policies can do is to redistribute unemployment among this unfortunate group, who are blamed for their joblessness. In an expansion, those job seekers who are not hired generally do have the characteristics identified with the behaviourist and structuralist arguments (since employers hire those with the most desirable characteristics first), hence concealing the true problem - a chronic job shortage.

Hyman Minsky (1986) and in earlier work always argued that public policy that favours education and training over job creation puts the cart before the horse and is unlikely to succeed. While Minsky is best known for his work on financial fragility, he also wrote many articles and chapters on employment policy, and was a consistent advocate for a Job Guarantee.

First, it lays the blame on the unemployed, which can be demoralising and can validate public perceptions regarding the undesirable characteristics supposedly endemic within the disadvantaged population. The message is that the poor must change their characteristics, including their behaviour, before they deserve to work. However, those without jobs might not view such changes as desirable or even possible.

Second, it can require a long time to see results; the gestational period to produce a worker is at least 16 years for developing nations and 25 years or more for highly developed nations. Further, as structuralists recognised, a dynamic economy is always leaving old skills behind and demanding new ones. At any point, there will be a permanent, sizeable, pool of those with

inappropriate skills and education, even if many individuals are able to transition out of the pool in a timely fashion.

Third, as mentioned, there is the danger that the retrained will face a job shortage so that at best they simply displace previously employed workers who will join the ranks of the unemployed.

For these reasons, jobs must be made available that can take workers as they are, regardless of their skills, education, or personal characteristics. Upgrading of these characteristics would be the second step - with much of the necessary training occurring on the job. The unemployed need jobs, not merely the promise of a job for those who successfully reform themselves. Note also that if welfare (including unemployment compensation) is offered as a substitute for a job, this has negative impacts on self-esteem, on public perceptions of the unemployed, and on the human capital (skills and experience) that deteriorates through lack of use. For these reasons, providing welfare rather than work to those who want to work is not only an admission of defeat (the labour market fails to provide enough jobs), but also wastes resources and generates social costs.

After WWII, the notion that **Keynesian policies** could keep aggregate demand at a sufficiently high level to promote robust growth came to dominate western thought. Further, it was believed that high growth would keep unemployment low and thereby reduce poverty rates. Aggregate demand was sustained in the West (especially the US) through spending on defence and public infrastructure investment, and through favourable treatment of private investment. During the early post-war period, economic growth was maintained at an above average pace, and unemployment and poverty rates seemed to fall as a result - apparently validating the Keynesian approach.

The high growth strategy was supplemented by a combination of behaviourist and structuralist labour market policies plus welfare (itself a labour market policy in the sense that one of the goals of welfare for families with children and retirement income for the aged was to reduce the size of the labour force). This was deemed necessary because growth was leaving behind pockets of poverty among disadvantaged groups - with poverty regionally and racially concentrated. However, these policies would not eliminate unemployment and poverty because they failed to provide job creation as a central feature. At best, they redistributed joblessness. Further, a high growth strategy would actually favour the more advanced sectors of the economy that is those with highly skilled and paid workers, which would increase income inequality. Finally, policy that favours high investment would prove to be unsustainable because it would generate macroeconomic instability - evidenced by inflation, currency devaluation, and financial fragility. For these reasons, policy induced recessions would be required to try to restore conditions favourable to macro stability. This led to a **stop-go pattern** of using fiscal stimulus in downturns and tight fiscal policy near a business cycle peak. As a result, expansions would be curtailed long before a sufficient supply of jobs would be created to achieve full employment and address poverty.

Keynesian policy fell out of favour during the **stagflationary** 1970s. In March 1973 the major currencies floated after the collapse of the fixed exchange rate, Bretton Woods system. The social safety nets adopted in the early post-war period were gradually either dropped or underfunded, and neo-liberalism (called neoconservative in the US) played a growing role in the developed economies, such as the US, UK and Australia.

Finally, recessions and financial crises returned after 1970 as the belief that high growth and low unemployment are inconsistent with price stability came to dominate policy formation. Financial

fragility appears to have risen over time, as evidenced by increasingly frequent and severe domestic and international financial crises. Policy makers have turned away from the use of fiscal policy to promote growth, and have largely relied on monetary policy. Monetary policy makers, in turn, generally deny responsibility for maintaining high employment and growth, except to the extent that low inflation promotes a strong economy. Substantial controversy surrounds all of these issues, but it is commonly accepted that attempts to fine-tune the economy through Keynesian style aggregate demand manipulation have proven to be largely unsuccessful. Even if these policies had been successful, there is little political will to return to them today. However, the chosen replacement, neo-liberalism, has not succeeded either.

### **Private sector incentives**

There still remain two main alternatives for the promotion of employment: indirect job creation through incentives given to the private sector, and direct employment by government. Both strategies have been experimented with in many economies.

There are several drawbacks to subsidised employment in the private sector. First, government needs to ensure that firms use the subsidies to create jobs, rather than to reduce private costs of existing employment. In a dynamic economy with jobs continually created and destroyed, this is difficult to police because profit-seeking firms will want to use government funds to subsidise existing jobs, resulting in a leakage of public spending.

Second, as unemployment is concentrated among disadvantaged workers, the policy should encourage firms to employ individuals they would not otherwise have hired. Again, this is difficult to monitor because firms will want to hire those job seekers with the most desirable characteristics that are allowed under program rules, rather than those with average (or lower) characteristics. Further, there is the danger that firms will hire eligible workers, displacing workers with similar characteristics but who do not quite qualify.

Third, there are questions about the time span permitted for eligibility. One goal of the program should be to take workers with little experience or skills in order to prepare them for non-subsidised work. However, if workers are permitted to stay in the program for only a specified period, there is a strong incentive for employers to replace workers at the end of their period of eligibility with newly eligible and subsidised workers. Workers who are forced to leave the program might not find unsubsidised work.

Fourth, the setting of the wage subsidy is not necessarily simple. The subsidy required to induce firms to hire a new worker presumably varies according to the perceived shortfall of the worker's employability relative to the pool of workers from which the firm normally recruits. A sliding scale subsidy might be most effective, but it could be difficult to establish the proper subsidy rates. The required subsidy will also vary according to the firm's need for new workers - in an economic boom, a small subsidy might be sufficient to induce an employer to hire one more worker than the firm would have otherwise employed. In a deep recession, even a 100 per cent wage subsidy might not induce a firm to hire one more worker.

Finally, the payment of wage subsidies necessarily leads to some distortion of the market - some firms will be able to take advantage of the scheme, while others will not. Some existing employees will have to compete with subsidised labour while others will not. Some lines of production will increase output because of additional workers, while others will not; and so on. While none of these potential problems - even if taken together necessarily implies that a

program of wage subsidies should not be tried, the potential problems would seem to lead to the conclusion that such a program probably cannot by itself solve the problem of joblessness and ensure the right to work.

Of course, private sector subsidies will not work without a private sector sufficiently developed that it is capable of offering employment to a significant portion of the population. In some developing nations, especially in rural regions, such a policy will have limited application. Only direct job creation by government will provide a sufficient supply of non-agricultural work to reduce joblessness and provide a living wage.

### **Direct job creation by government**

**We conclude that raising aggregate demand, increasing human capital, and raising the incentives to private employers will fall short of ensuring the right to work.** While each of these policies might be desirable in its own right, they must be supplemented by direct job creation by government. Most governments engage in some form of job creation for the purpose of relieving unemployment. Arguably, the nations that achieved anything close to full employment in the post-war years used a variety of such programs to keep unemployment low. They all maintained, in one form or another, a buffer of jobs that were inclusive to the least skilled workers that were likely to be unemployed otherwise.

The main criticism of government job creation schemes is that, unlike the JG which we have outlined in detail, they typically do not provide ongoing employment under normal working conditions and their coverage is limited to particular groups, such as rural workers (Mahatma Gandhi National Rural Employment Guarantee Scheme in India), heads of household (Jefes de Hogar program in Argentina) and youth (Youth Guarantee in EU countries).

## **Appendix: Advanced Material**

### **Measuring the costs of unemployment buffer stocks**

Under inflation targeting monetary policy regimes, central bankers use the persistent pool of unemployed (and other forms of labour underutilisation, for example, underemployment) as a buffer stock to achieve a desirable inflation outcome. If their inflation outlook rises above their target rate they will induce higher rates of unemployment by increasing in interest rates until they are satisfied their inflation target is being met (see Chapter 15).

While some extreme elements of the profession, who still consider rational expectations to be a reasonable assumption, will deny any real output effects, most economists acknowledge that any disinflation engendered by this approach will be accompanied by a period of reduced output and increased unemployment (and the related social costs) because a period of (temporary) slack in the economy is required to break inflationary expectations.

The real question then is how large are the output losses following a discretionary disinflation? There is overwhelming evidence to suggest that the cumulative costs of this strategy in real terms have been substantial.

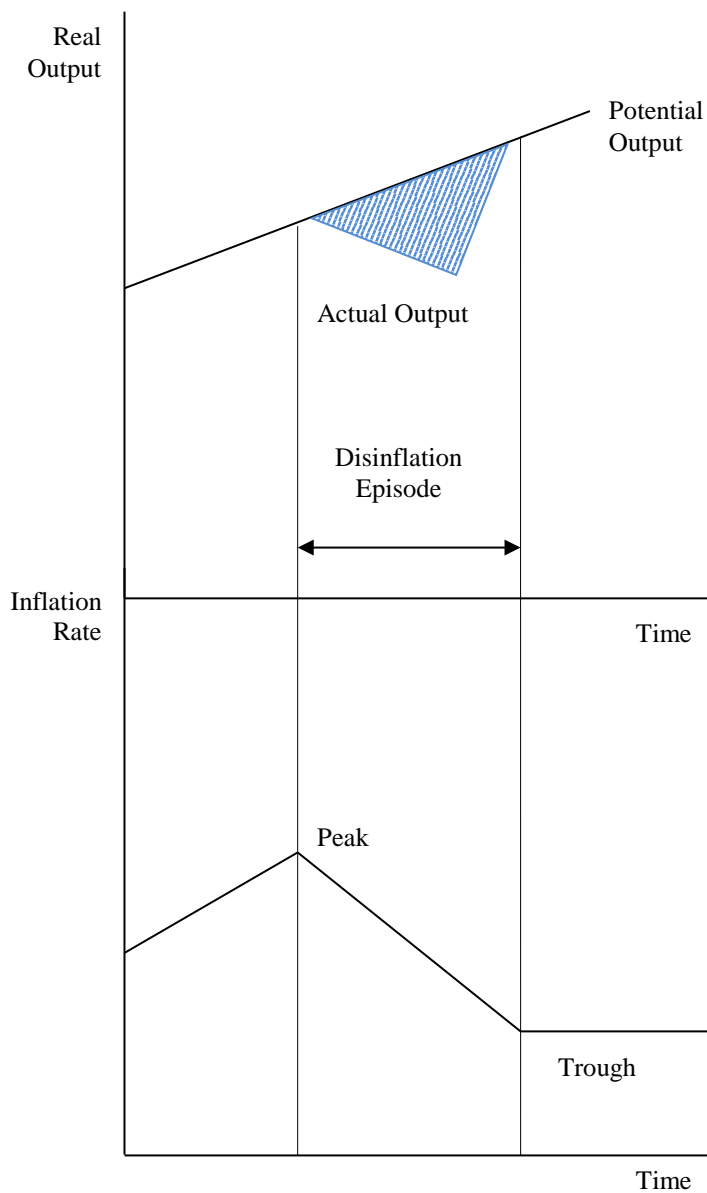
Economists measure these real costs in terms of a **sacrifice ratio**, which is the accumulated loss of output during a disinflation episode expressed as a percentage of initial output divided by the overall reduction in the inflation rate.

For example, if the sacrifice ratio was two it would mean that a one-point reduction in the trend inflation rate is associated with a GDP loss equivalent to 2 per cent of initial output.

Figure 12.1 is a simple graphical depiction of the sacrifice ratio concept and captures the way most empirical studies have pursued its estimation.

**The cumulative output loss resulting from actual output falling below potential output is depicted by the shaded area.** In Figure 12.1, we have deliberately constructed output to resume at its potential level at the exact end of the disinflation period (defined as the period between the peak inflation and the trough inflation). This is the normal assumption adopted in empirical studies.

**Figure 12.1 The sacrifice ratio and disinflation episode**





The depiction in Figure 12.1 assumes that the disinflation episode has a relatively finite, short-term impact on real GDP growth and before long the actual growth path converges on the potential path (that was unchanged despite the introduction of the disinflationary policy).

However, in the real world, it is clear that a prolonged period of reduced real GDP growth lasts beyond the formal disinflation period and that the potential real GDP growth path also declines as the collateral damage of low confidence among firms curtails investment (which slows down the growth in productive capacity and hence potential output).

As a consequence, the estimates of sacrifice ratios, based on the conception shown in Figure 12.1, will be biased downwards because they ignore the impacts of **output persistence and hysteresis**.

The concept of persistence means that actual output remains below its potential after the disinflation period has finished. The longer this output gap exists, the longer is the persistence.

Relatedly, hysteresis (in this context) refers to the permanent losses of potential output that arise as a consequence of the disinflation policy.

The important point is that to accurately estimate the sacrifice ratio, researchers must not only consider the short-term losses but also the longer-term losses arising from persistence and hysteresis.

Figure 12.2 stylises the impacts of persistence and hysteresis arising from a disinflationary policy stance. It is clear that the real output losses are much greater than would be estimated using the restricted concept shown in Figure 12.1.

From the inflation peak, real output falls immediately as before. But after a time, the reduced levels of economic activity erode confidence among consumers and firms. Consumers fearing even higher unemployment restrict consumption spending and firms respond to the lack of sales order by cutting investment plans.

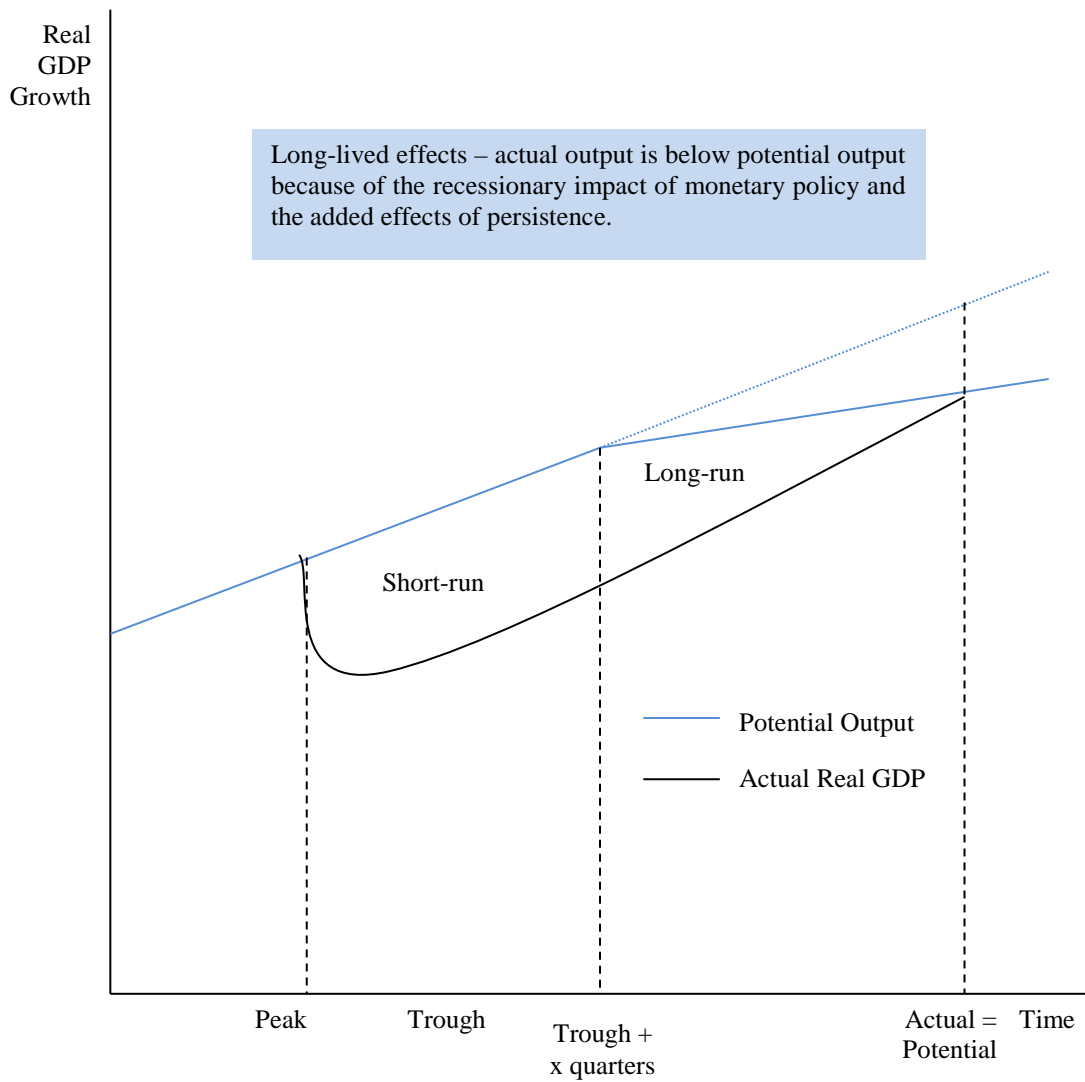
Two impacts occur as a result: (a) the potential real output path falls (from Trough + x quarters on Figure 12.2), reducing the growth capacity of the economy; and (b) actual real output deviates from its potential path for much longer than otherwise would have been the case.

As a result, the estimated costs of the disinflation are much larger. Eventually actual and potential output paths may converge but at that point there is less output and national income and almost certainly, persistently higher unemployment.

Mass unemployment was initially caused by the deliberate cutting of aggregate demand due to the contractionary policy stance of the government, but the subsequent expansion of output can become capacity constrained as a result of a slow growing or falling potential output level, due to the weak inducement to invest.

The fiscal austerity policies pursued by governments during the global financial crisis also have had this impact. It is much harder to then restore robust growth because it takes longer to ensure there is also potential capacity to support it without triggering inflation.

**Figure 12.2** Sacrifice ratios with persistence and hysteresis



Mitchell and Muysken (2008) drew on an extensive literature analysis and their own empirical work to conclude the following:

- Formal econometric analysis does not support the case that inflation targeting delivers superior economic outcomes in terms of reducing the costs of disinflation. Both targeters and non-targeters enjoyed variable outcomes and there is no credible evidence that inflation targeting improves performance as measured by the behaviour of inflation, output, or interest rates.
- There is no credible evidence that central bank independence and the alleged credibility bonus that this brings, leads to the faster adjustment of inflationary expectations to the policy announcements. There is no evidence that targeting affects inflation behaviour differently.
- Estimates of sacrifice ratios confirm that disinflations are not costless. Significantly, the average estimated GDP sacrifice ratios have increased over time, from 0.6 in the 1970s to

1.9 in the 1980s and to 3.4 in the 1990s. That is, on average a reduction of trend inflation by one percentage point resulted in a 3.4 per cent cumulative loss in real GDP in the 1990s.

- Australia, Canada, and the UK, who announced formal policies of inflation targeting in the 1990s, do not have substantially lower sacrifice ratios compared to G7 countries that did not announce such policies. Australia does appear to record a lower average ratio during the targeting period than in the 1980s. However, this figure is not lower than the average for all previous periods. Canada records a higher sacrifice ratio in the 1990s of 3.6 per cent. The ratio for the UK during inflation targeting is significantly higher at 2.5 per cent (relative to quite low sacrifice ratios in previous periods). Italy, Germany, Japan and the US, average 0.6, 2.3, 2.9 and 5.8, respectively.

The evidence is clear that inflation targeting countries have failed to achieve superior outcomes in terms of output growth, inflation variability and output variability; moreover there is no evidence that inflation targeting has reduced the persistence of inflation.

Other factors have been more important than targeting *per se* in reducing inflation. Most governments adopted fiscal austerity measures in the 1990s in the mistaken belief that fiscal surpluses were the exemplar of prudent economic management and provided the supportive environment for monetary policy.

The fiscal cutbacks had adverse consequences for unemployment and generally created conditions of labour market slackness. Labour underutilisation defined more broadly to include, among other things, underemployment, rose in the same countries.

Further, the comprehensive shift to active labour market programs, welfare-to-work reform, dismantling of unions and privatisation of public enterprises also helped to keep wage pressures down.

It is clear from statements made by various central bankers (in addition to their formal obligations) that a belief in the long-run trade-off between inflation and employment embodied in the NAIRU has led them to pursue an inflation-first strategy at the expense of unemployment.

Disinflations are not costless irrespective of whether targeting is used or not. An average sacrifice ratio of 3.5 in the 1990s means that any attempt to bring down inflation nowadays with 1 per cent-point will result in a cumulative loss in GDP of 3.5 per cent on average. In terms of unemployment the latter can be interpreted roughly speaking as a cumulative increase by 7 per cent.

The increase in the sacrifice ratio over time illustrates that reduced inflation variability allows more certainty in nominal contracting with less need for frequent wage and price adjustments. The latter in turn means less need for indexation and short-term contracts and leads towards a flatter short-run Phillips curve. Thus a consequence of inflation targeting is that the costs of disinflation become higher.

The late Franco Modigliani (2000: 3), who was one of the economists who coined the term NAIRU, reflected on the legacy he had created:

Unemployment is primarily due to lack of aggregate demand. This is mainly the outcome of erroneous macroeconomic policies ... [the decisions of Central Banks] ... inspired by an obsessive fear of inflation, ... coupled with a benign neglect for unemployment ...

have resulted in systematically over tight monetary policy decisions, apparently based on an objectionable use of the so-called NAIRU approach. The contractive effects of these policies have been reinforced by common, very tight fiscal policies.

One of the major problems of inflation targeting as a policy paradigm is that it has been accompanied by a view that fiscal policy has to be passive and not compromise the inflation target. As a result, economies have tolerated persistently high rates of labour underutilisation despite having achieved low inflation.

As noted earlier in the Chapter, persistent unemployment results not only in massive losses of real output and national income, but other real costs are also endured by the nation, including the depreciation of human capital, family breakdowns, increasing crime, and increasing medical costs.

These additional costs, in particular the depreciation of human capital, also mean that the effectiveness of the unemployed pool as a price anchor deteriorates over time, with ever larger numbers of fresh unemployed or underemployed required to function as a price anchor that stabilises wages.

Given the scale of these costs, it is unlikely that using a persistent pool of unemployed (or casualised underemployed) is the most cost effective way to achieve price stability.

### **Buffer Stocks in agriculture**

The JG bears many similarities to (and a significant difference from) agricultural price support buffer stock schemes that governments have regularly used to stabilise prices and incomes in the agricultural sector.

For example, in November 1970, the Australian Government introduced the Wool Floor Price Scheme. The scheme was relatively simple and worked by the Government establishing a floor price for wool after hearing submissions from the Wool Council of Australia and the Australian Wool Corporation (AWC).

The aim of the system was to stabilise farm incomes and led to an agreed price for wool being paid to the farmers. The Government then stabilised the price at this guaranteed level by using the AWC to purchase stocks of wool in the auction markets if demand was low and selling it if demand was high.

By being prepared to hold buffer wool stocks in times of low demand and release them again in times of high demand, the government was able to guarantee incomes for the farmers around the stable price.

The contention that ultimately led to the demise of the system was whether the guarantee constituted a reasonable level of output in a time of declining demand. Farmers clearly had an incentive to over-produce wool knowing that the government would buy any excess not demanded by the auction markets.

The JG approach is also based on the maintenance of a variable buffer stock of jobs in line with fluctuations in private demand. However, the weaknesses of the agricultural scheme do not apply to a JG.

First, if there is a price guarantee (the JG wage) below the prevailing market price and a buffer stock of working hours constructed to absorb the excess supply at the current market price, then a form of full employment can be generated without tinkering with the price structure.

Second, the incentives to over-production in commodity buffer stock systems do not apply to maintaining a labour buffer stock because no one is concerned that employed workers would have more children than unemployed workers.

Benjamin Graham wrote in the 1930s about the idea of stabilising prices and standards of living by surplus storage. He documents how a government might deal with surplus production in the economy: “[The] State may deal with actual or threatened surplus in one of four ways: (a) by preventing it; (b) by destroying it; (c) by dumping it; or (d) by conserving it.” Graham (1937: 18).

In the context of an excess supply of labour, governments now tend to choose the dumping strategy via the unemployment buffer stock approach (the NAIRU). However, it is less wasteful to use the conservation approach, which is reflected in the JG framework.

Graham (1937: 34) noted that:

The first conclusion is that wherever surplus has been conserved primarily for future use the plan has been sensible and successful, unless marred by glaring errors of administration. The second conclusion is that when the surplus has been acquired and held primarily for future sale the plan has been vulnerable to adverse developments.

This distinction is important when we conceive of the way employment buffer stock models might work in practice. The Australian Wool Scheme was an example of storage for future sale and was not motivated to help the consumer of wool but the producer.

The JG policy is an example of storage for use where the “reserve is established to meet a future need which experience has taught us is likely to develop” (Graham, 1937: 35). Since government spending on the JG program would pay a basic wage, it establishes a stable value for the currency: the currency will be worth the amount of labour it can buy on the margin, which is the JG wage. This provides a backing to the currency - and a much better backing than gold.

We won't go into it in detail, but a currency backed by gold (or wheat, or wool) will keep gold (or wheat or wool) fully employed in the sense that government stands ready to buy it for its reserves. No reasonable person believes there is a public interest in keeping gold fully employed. It is hard to see why government should even care what the price of gold is. Thus, why it should stand ready to buy it if its price starts to fall (because supply exceeds demand). One might make an argument that a buffer stock program for wheat or wool does make sense because it helps the producers of those commodities, which might help to stabilise the economy if these commodities are a large part of the nation's production. Stabilising incomes of wheat farmers or sheep herders could be beneficial.

### **A plausible adjustment path**

A plausible story to show the dynamics of a JG economy compared to a NAIRU economy would begin with an economy with two labour sub-markets: Sector A (primary) and Sector B (secondary) which broadly correspond to the dual labour market depictions in the literature, which distinguish between stable, well-paid primary jobs and low-paid, precarious secondary jobs.

Assume as before that firms set prices according to mark-ups on unit costs in each sector.

Wage setting in Sector A is contractual and responds in an inverse and lagged fashion to relative wage growth (Sector A / Sector B) and to the wait unemployment level (displaced Sector A workers who think they will be re-employed soon in Sector A).

So when the ratio of Sector A wages to Sector B falls, workers in Sector A will eventually seek to reinstate the past relativity, which reflects their sense of worth in the wage structure and their bargaining capacity as skilled workers. Increasing numbers of unemployed workers waiting for work in Sector A (but not taking Sector B jobs) also depress wages growth in Sector A.

In a non-JG economy, a government stimulus increases output and employment in both sectors immediately. Wages are relatively flexible upwards in Sector B and respond immediately. The compression of the Sector A / Sector B wage relativity stimulates wage growth in Sector A after a time.

Wait unemployment falls due to the rising employment demand in Sector A, but also rises due to the increased probability of getting a job in Sector A. That is, workers who had previously taken Sector B jobs in desperation or were classified as being outside the labour force may leave their Sector B jobs or re-enter the labour force in expectation of a prospect of a better paying Sector A job, which is more in line with their skill levels. The net effect of these two movements is unclear at the conceptual level.

The total unemployment rate falls after participation effects are absorbed. The wage growth in both sectors may force firms to increase prices, although this will be attenuated somewhat by rising productivity as utilisation increases.

A combination of wage-wage and wage-price mechanisms in a soft product market can then drive inflation. These are the type of adjustments that are described in a Phillips curve economy.

To stop inflation, the government has to repress demand. The higher unemployment brings the real income expectations of workers and firms into line with the available real income and the inflation stabilises. This is a typical NAIRU story.

Now consider what would be different in a JG economy. Introducing the JG policy into the depressed economy puts pressure on Sector B employers to restructure their jobs in order to maintain a workforce.

For given productivity levels, the JG wage constitutes a floor in the economy's cost structure. The dynamics of this economy change significantly.

The elimination of all but wait unemployment in Sector A and frictional unemployment does not distort the relative wage structure so that the wage-wage pressures arising from variations in the Sector A / Sector B relativity that were prominent previously, are now reduced.

The wages of JG workers (and hence their spending) represents a modest increment to nominal demand given that the state was typically already supporting them on unemployment benefits. It is possible that the rising aggregate demand tightens the product market, and the demand for labour rises in Sector A.

But there are no new problems faced by employers who wish to hire labour to meet the higher sales levels in this environment. They must pay the going wage rate, which is still preferred to the lower JG wage by the appropriately skilled workers. The rising aggregate demand *per se*

does not invoke inflationary pressures if firms can increase capacity utilisation to meet the higher sales volumes.

With respect to the behaviour of workers in Sector A, one might think that the provision of the JG will lead to workers quitting bad private sector employers. It is clear that with a JG, wage bargaining is freed from the general threat of unemployment.

However, it is unclear whether this will lead to higher wage demands than otherwise. In professional occupational markets, some wait unemployment will remain. Skilled workers who are laid off are likely to receive payouts that forestall their need to get immediate work.

They have a disincentive to immediately take a JG job, which is a low-wage and possibly stigmatised option. Wait unemployment disciplines wage demands in Sector A. However, demand pressures may eventually exhaust this stock, and wage-price pressures may develop.

A crucial point is that the JG does not rely on the government spending at market prices which then exploits the expenditure multiplier to achieve full employment as is characteristic of traditional Keynesian pump-priming. In this sense, traditional Keynesian remedies fail to provide an integrated full employment-price anchor policy framework.

From the above analysis it is clear that the introduction of a JG eliminates the traditional Phillips curve trade-off.

Consider Figure 12.3. In a Phillips curve world, imagine that the unemployment rate was at  $UR_A$  and the inflation rate was  $I_A$ .

The full employment unemployment rate is  $UR_{FULL}$ , which denotes frictional unemployment.

The government is under pressure to reduce the excessive unemployment and if it increased aggregate demand, wage-wage and wage-price pressures would drive the inflation rate up to  $I_B$  (a movement along the Phillips curve from Point A to Point B) and achieve full employment.

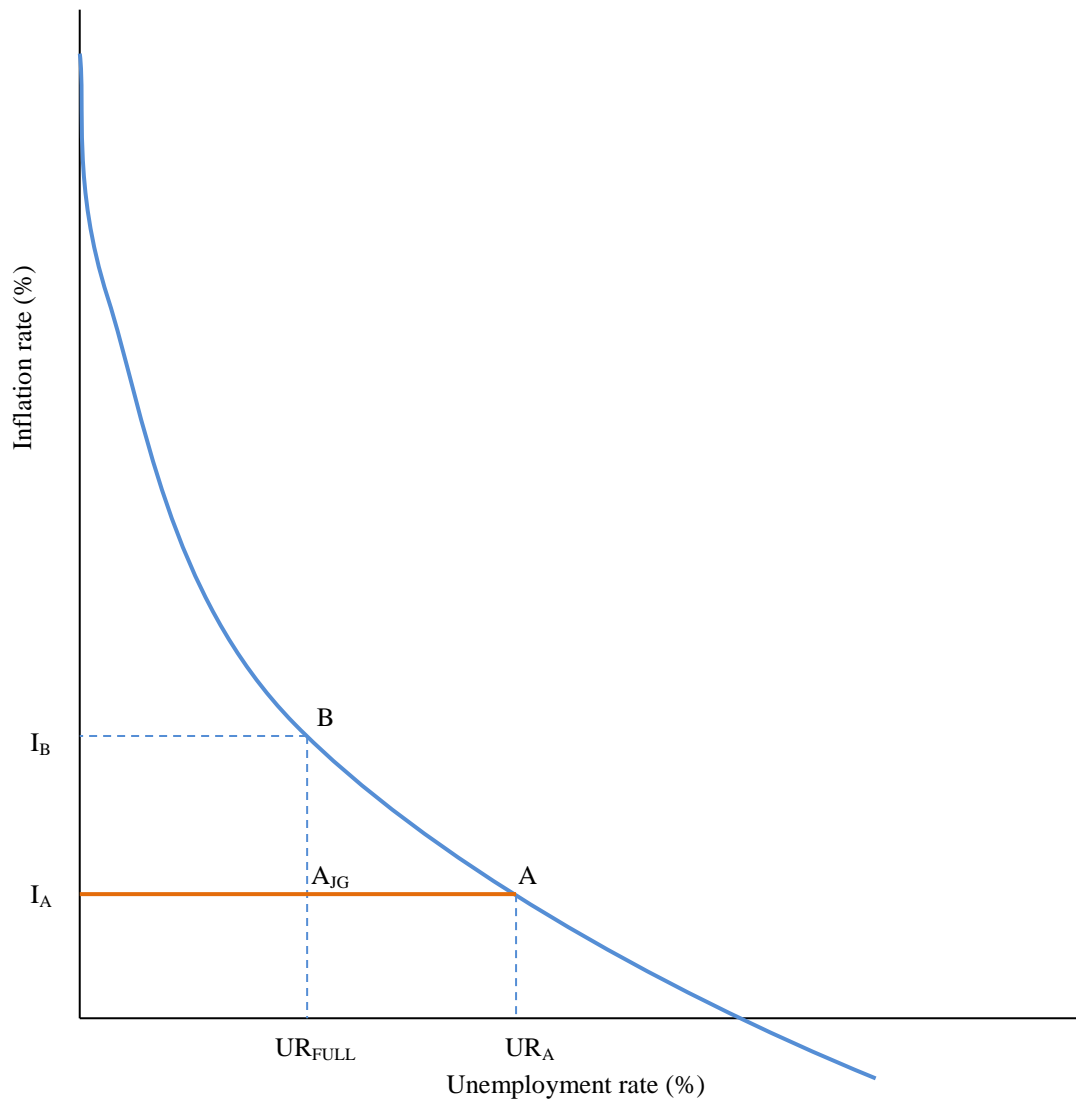
However, there is no guarantee that the inflation rate would remain stable at  $I_B$ . Certainly, the NAIRU model would predict that bargaining agents would incorporate the new higher inflation rate into their expectations and the Phillips curve would start moving out. Whether that happens is not relevant here and we considered those issues in Chapter 11.

If the government initially responded to the excessive unemployment at Point A by introducing a Job Guarantee it could absorb workers in jobs commensurate with the difference between  $UR_A$  and  $UR_{FULL}$ , although in reality as more work was available, workers from outside the labour force (the hidden unemployed) would also take JG jobs in preference to remaining without income.

But whatever the quantum of workers that would initially be absorbed in the JG pool, the economy would move from A to  $A_{JG}$  rather than from A to B.

**In other words, the introduction of the JG eliminates the Phillips curve. The macroeconomic opportunities facing the government are not dictated by a perceived unemployment and inflation trade-off and any fear that that trade-off might be unstable (as in a NAIRU world).**

**Figure 12.3 The Job Guarantee and the Phillips curve**



Rather full employment and price stability go hand in hand.

**Would the NAIBER be higher than the NAIRU?**

We have learned that the NAIRU defines the unemployment buffer stock associated with stable inflation whereas in the employment buffer stock approach to price stability, the **Non-Accelerating Inflation Buffer Employment Ratio (NAIBER)** is the BER that results in stable inflation via the redistribution of workers from the inflating private sector to the fixed price JG sector.

The main principle of a buffer stock scheme like the JG is straightforward – it buys off the bottom at zero bid, which means that the worker has no other employer bidding for their services and cannot put pressure on wages that are above this floor. The choice of the wage floor may have a once-off effect on the price level.



An interesting question to explore relates to the relative sizes of the NAIBER vis-à-vis the NAIRU. There are two arguments that might be used to argue that the NAIBER would have to be larger than the NAIRU for an equivalent amount of inflation control.

First, the intuitive, but somewhat inexact view is that because JG workers will have higher incomes (than when they were unemployed) a switch to this policy would always see demand levels higher than under a NAIRU world.

As a matter of logic then, if the NAIRU achieved output levels commensurate with price stability, then other things equal, a higher demand level would have to generate inflationary impulses. So according to this view, the level of unemployment associated with the NAIRU is intrinsically tied to a unique level of demand at which inflation stabilises.

It should be noted that while it is clear that JG workers will enjoy higher purchasing power under a JG compared to their outcomes under a NAIRU policy, it is not inevitable that aggregate demand overall would rise with the introduction of JG.

But assuming aggregate demand is higher when the JG is introduced than that which prevailed in the NAIRU economy, we might wonder why inflation is not inevitable as we replace unemployment with (higher paying) employment.

Rising demand *per se* does not necessarily invoke inflationary pressures, because by definition the extra liquidity is satisfying a net savings desire by the private domestic sector.

Additionally, in demand constrained economies, firms are likely to increase capacity utilisation to meet the higher sales volumes rather than risk losing market share by increasing prices. There would be no obvious cost pressures forcing the firms to increase prices.

Further, the aggregate demand impulse required to return the economy to what we might call loose full employment under the JG is less than would be required in a NAIRU economy where the government would have to pay market prices to bring the idle resources back into productive use.

In that context, it is clear that if there were any demand-pull inflation it would be lower under the JG. So there are no new problems faced by employers who wish to hire labour to meet the higher sales levels.

Additionally, any initial rise in demand will stimulate private sector employment growth while reducing JG employment and spending.

Second, and related, it is claimed that the introduction of the JG reduces the threat of unemployment which serves to discipline the wage setting process. The impact on the price level by the introduction of the JG will also depend on qualitative aspects of the JG pool relative to the NAIRU unemployment buffer.

In the NAIRU logic, workers may consider the JG to be a better option than unemployment. Without the threat of unemployment, wage bargaining workers then may have less incentive to moderate their wage demands notwithstanding the likely disciplining role of wait unemployment in skilled labour markets.

However, when wait unemployment is exhausted private firms would still be required to train new workers in job-specific skills in the same way they would in a non-JG economy.

The functioning and effectiveness of the buffer stock in question is critical to its operation as a price anchor. There is overwhelming evidence that long-term unemployment generates costs far in excess of the lost output that is sacrificed every day the economy is away from full employment.

It is clear that the more immediately employable are the unemployed, the better the price anchor will function. After an extended downturn the unemployment buffer stock will be composed of a significant proportion of long-term unemployed.

JG workers are far more likely to have retained higher levels of skill than those who are forced to succumb to lengthy spells of unemployment. It is thus reasonable to assume that an employer would consider a JG worker, who is already demonstrating a commitment to working, a superior training prospect relative to an unemployed and/or hidden unemployed worker.

The JG policy would thus reduce the hysteretic inertia embodied in the long-term unemployed and allow for a smoother private sector expansion. Therefore JG workers would constitute a more credible threat to the current private sector employees than, say, the long-term unemployed.

When wage pressures mount, an employer would be more likely to exercise resistance if they knew they could hire from the fixed-price JG pool.

This changes the bargaining environment rather significantly because firms now have reduced hiring costs. Previously, the same firms would have lowered their hiring standards and provided on-the-job training and vestibule training in as the labour market tightened.

As a consequence, longer term planning with cost control would be enhanced. So in this sense, the inflation restraint exerted via the NAIBER is likely to be more effective than using a NAIRU strategy.

In summary, the JG buffer stock is likely to be a qualitatively superior inflation fighting pool than the unemployed stock under a NAIRU. In that sense, the NAIBER will be lower than the NAIRU, which means that private sector employment can be higher before the inflation barrier is reached.

Another associated factor relates to the behaviour of professional occupational markets. In those markets, while any wait unemployment will discipline wage demands, the demand pressures may eventually exhaust this stock and wage-price pressures may develop.

With a strong and responsive tertiary education sector combined with strong firm training processes, skill bottlenecks can be avoided more readily under the JG than with an unemployed buffer stock in place. The JG workers would already be maintaining their general skills as a consequence of an on-going attachment to the employed workforce.

The qualitative aspects of the unemployed pool deteriorate with duration making the transition back in the labour force more problematic. As a consequence, the long-term unemployed exert very little downward pressure on wages growth because they are not a credible substitute

## References

- Beveridge, W. (1945) Full Employment in a Free Society: A Summary, *New Statesman and Nation and Reynolds News*, London. Accessed 25 February 2016 from [http://lib-161.lse.ac.uk/archives/beveridge/9A\\_79\\_Full\\_employment\\_in\\_a\\_free\\_society.pdf](http://lib-161.lse.ac.uk/archives/beveridge/9A_79_Full_employment_in_a_free_society.pdf).
- Forstater, M. (2006) Working for a Better World: The Social and Economic Benefits of Employment Guarantee Schemes, Center for Full Employment and Price Stability, University of Missouri-Kansas City and the Levy Economics Institute of Bard College, available at: [http://www.economistsforfullemployment.org/knowledge/presentations/Session1\\_Forstater.pdf](http://www.economistsforfullemployment.org/knowledge/presentations/Session1_Forstater.pdf).
- Graham, B. (1937) *Storage and Stability*, McGraw Hill, New York.
- Kaboub (2008) “Employer of Last Resort Schemes” in P. A. O’Hara (ed), *International Encyclopedia of Public Policy: Governance in a Global Age*, GPERU, Perth.
- Minsky, H. (1986) *Stabilizing an Unstable Economy*, Yale University press, New Haven, USA.
- Mitchell, W.F. and Muysken, J. (2008) *Full Employment abandoned*, Edward Elgar, Cheltenham UK and Northampton USA.
- Modigliani, F. (2000) ‘Europe’s Economic Problems’, *Carpe Oeconomiam Papers in Economics*, 3rd Monetary and Finance Lecture, Freiburg, April 6.
- Ormerod, P. (1994) *The Death of Economics*, Faber and Faber, London.
- Petty, W. (1662) “*A Treatise of Taxes and Contributions*”, in Murphy, A.E. (ed.) (1997) *Monetary Theory 1601-1758*, Vol. 2, Routledge, London and New York.
- Sen, A. (1997) “Inequality, Unemployment and Contemporary Europe.” *International Labour Review*, 136(2), 155-172

# Chapter 13: Introduction to Monetary and Fiscal Policy Operations

## Chapter Outline

- 13.1 Introduction
- 13.2 The Central Bank
  - The payments system, reserves and the interbank market
- 13.3 The Treasury
  - Government and private financial accounting
  - Sectoral balances
- 13.4 Coordination of Monetary and Fiscal Operations
  - A numerical example using balance sheets
  - Is there a sufficient demand for treasury debt?
- 13.5 Taxes and Sovereign Spending
- 13.6 Currency Sovereignty and Policy Independence
- Appendix: Advanced Material
  - Monetary policy in the open economy, causes and consequences of capital flows

## Learning Objectives

1. Understand the roles of the treasury and the central bank.
2. Recognise why and how liquidity management by the central bank accompanies the operation of fiscal policy.
3. Acknowledge that the design of the taxation system should be motivated by equity and behavioural objectives and not revenue raising.
4. Appreciate that a necessary condition for the independence of macroeconomic policy is the sovereignty of the domestic currency.

## 13.1 Introduction

In this chapter we will address three main topics:

1. How is monetary policy conducted?
2. How is fiscal policy conducted?
3. How do the central bank and the treasury coordinate their operations to enable the sovereign government to spend?

We start this chapter by briefly exploring the roles of the central bank and treasury in an economy with its own floating sovereign currency. We again highlight that, while accounting principles are universal, their application to households and firms as currency users differs fundamentally from the central bank as a currency issuer.

Our focus in this chapter is operational practice. For many years there has been a disconnect between the textbooks' treatment of the implementation of fiscal policy and the institutional arrangements in different countries, such as the USA, UK and Australia, which while following similar principles, do differ in the way they conduct fiscal policy. We provide a generic and simplified description of how fiscal policy is conducted. However in the literature it is shown that the self-imposed constraints on central banks buying treasury debt on the primary market does not affect the operational outcomes in a meaningful way (see Lavoie, 2013; and Tymoigne, and Wray, 2013).

We then revisit the role of taxation in a modern monetary system. The chapter concludes with a further discussion of the crucial importance of a sovereign floating currency for policy independence. In the Appendix we extend the analysis of central bank operations to take account of an open economy.

Chapter 14 then examines fiscal policy in more detail. We explore some key debates about the conduct of fiscal policy, which include whether the reliance on fiscal policy to achieve full employment could cause (i) crowding out; (ii) high rates of inflation or even hyperinflation, and (iii) unsustainable deficit and debt to GDP ratios.

## 13.2 The Central Bank

Modern governments operate with a central bank. In some nations, the central bank is formally independent of the treasury although, typically the elected government still appoints the senior management ('the board') and maintains the right to overrule monetary policy decisions. But that political control aside, most central bank managers (the Board of Governors, in the case of the US; and the Monetary Policy Committee in the UK) have some independence from elected representatives and as well from the administration, because they set a target overnight or interbank interest rate, which is now the primary tool to implement monetary policy. The presumption is that an independent body charged with formulation of monetary policy will make better decisions.

In practice, the central bank's independence is not great - for a variety of reasons. For example, in the US, the Federal Reserve Bank (known as the 'Fed') is a 'creature of Congress', subject to the laws that Congress has promulgated; indeed, the Fed was created by an Act of Congress (the

1913 Federal Reserve Act-FRA), and Congress has periodically mandated changes to Fed operations. Similarly in the UK and Australia, the Bank of England and Reserve Bank of Australia are subject to legislation passed in their respective Parliaments. The Bank of England is subject to a CPI inflation target of 2 per cent per year, whereas in Australia, the CPI inflation target has been set between 2 and 3 per cent per year since 1993. The US Fed is not the subject of a specific inflation rate target, but the Federal Open Market Committee (FOMC, 2016) states that a CPI inflation rate target of 2 per cent is “most consistent over the longer run with the Federal Reserve’s statutory mandate.” Low, stable inflation is claimed to provide a more certain environment for the private sector to make spending decisions.

There is a more important reason to doubt the central bank’s independence: its interest rate (liquidity) management operations are largely accommodative because it responds to the needs of private banks as well as to the actions taken by the treasury, which necessitates close coordination with treasury. In this section we will provide a summary of the central bank’s operations. In Chapter 15, we will go into more detail, by addressing issues related to control of the money supply, setting of interest rates, and provision of reserves through lender of last resort operations.

### **The payments system, reserves and the interbank market**

As we have mentioned, central banks in most developed economies conduct monetary policy by setting and announcing a target overnight (interbank) interest rate each month. The target rate must be achieved; otherwise the financial system would be subject to major uncertainty. The **interbank rate** is the rate at which banks lend reserves to each other.

The central bank needs to add or drain reserves to ensure that the banking system has just the desired amount of reserves (or the required amount in those nations like the USA that have a legal reserve requirement), which is consistent with the target rate. In Australia, the only requirement on banks is to hold positive reserves.

If there are excess reserves, then market forces will drive the interbank interest rate below its target level as banks alter the price they are willing to loan reserves. Likewise, in the case of a shortage of reserves, the interbank interest rate will be driven above its target level.

Reserves are added through discount window loans, through open market purchases by the central bank of government bonds, and through purchases of gold, foreign currencies, or even private sector financial assets. In other words, banks can secure additional reserves, when there is a (banking) system shortage, either through lending by the central bank or by selling assets to the central bank. In either case, the central bank credits their reserve deposit accounts at the central bank, which addresses the system shortage.

The central bank reverses these actions in the case of excess reserves in the banking system, that is, when the banks have more reserves than they wish to hold (or are required to hold). Banks with excess reserves can pay down loans at the discount window, or they can buy assets (usually treasury debt, although possibly foreign currency or private assets) from the central bank. The central bank will then debit their reserves.

The central bank must estimate and predict reserve supplies and demands, but it is easy to determine whether the banking system faces excess or deficient reserves. The overnight rate will move away from its target, triggering a nearly automatic offsetting addition or drain of reserves

by the central bank. We outline these practices in detail in Chapter 15. Note that if a central bank pays interest on reserves equal to the target rate, then excess reserves will not push the overnight rate below the rate paid by the central bank.

As noted in Chapter 6, private banks hold reserves at the central bank to enable the payments system to function efficiently. Payments for goods and services by customers to retailers of different banks are resolved by, not only the respective bank deposit accounts of these buyers and sellers being adjusted, but also reserves being transferred between the banks of customers to the banks of the retailers. The operation of the payments system can leave some banks with a shortage of reserves, if their customers have engaged in relatively high levels of spending, whereas other banks will have excess reserves. The deficit banks will seek to borrow reserves in the interbank market from the surplus banks and will pay the prevailing interbank interest rate, which should coincide with the target rate.

As noted, in normal times, central banks accommodate the private bank demand for reserves so that they can control the overnight interest rate. This quantity of reserves is nondiscretionary from the point of view of the central bank. It is the interest rate target that is discretionary. In a crisis, the demand for central bank reserves can rise suddenly. The central bank must supply the additional reserves. **It is important to understand that banks do not loan out reserves to customers, contrary to the impression given by mainstream monetary theory.**

The central bank needs to have good lines of communication with treasury to ensure that the financial system is not disrupted by the treasury running either surpluses or deficits, both of which impact on bank reserves. In the next two sections we outline the operation of fiscal policy and how it impacts on reserves.

Finally, we need to recognise that central banks perform other functions, including acting as a lender of last resort. For example, a bank in financial difficulty may not be able to borrow reserves in the interbank lending market, even if excess reserves exist at the aggregate level. Central banks also regulate and supervise private banks and other financial institutions. For example, the central bank might prohibit banks from making certain kinds of loans (that is, credit controls) or from issuing some kinds of deposits. In many nations, the central bank plays some role in ensuring ‘safety and soundness’ of individual banks as well as of the financial system as a whole. Such functions are also performed by other bodies, including divisions of the treasury as well as state or provincial government offices - and even by independent financial sector regulators. In addition, many nations enforce international guidelines on the behaviour of financial institutions, such as those adopted in the Basel Accords.

A detailed examination of bank regulation and supervision is beyond the scope of a macroeconomics textbook. We will have a little bit to say on these matters later in the text, when we discuss financial instability and global financial crises.

### 13.3 The Treasury

The treasury is the fiscal agent of the elected government that is it sets fiscal policy by varying government spending and taxation.

In the distant past, the treasury would spend directly through the issue of money-denominated IOUs, whether these were tally sticks, metallic coins, or paper money. The administrative branch would spend up to the specified amount, with provisions being made for overrides.

The treasury would also be responsible for collecting taxes in the form approved by elected representatives. Normally, this would include the money-denominated IOUs issued by treasury in its spending. In addition, however, the treasury would sometimes be permitted to accept other IOUs, including currency issued by other nations, other types of domestic government IOUs, or even some types of privately-issued IOUs denominated in the domestic currency.

### **Government and private financial accounting**

Even though some principles of accounting are universal, federal financial accounting has never followed, and should not follow, the procedures adopted by households or business firms. We outlined the arguments in Chapter 2 and summarise and supplement these arguments here.

First, the government's objective should be the pursuit of public purpose that is, general welfare (see also Chapters 1 and 12). There is no necessary correlation between this objective and the achievement of a fiscal surplus or deficit, or higher or lower indebtedness.

Second, the government is sovereign. This fact gives to the government an authority that households and firms do not have. In particular, government has the power to tax and to issue money. The power to tax means that government does not need to sell products, and the power to issue currency means that it can make purchases by dispensing IOUs.

In short, **governments like those in Britain, the United States, Japan and Australia, can never run out of money.** These governments can purchase whatever goods and services are for sale in the currency they issue. The government has to consider how best to deploy the real resources that are available to the economy, but there are no intrinsic 'financial' constraints that are relevant to a currency-issuing government. While it is common to regard government tax revenue as income, this revenue is not comparable to that of firms or households. Government can choose to impose new taxes or raise tax rates.

There is no operational procedure through which the national government uses tax receipts or borrowings for its spending. If households choose to pay taxes in cash, the treasury simply issues a receipt and can choose to shred the cash. It does not need these tax receipts in order to spend. Thus it is a mistake to look at the national government's tax receipts as an equivalent concept to the income of households or firms.

Also, **fiscal surpluses (taxation revenue greater than government spending) today do not provide governments with a greater capacity to meet future spending needs, nor do fiscal deficits (taxation revenue less than government spending) erode that capacity.** Indeed, there is no evidence, nor any economic theory, behind the proposition that national government spending ever needs to match national government tax receipts - over any period, short or long. The deficit per unit of time is the difference between taxing and spending over that time. To repeat, taxing, on the one hand, and spending, on the other, are operationally independent procedures.

On the other hand, private firms are constrained because they cannot force buyers to purchase their products or their debt. Even firms with market power recognise that consumers will find substitutes if prices are raised too much. However, taxation creates a demand for public spending in order to make available the currency required to pay the taxes. No private firm can generate demand for its output in this way. Neither firms nor households can live beyond their financial means indefinitely by accumulating debt. Eventually they have to sacrifice spending to pay the



debts back. Thus firms, households, and even state and local governments require income or need to borrow in order to spend.

These statements are not controversial. They are matters of fact. Nor should they be construed as implying that government should raise taxes sky-high or spend without limit. However, they do imply that financing federal spending is different to private budgeting, which is necessary to plan future expenditures.

MMT teaches us that experience in managing a household budget provides no guidance about the management of the government fiscal position, yet, on a daily basis, the message delivered by the media and most politicians is that the same principles apply. However households and firms are users of the currency that the government issues.

### Sectoral balances

The difference between microeconomic and macroeconomic accounting is pertinent. An individual household or firm has a balance sheet that consists of assets and liabilities. The spending of that household or firm is constrained, by its income and balance sheet - by its ability to sell assets or to borrow against them. Its ability to deficit-spend is constrained. A household must get the approval of a bank before its spending can exceed income, (unless it has assets to sell) and therefore borrowing is subject to banking norms.

On the other hand, if we consider households and firms in aggregate, the situation is different. The private sector's ability to deficit-spend, that is spend more than its income, depends on the willingness of another sector to spend less than its income. For **one sector to run a deficit, another must run a surplus** (as this textbook has emphasised from the outset, see Chapter 5). This surplus is saving, that is claims against the deficit sector. In principle, there is no reason why one sector cannot run perpetual deficits, so long as at least one other sector wants to run surpluses.

In the real world, we observe that the U.S. federal government (like most national governments, including the UK and Australia) tends to run persistent deficits. This is matched by a persistent tendency of the non-government sector to save overall – that is, spend less than its total income. The non-government sector accumulates net claims on the government; the **non-government sector's overall saving is equal (by identity) to the government's deficits**. At the same time, the non-government sector's net accumulation of financial assets (or **net financial wealth**) equals, exactly, the government's total net issue of debt - from the inception of the nation. Debt issued between private parties cancels out in net terms; but that between the government and the private sector remains, with the private sector's net financial wealth consisting of the government's net debt.

This identity does not change once we allow for a foreign sector, which is just a part of the non-government sector. Since nations such as the US, the UK and Australia, have in recent decades run persistent current account deficits, the foreign sector has been accumulating net financial claims on those nations in the currency issued by each. These are initially held in the form of cash or reserve balances at the respective central banks, but they are then typically exchanged for government debt in order to earn interest.

Sectoral balances are linked by an identity, so it is a matter of definition that government deficits equal non-government surpluses, and government debt equals non-government net financial

wealth. Yet, as we noted in Chapter 5, these macroeconomic relations are not obvious when one looks to individual firms or households.

### 13.4 Coordination of Monetary and Fiscal Operations

The first central banks were created near the end of the 17th century, namely the Swedish Sveriges Riksbank, or simply Riksbanken, and then the Bank of England, but the US did not get its central bank, the Fed, until 1913. A consolidated government, consisting of a treasury and a central bank, was typically the model that was initially adopted. Even today, there are many countries that operate without a clear division of responsibilities between the central bank and treasury. Hence, the consolidated government is of some theoretical interest.

While it has been common in the MMT literature to begin with an analysis that **consolidates the central bank and treasury into a sovereign government**, we will maintain the division of responsibilities between the central bank and treasury, which is shown below, but we will make a simplifying assumption, when we provide a numerical example of the implementation of fiscal policy.

#### Duties of the central bank:

- Issues currency notes and in many nations coins.
- Issues reserves (at discount window and through open market purchases of government bonds in secondary markets).
- Sets overnight interest rates, operates clearing for interbank payments, and operates clearing for bank settlements with the treasury.

The separation of responsibilities between the central bank and treasury leads to one extra function for the central bank, namely acting as an intermediary with respect to payments made to and from private banks to treasury. This arises because private banks do not have accounts at the treasury (they hold their reserve deposit accounts at the central bank). Thus the analytical simplification associated with consolidation of the treasury and the central bank is minimal.

#### Duties of the treasury:

- Issues coins (in the US).
- Makes payments to the non-government sector.
- Receives tax payments from the non-government sector.
- Issues new government bonds (usually through a specialised public debt management agency).

The two main **voluntary, operational** rules, which are typical of many countries is that:

1. The treasury writes cheques on its account at the central bank, when engaged in spending. It must have sufficient deposits in that account before it writes a cheque.
2. The treasury cannot sell newly issued bonds to the central bank on the primary market; it must sell them to private banks or other investors. However, the central bank can buy these bonds from private banks or other investors on the secondary market.

Thus the treasury typically is prevented by legislation or other rules from selling bonds to its own bank (the central bank on whose account it draws to spend), but it can sell bonds directly to private banks. It is important to understand that these restrictions are not intrinsic but are voluntarily imposed by the government on itself.

The reason for these restrictions has more to do with satisfying ideological preferences that make it hard for governments to spend, rather than any economic or financial necessity or sound fiscal practice. We see that in emergencies, the restrictions are often quickly relaxed to provide more flexibility to government to use its currency-issuing capacity to meet the challenge of the crisis (for example, during the GFC).

When it spends, the treasury injects currency into the economy by crediting the deposit accounts of the sellers of goods and services at the private banks. The private banks send the treasury cheques to the central bank, which credits the private banks' reserves. This means that the banks must have reserve accounts at the central bank to be credited.

Further assume that treasury accepts only currency, in the form of a cheque or transfer, for the payment of taxes. This means that the electronic bank entries of taxpayers are debited, and the central bank debits the reserves of the taxpayers' banks.

Deficit spending means net currency emission; budget surpluses mean the stock of net currency (that is, bank deposits and any holdings of notes and coins) of the non-government sector is reduced.

The view that a central bank might **choose to print money to finance a fiscal deficit is flawed**. If the government runs a deficit, it inevitably net credits bank accounts (credits exceed debits) in the first instance.

However, if the central bank operates with a positive overnight target interest rate, it must have a debt instrument that is a financial asset, on which it pays interest. Hence, the central bank sells bonds, typically treasury debt, as an attractive interest-earning alternative to bank reserves, which can be used to mop up excess reserves, generated by net treasury spending. This is referred to as open market operations (OMO). The banks are making choices about the composition of their portfolios, taking into account the requirement and their need to hold reserves, but also the higher interest rates offered on treasury debt.

These activities are coordinated with the treasury, which will usually issue new bonds more or less in step with its deficit spending. This is because the central bank would run out of bonds to sell. The important point, however, is that such central bank operations are not discretionary, but rather are required to ensure they can consistently achieve their interest rate targets. The quantity of 'liquidity', (reserves), is not discretionary.

### **A numerical example using balance sheets**

We now provide a simplified analysis in which treasury engages in net government spending ( $G + iB > T$ ) of \$100, where  $iB$  denotes interest payments on existing treasury debt, and  $i$  denotes the nominal interest rate. We assume that the interest rate paid by the central bank on reserves held by the private banks is zero, but the target interbank rate is positive. Treasury has an account at the central bank. Under its voluntary rules, it must have sufficient deposits at the central bank to enable its planned spending. For simplicity and to reflect the notion of the

treasury and central bank being consolidated into one entity, we will assume that treasury has sufficient balances at the central bank for its planned net spending.

When treasury net spends, there is an overall increase in private sector bank deposits of \$100, representing payments for goods and services sold to the government. At the same time the reserves held by the private banks at the central bank increase by \$100, which represents an additional asset for the private banks and a liability for the central bank. The rise in the liabilities of the private banks, via the rise in deposits of the private sector, is matched by their increased holdings of reserves at the central bank. Thus the net positions of the central bank and the banks are unchanged (see Figure 13.1).

**Figure 13.1 Balance sheets associated with net government spending**

**Central Bank**

<b>Assets include</b>	<b>Liabilities include</b>	
Treasury Debt held by CB	Reserves of Private Banks	+\$100
	Treasury Deposits	-\$100

**Private Banks**

<b>Assets include</b>	<b>Liabilities include</b>		
Reserves of Banks	+\$100	Deposits of Private Sector	+\$100
Treasury Debt held by Banks			

Note: \$ amounts in bold indicate net changes to balance sheet items after treasury spends \$100.

The reserves held by the private banks at the central bank, which enable the operation of the payments system, have risen by \$100 and while economic activity has marginally increased, private banks may be reluctant to hold an additional \$100 in reserves. Assume they have a leverage ratio of 0.1. Those private banks holding excess reserves will try to lend their excess reserves - that is, a total of \$90, to other banks. Given that there is a **system-wide excess**, that is, an overall excess supply of reserves, the interbank rate would be driven down below its target level in the absence of central bank action.

The central bank will offer \$90 worth of treasury debt, which will attract an interest rate in excess of the target interbank rate, to the private banks. Banks holding the \$90 of excess reserves

will thus have the incentive to buy this treasury debt. This action by the central bank will remove the downward pressure on the interbank rate and thus protect the integrity of monetary policy, which entailed the setting of a target interbank rate. The final balance sheets are shown in Figure 13.2. Thus the coordination of central bank and treasury operations is required to implement this program of government deficit spending.

**Figure 13.2 Balance sheets associated with net government spending**

<b>Central Bank</b>			
<b>Assets include</b>		<b>Liabilities include</b>	
Treasury Debt held by CB	-\$90	Reserves of Private Banks	+\$10
		Treasury Deposits	-\$100

<b>Private Banks</b>			
<b>Assets include</b>		<b>Liabilities include</b>	
Reserves of Banks	+\$10	Deposits of Private Sector	+\$100
Treasury Debt held by Banks	+\$90		

Note: \$ amounts in bold indicate net changes to balance sheet items after treasury spends \$100 and the central bank engages in OMO, selling \$90 of treasury debt to the private banks.

The **Monetary Base** or High Powered Money (HPM), which is defined as the total bank reserves, held by the central bank plus currency held by the non-government sector (that is, banks, non-bank firms and households) has risen by \$10, since the non-government sector's holding of notes and coins is unchanged. The **net financial assets** held by the non-government sector, which here, for simplicity, we are identifying with the domestic private sector, are defined as their holdings of net financial assets plus HPM. This has increased by a total of \$100, with the private banks' holdings of treasury debt rising \$90.

Thus, this **vertical transaction** of \$100 arising from net spending by treasury in the domestic economy has, as expected, increased the net financial assets of the domestic private by an equivalent amount (see the distinction between horizontal and vertical transactions in Chapter 5).

The balance sheet analysis of this vertical transaction can be extended to incorporate the desire of the private sector to hold additional cash, say \$20. The IOUs of the central bank in the form of cash, which are currently not shown in Figures 13.1 and 13.2 will increase by \$20, which will

impact on the private banks' balance sheet and in turn the cash holdings of the private sector. Also, the impact of the central bank paying an interest rate on reserves equal to the target interbank rate can be considered.

Finally, it is important to note that through this numerical example, we have illustrated the operation of the fundamental relationship between sectoral balances in a closed economy. Further, if we denote the change in HPM by  $\Delta M_h$  and the change in the stock of public debt held by the private sector as  $\Delta B$ , we can rewrite the sectoral balance identity as

$$(13.1) \quad G + iB - T \equiv \Delta B + \Delta M_h$$

A fiscal deficit (left hand side  $>0$ ) gives rise to a private sector surplus, which takes the form of a change in the stocks of HPM and treasury debt, held by the non-government sector.

We will see in the next Chapter that this identity is also referred to as the 'Government Budget Constraint' and is interpreted by mainstream economists as an *ex ante* financial constraint on government, based on the need to finance fiscal deficits. In fact, it is just **an accounting identity**, which tells us *ex post* what the changes in the financial aggregates are as a result of discretionary government policy choices and the state of the economy. A currency-issuing government is not financially constrained in the way the mainstream textbooks suggest.

### Is there a sufficient demand for treasury debt?

In most developed economies, and as a result of these voluntary constraints and accounting conventions that the government places on it, if the treasury does not have sufficient deposits to cover mandated spending, it must first sell bonds. The important question is whether there will be sufficient demand for them from the banks.

When banks buy bonds from the treasury, their reserves at the central bank are debited. If a bank that wants to buy bonds has no excess reserves to debit, then it will either go to the interbank market to borrow them from banks with extra reserves, or it will borrow them from the central bank (at the discount window, see Chapter 15).

We know that even if the banking system has no excess reserves, the central bank will respond to any pressure on overnight interest rates that might be created by banks trying to borrow reserves in order to buy the bonds. With an interest rate target the central bank is always accommodating. Thus banks will always be able to get the reserves they need in order to buy bonds. The banks want the bonds rather than reserves, because the interest rate paid is higher.

Specialised financial institutions are ready to buy domestic bonds in most countries. For example, in the US there are 21 primary dealers who are obligated to bid at U.S. government debt auctions. Likewise in the UK, treasury bill primary participants are financial institutions that have agreed, subject to their own due diligence, to bid at UK treasury bill tenders on behalf of investors. These institutions also operate in secondary markets. The evidence reveals that bond issues are typically over-subscribed. In other words, the demand for new issues of treasury bonds is elastic.

## 13.5 Taxes and Sovereign Spending

Previously we have argued that the imposition of a tax that is payable in the national government's own currency will create demand for that currency. We have also seen that sovereign government does not really need revenue in its own currency in order to spend.

This sounds shocking because we are so accustomed to thinking that taxes pay for government spending. This is true for local governments and states that do not issue the currency. It is also not too far from the truth for nations that adopt a foreign currency or peg their own to gold or foreign currencies. When a nation pegs, it needs stocks of gold or foreign currency to which it promises to convert its currency on demand. Taxing removes its currency from circulation making it harder for anyone to present it for redemption in gold or foreign currency. Hence, a prudent practice would be to constrain government spending to the level of tax revenue.

But in the case of a government that issues its own sovereign currency without a promise to convert at a fixed value to gold or foreign currency (that is, the government floats its currency), we need to think about the role of taxes in an entirely different way. Taxes are not needed to pay for government spending. Further, the logic is reversed: government must spend (or lend) the currency into the economy before taxpayers can pay taxes in the form of the currency. **Spend first, tax later is the logical sequence.**

Some who hear this for the first time jump to the question: 'Well, why not just eliminate taxes altogether?' There are several reasons. First, it is the tax that drives the currency. If we eliminated the tax, people probably would not immediately abandon use of the currency, but the main driver for its use would be gone.

The second reason to have taxes is to reduce aggregate demand. Taxes create real resource space in which the government can spend to fulfil its socio-economic mandate. Taxes reduce the non-government sector's purchasing power and hence its ability to command real resources. Take a situation where the national government spending is around 30 per cent of GDP, while tax revenue is somewhat less - say 27 per cent. The net injection of spending coming from the national government is thus about 3 per cent of GDP. If we eliminated taxes (and held all else constant) the net injection might rise toward 30 per cent of GDP. That is a huge increase of aggregate demand, and could cause inflation. Taxes thus free up real resources in the economy (labour and capital), which otherwise would have been used by the non-government sector for private ends. They thus allow the government to spend without coming up against an inflation constraint.

Ideally, it is best if tax revenue moves counter-cyclically - increasing in an expansion and falling in a recession. That helps to make the government's net contribution to the economy counter-cyclical, which helps to stabilise aggregate demand. In this case the fiscal outcome operates as an automatic stabiliser.

All of this was recognised by the American Beardsley Ruml, who chaired the US Federal Reserve Bank in the 1940s. He also wrote two important papers on the role of taxes (*Taxes for Revenue are Obsolete* and *Tax Policies for Prosperity* (Ruml 1946a and 1946b). Let us first examine his cogent argument that sovereign government does not need taxes for revenue, and then turn to his views on the role of taxes.

In a later retrospective article, he emphasised that (Ruml, 1946b: 82-83):

We must recognise that the objective of national fiscal policy is above all to maintain a sound currency and efficient financial institutions; but consistent with the basic purpose, fiscal policy should and can contribute a great deal toward obtaining a high level of productive employment and prosperity.

This view is similar to that propounded in our textbook.

He goes on to say that the US government gained the ability to pursue these goals after WWII due to two changes of great consequence. “The first of these changes is the gaining of vast new experience in the management of central banks. The second change is the elimination, for domestic purposes, of the convertibility of the currency into gold or into any other commodity” (Ruml, 1950: 91).

With those two conditions, “[i]t follows that our federal government has final freedom from the money market in meeting its financial requirements ... National states no longer need taxes to get the wherewithal to meet their expenses.” (Ruml, 1946b: 84)

Why, then, does the national government need taxes? He provides four reasons (Ruml, 1946: 84):

1. As an instrument of fiscal policy to help stabilize the purchasing power of the dollar;
2. To express public policy in the distribution of wealth and of income as in the case of the progressive income and estate taxes;
3. To express public policy in subsidizing or in penalizing various industries and economic groups; and
4. To isolate and assess directly the costs of certain national benefits, such as highways and social security.

The first of these is related to the inflation issue that we discussed above. The second purpose is to use taxes to change the distribution of income and wealth. For example, a progressive tax would reduce income and wealth at the top, while imposing minimal taxes on the poor. The third purpose is to discourage bad behaviour: pollution of air and water, use of tobacco and alcohol, or to make imports more expensive through tariffs (essentially a tax to raise import costs and thereby encourage purchase of domestic output). These are often called ‘sin’ taxes - whose purpose is to raise the cost of the ‘sins’ of smoking, gambling, purchasing luxury goods, and so on. The fourth is to allocate the costs of specific public programs to the beneficiaries. For example, it is common to tax gasoline so that those who use the nation’s highways will pay for their use (tolls on freeways are another way to do this).

Note that while many would see these taxes as a means to pay for government spending, Ruml (1946a) vehemently denies that view in the title of his article, ‘Taxes for Revenue are Obsolete’. Government does not need the gasoline tax to pay for highways. That tax is designed to make those who will use highways think twice about their support for building them. Government does not need the revenue from a cigarette tax, but rather wants to raise the cost to those who will commit the sin of smoking.

Many would say that it is only fair that those who smoke will pay for the costs their smoking imposes on society (in terms of hospitalisations for lung cancer, for example). From Ruml’s



perspective this is not far from the truth. The hope is that the high cost of tobacco will convince more people never to smoke, which thereby reduces the cost to society.

However, the point is not the revenue to be generated. Government can always find the money to pay for hospital construction and operation. Rather, it is to reduce the waste of real resources that must be devoted to caring for those who smoke. The ideal cigarette tax would be one that eliminated smoking, not one that maximised revenue to government. Ruml (1946b: 84) said “The public purpose which is served ... [by the tax] ... should never be obscured in a tax program under the mask of raising revenue.”

Ruml concluded both of his articles by arguing that once we understand what taxes are for, then we can go about ensuring that the overall tax revenue is at the right level.

He concluded (1946b: 85):

Briefly the idea behind our tax policy should be this: that our taxes should be high enough to protect the stability of our currency, and no higher.... Now it follows from this principle that our tax rates can and should be lowered to the point where the federal budget will be balanced at what we would consider a satisfactory level of high employment.

This principle is also one adopted in this textbook, but with one caveat. Ruml was addressing the situation in which the external sector balance could be ignored (which was not unreasonable in the early post war period). In today’s world, in which some countries have very high current account surpluses and others have high current account deficits, the principle must be modified.

We would restate it as follows: **tax rates should be set so that the government’s fiscal outcome (whether in deficit, balanced, or in surplus) is consistent with full employment.**

Nations that will typically have a current account deficit at full employment (such as Australia, the US, the UK) will normally have a fiscal deficit at full employment (equal to the sum of the current account deficit and the domestic private sector surplus).

Countries like Japan (with a current account surplus at full employment) will have a relatively smaller fiscal deficit at full employment (equal to the domestic private sector surplus less the current account surplus). Countries with larger current account surpluses at full employment, such as Norway, will typically have a fiscal surplus at full employment, so as not to push the economy past the inflation barrier.

## 13.6 Currency Sovereignty and Policy Independence

Currency-issuing nations such as the US, the UK, Australia, Japan, Turkey, and Argentina after it abandoned the currency board or Italy before it joined the Eurozone created **a currency for domestic use**. The government, itself (including the treasury and the central bank), issues, spends, and lends high powered money (HPM), including coins, notes and bank reserves as its liability.

These governments do not promise to convert their HPM to any other currency, nor to gold or any other commodity, at any fixed exchange rate. The flexible exchange rate is a key to maintaining fiscal and monetary policy independence - what we will call sovereignty, although governmental sovereignty certainly has other dimensions as well.

By contrast, as we noted earlier, if a country pegs its exchange rate, it must hold sufficient foreign currency reserves to maintain the peg, which means that it must subsume domestic policy independence to the overriding necessity of accumulating reserves. It thus surrenders monetary sovereignty and hence domestic policy independence in the name of external balance. This is why **a floating exchange rate is a necessary component of policy independence.**

But there is more to it than a flexible exchange rate. The sovereign government spends (buys goods, services, or assets, or makes transfer payments) by issuing a treasury cheque, or, increasingly, by simply crediting a private bank deposit. In either case, however, credit balances (HPM) are created when the central bank credits the reserve account of the receiving bank.

Analogously, when the government receives tax payments, it reduces the reserve balance of a member bank (and, hence the quantity of HPM). Simultaneously, the taxpayer's bank deposit is debited, and their bank's reserves at the central bank are reduced.

While it is usually supposed that the operation is reversed, with a government needing to first receive tax revenue, and then spending that revenue, this sequence is not necessary for any sovereign government. If a government spends by crediting a bank account (issuing its own IOU - HPM) and taxes by debiting a bank account (and eliminating its IOU - HPM), then it is not, as a matter of logic spending tax revenue. In other words, with a floating exchange rate and a domestic currency, the sovereign government's ability to make payments is not revenue constrained precisely because it spends by emitting IOUs.

Note that the sale of its own debt by a sovereign government should not be thought of as a borrowing operation, even though it is frequently described as such. As discussed in the previous sections, the operational effect of government bond sales (whether by the treasury in the new issue market, or by the central bank in open market operations) is to drain any excess reserves created (mostly) by treasury deficit spending. If the bond sales were not undertaken to drain excess reserves, the overnight rate would fall.

The treasury and the central bank work together to ensure that the overnight interest rate target (set by monetary policy) is maintained. They do this through sales or purchases of government bonds to drain or add reserves as necessary to allow the monetary authorities to manage liquidity (reserves) and balance supply and demand for reserves at the desired target interest rate.

When a household or non-sovereign government borrows, it issues an IOU and obtains a bank IOU that it needs in order to spend. The sovereign government, on the other hand, has no need to obtain a deposit before it spends its own currency. It can spend by issuing currency or (today) by crediting a private bank account. It sells a security, not to finance its expenditures but to reduce the outstanding stock of HPM, offering to substitute one of its interest-paying liabilities (the security) for a non- or low- interest-paying liability (the HPM that is then debited from bank accounts).

This is really an interest rate management operation - reducing bank reserves in order to eliminate (non-interest earning) excess reserves that would otherwise place downward pressure on overnight interest rates. As such, **bond sales are really a part of monetary policy, not a required part of fiscal policy.**

The final point to be made regarding such operations by a sovereign government is that the interest rate paid on government bonds is not subject to normal market forces. The sovereign government could always choose to leave excess reserves in the banking system, in which case

the overnight rate would fall toward zero (or the support rate - the interest rate paid by the central bank on reserves).

When the overnight rate is zero, the treasury can always offer to sell bonds that pay a few basis points above zero and will find willing buyers because such bonds offer a better return than the alternative (zero). This drives home the point that a sovereign government with a floating currency can issue bonds at any rate it desires - normally a few basis points above the overnight interest rate target it has set.

There may be economic or political reasons for keeping the overnight rate above zero (which means the interest rate paid on bonds will also be above zero). But it is not correct to argue that the size of a sovereign government deficit affects the interest rate paid on securities.

Not understanding this, treasuries sometimes try to play the yield curve, issuing longer maturities when interest rates are low on them, or reversing course and issuing short maturities when the yield curve is steep. While it could be true that market forces of supply and demand enter into maturity spreads, if treasuries understood that the purpose of bond sales is to drain excess reserves so that the central bank can hit its overnight interest rate target, they would not issue long maturity debt at all.

Indeed, paying interest on reserves is an adequate substitute for treasury debt issue - as the overnight rate cannot fall below the interest rate on reserves.

## **Appendix: Advanced Material**

### **Monetary policy in the open economy, causes and consequences of capital flows**

Here we extend the analysis of central bank operations to the open economy. We must take account of international payments and possible effects on bank reserves. We will conclude that while this complicates the exposition, it does not significantly change the results.

#### *Sterilisation of capital flows*

The first issue explored is a practice called **sterilisation of capital flows**. The basic idea is that international payment flows can affect bank reserves. The question then becomes: should the central bank offset these impacts? That is, should it sterilise those impacts? If these flows increase reserves, should the central bank take the extra reserves out of the banks? If the flows reduce reserves, should the central bank restore reserves?

In the orthodox literature, this is presented as a choice. The fear is that if the flows increase reserves, and if the central bank does not move to take those extra reserves out, then the banks will increase domestic lending - creating money that could cause inflation. Hence the orthodox recommendation is that the flows should be sterilised.

Notwithstanding the observation that banks do not loan out reserves to customers, our response is that **sterilisation is more-or-less automatic**. As we'll see, this is because a central bank that targets overnight interest rates must accommodate reserve demand or it will miss its target. Let us turn to the detailed argument.

If international payments flows (or domestic fiscal actions) leave banks with excess reserves, the central bank has no choice but to drain the excess, unless it is willing to allow the overnight rate to fall towards zero or it pays an interest rate on reserves equal to the target interbank rate. As we

have learned, draining reserves is accomplished through open market bond sales, unwinding discount window lending, or sales of foreign currency reserves.

On the other hand, if international payment flows or domestic fiscal actions leave banks with insufficient reserves, overnight rates would rise above target - triggering the opposite interventions.

For this reason, sterilisation is not a discretionary operation. For example, China currently runs a large trade surplus with the US. Chinese importers want to convert their dollar-denominated receipts to yuan, an operation that is facilitated by the central bank when it buys dollars and creates yuan reserves.

If this leads to excess reserves in the Chinese banking system, the central bank then drains the excess through, for example, a sale of Chinese government debt. It cannot choose however, to leave excess reserves in the banking system, unless it is prepared to see the overnight interest rate fall toward zero. Any sterilisation of yuan reserves is automatic, a result of the interest rate targeting procedure.

#### *Government deficit finance and capital inflows*

It is sometimes claimed that a government's deficit spending as well as a nation's external position are constrained by the portfolio preferences of savers. For example, many believe that government faces a fiscal constraint, according to which its spending must be financed by a combination of tax revenues, bond sales (borrowing), or money creation. The form that financing of fiscal deficits takes is thus supposed to depend on the portfolio preferences of savers.

It is claimed that once households and firms have accepted all the new money desired, government must sell bonds, and the interest rate required to get the public to hold the bonds will be determined by their preferences. This supposedly applies even more forcefully to external constraints on national government fiscal deficits. For example, it is often erroneously claimed that the foreign sector (particularly, China) is financing the US fiscal deficit by lending US dollars to the government.

It is feared that once the ROW (Rest of World) has all the US government bonds that it desires, the US government won't be able to finance its deficit except at rising interest rates. Finally, it is argued that the ROW might even turn against the US dollar, refusing to hold them or US government debt, resulting in a financing crisis for the US and its government. Similar arguments have been made about other nations running external deficits.

This thinking reflects several different types of confusion. First, it conflates saving with portfolio allocation decisions. Second, it inappropriately equates the position of the issuer of the currency (the sovereign government) with the user of a currency (domestic households and firms plus foreigners). Third it applies an analysis that might be appropriate for a nation on a fixed exchange rate regime to a nation operating with a floating currency.

A sovereign government running a floating rate regime spends by crediting bank accounts, so the government fiscal constraint is nothing more than an *ex post* accounting identity (see Chapter 14). If a deficit results, government drains any excess reserves through bond sales as part of its interest rate targeting procedure. Again, a nation that pays interest on reserves never needs to sell bonds because the interest earning reserves serve the same purpose as interest-paying bonds.

The public makes its portfolio preferences apparent as excess reserves drive the overnight interest rate below the target rate, and will accept government bonds until all undesired reserves are drained. The demand for reserves is highly interest inelastic, but even if it were not, government can set the overnight rate at any positive level desired simply by ensuring that the banking system has no more reserves than it wants.

Whether the *ex post* fiscal identity will record a deficit after government increases its spending depends largely on the reaction of the other sectors. In other words, the government can decide how much it will increase spending and after the fact we will observe some combination of increased tax revenue, increased bonds held by the non-government sector, and increased high powered money holdings (reserves held by banks and cash held by the non-banking private sector).

The degree to which taxes rise will depend on the responsiveness of tax revenue to rising aggregate spending and income; the additions of bonds and high powered money to non-government portfolios will equal (by identity) the fiscal deficit, and the split between the two will depend on preferences for interest-earning assets, given the overnight interest rate set by central bank policy.

**The saving propensities of both the domestic and external sector go into determining the financial balances of all three sectors: the domestic private sector, the foreign sector, and the government sector.** Higher domestic private sector saving represents a leakage out of domestic income that is matched by some combination of a bigger government deficit and a smaller current account deficit. Higher ROW saving is matched by a combination of a larger US government deficit and greater US current account deficit.

We cannot observe saving or import propensities and our three-sector balances identity cannot tell us the complex causalities that lie behind the resulting balances. However, we should understand that **the fiscal outcome for a currency-issuing government is largely a residual**, rising when private domestic and foreign demand shrinks and falling when demand is rising. By the same token, a nation's current account deficit is largely a function of the ROW desire to spend.

Unfortunately, most analysts incorrectly interpret the causal forces involved, adopting a loanable funds approach according to which saving 'finances' investment, fiscal deficits, and current account deficits. Actually the causation is the reverse: it is the investment spending, the government spending, and the export spending that together creates the domestic saving of the private sector and the foreign saving in the form of the currency of issue. **Spending brings forth its own saving** as a result of income growth.

A moment's reflection about bank balance sheets will confirm that this must be true. A saver cannot simply ask their bank to credit their savings account with more dollars, but an investor can approach a bank for a loan, in which case the investor's deposit account is credited and this transaction is offset on the bank's balance sheet by the loan, which is the bank's asset. When the investor purchases plant and equipment, that deposit account is drawn down and a saver's account is credited.

Similarly, a foreigner cannot save more dollars until a local importer has purchased foreign output (or purchased foreign assets, including direct investment). Again, it is the importer's willingness to take out a loan to finance this purchase that results in a new dollar credit to the

account of the foreign saver. Hence, the notion that a nation is borrowing its local currency (for example, US dollars in the case of America) from abroad to finance government and trade deficits is erroneous. Rather, it is more revealing to think of the nation's fiscal deficit and the current account deficit as financing the ROW saving in that currency.

The decision to save is a decision to 'not spend'. For example, when the Japanese domestic sector taken as a whole produces more than its government and non-government sectors wish to purchase, it can save in financial form - but only if it can find external buyers so that it can export. Otherwise, saving takes the form of undesired inventory accumulation - which would then probably depress future production, employment, and income.

Let us assume Japan sells the excess production to Americans, in which case the savings are initially in US dollars. Portfolio decisions then come into play when savers decide how to hold the savings. Most of the US dollars will be exchanged for yen, used to purchase yen assets (financial and real). The Bank of Japan will usually facilitate this process as domestic banks offer US dollar reserves for yen reserves. As discussed above, if excess yen reserves result, these can be drained by the Bank of Japan so as to maintain a positive overnight interest rate. However, as Japan currently operates with a zero interest rate target, it leaves some excess reserves in the banking system.

In this situation, the portfolio decisions of foreigners (including, importantly, those decisions of ROW central banks) place no direct pressure on the US overnight interest rate. However, they can affect the exchange rate of the US dollar. It is commonly believed that a nation that runs a trade deficit must eventually see its currency depreciate in foreign exchange markets, although it is well recognised that empirical studies have not been able to systematically link exchange rates to the usual set of variables thought to be important determinants of exchange rates, including the trade balance.

In any case, this is a separate issue from the concerns with interest rate setting by the central bank or 'financing' of external and fiscal deficits. A country with a sovereign currency on a floating exchange rate can set its policy interest rate at any level desired, and can run fiscal deficits at any level desired, without worrying about impacts of foreign saving propensities or portfolio preferences on 'financing'. The country might, if desired, adjust interest rates or fiscal policy in response to actual or supposed pressure on exchange rates, but that is again, a separate issue from 'financing'.

We conclude: allowing for open economy impacts does not change our results. Central banks will sterilise any impacts on banking system reserves in order to hit overnight interest rate targets. Foreign portfolio preferences can impact current account outcomes - if the ROW wants US and Australian dollar assets, it exports output to the US and to Australia. In that case, the US and Australia might record current account deficits and the respective currencies will flow out to foreigners. On the other side of the coin, they will record capital account surpluses as dollars (US and Australian, respectively, flow in).

## References

Federal Open Market Committee (FOMC) (2016) “Statement on Longer-Run Goals and Monetary Policy Strategy” accessed 19 February, 2016 from [http://www.federalreserve.gov/monetarypolicy/files/FOMC\\_LongerRunGoals\\_20160126.pdf](http://www.federalreserve.gov/monetarypolicy/files/FOMC_LongerRunGoals_20160126.pdf).

Lavoie, M. (2013). “The Monetary and Fiscal Nexus of Neo-Chartalism: A Friendly Critique.” *Journal of Economic Issues*, 47 (1), 1–31.

Ruml, B. (1946a) “Taxes for Revenue are Obsolete”, *American Affairs*, Winter Number, Vol VIII, No. 1, available at: [http://www.constitution.org/tax/us-ic/cmt/ruml\\_obsolete.pdf](http://www.constitution.org/tax/us-ic/cmt/ruml_obsolete.pdf).

Ruml, B. (1946b) “Tax Policies for Prosperity”, *The Journal of Finance*, Vol. 1, Issue 1, p81-90.

Ruml, B. (1950) “The New Economic Insight”, *American Affairs*, XII(2), p90-93.

Tymoigne, E., and Wray, L.R. (2013) “Modern Monetary Theory 101: A Reply to Critics.” Levy Economics Institute of Bard College, Working Paper no. 778, November.

# Chapter 14: Fiscal Policy in Sovereign Nations

## Chapter Outline

- 14.1 Introduction
- 14.2 Functional Finance versus Sound Finance
  - The fiscal constraint and the views of deficit hawks, doves, and owls
  - Functional finance
- 14.3 Fiscal Policy Debates: Crowding Out and (Hyper)Inflation
  - Crowding out?
  - Competing theories of inflation
  - Inflation and sovereign fiscal policy
  - Alternative explanations of hyperinflation
  - Real world hyperinflations
  - Conclusion
- 14.4 The Debt Sustainability Debate
- 14.5 Conclusion: MMT and Fiscal Policy

## Learning Objectives

1. Comprehend the competing perspectives about the conduct of fiscal policy (hawks, doves and owls).
- 2 Understand the debates about fiscal policy as the alleged cause of crowding out and (hyper)inflation.
- 3 Recognise that a government operating with a sovereign currency will never face a crisis associated with public debt sustainability.



## 14.1 Introduction

In the previous chapter, we provide a detailed account as to how fiscal policy is conducted and the role of the central bank in interest rate (or liquidity) management through buying treasury debt from or selling it to the private banks.

In this chapter we continue our analysis of fiscal policy in sovereign currency issuing nations, by first contrasting the MMT view that the implementation of fiscal policy should be based on Abba Lerner's **functional finance principles** with the orthodox perspective which is based on the adoption of **sound finance principles**. We will see that the orthodox arguments about constraints on government spending follow from an inappropriate application of the **household budget constraint** to a sovereign government. This analysis then leads to an important policy recommendation that government can and should design its program of spending and taxation with the objective of achieving and maintaining full employment.

We then turn to key debates, which arise from this central role for fiscal policy. We briefly explore the claim that expansionary fiscal policy crowds out private sector spending. In an earlier chapter we discussed in detail a program that can achieve full employment - the Job Guarantee. The great fear is that full employment will cause inflation and some orthodox economists claim that there is the possibility of hyperinflation. We acknowledge that in the absence of appropriate oversight, a government can maintain an excessive rate of expenditure, which leads to rising inflation. Two examples, the Weimar Republic and Zimbabwe, are provided and it shown that supply constraints played a major role in the emergence of hyperinflation.

Finally we use a technical exposition to explore the claim that reliance on fiscal policy to sustain full employment is likely to lead to a rising treasury debt to GDP ratio, which could be unsustainable.

## 14.2 Functional Finance versus Sound Finance

### The fiscal constraint and the views of deficit hawks, doves, and owls

Following the GFC, many developed economies experienced a rise in their fiscal deficit to GDP ratios, due to the operation of automatic stabilisers, following the collapse of non-government expenditure. These fiscal deficits translated into rising debt to GDP ratios. Since 2009 multi-national agencies including the OECD and IMF, have preached the principles of sound finance through the adoption of austerity measures. By austerity measures we mean policies of cutting government expenditure and/or raising taxes to reduce the fiscal deficit. This represents the typical orthodox position on fiscal policy, with the **government fiscal constraint** viewed as an analogue to the household budget constraint derived from neoclassical microeconomic theory. The interpretation of the so-called constraint on government net spending is based on the premise that the government has three sources of finance for its spending as shown on the right hand side of the following identity:

$$(14.1) \quad G + iB \equiv T + \Delta B + \Delta M_h$$

where  $G$  stands for government spending;  $iB$  are the interest payments on existing public debt;  $T$  is tax revenue,  $\Delta B$  is new borrowing based on selling government bonds, and  $\Delta M_h$  is new money (HPM) creation. This relationship was introduced in the previous chapter and linked the

acquisition or loss of net financial assets of the domestic private sector to the government sectoral balance in a closed economy.

According to the orthodox interpretation of the **government fiscal constraint**, if the government runs a fiscal deficit, then it will have to borrow ( $\Delta B$ ) by selling public debt and/or create additional high-powered money ( $\Delta M_h$ ). Thus the GFC is alleged to represent an *ex ante* constraint on government spending. In other words, if the fiscal outcome was known in advance, then Equation (14.1) would be a guide to how it could be financed.

Few economists argue that government must or even should continuously achieve fiscal balance, although there are occasionally some politicians and other extremists who want to legislate such a requirement. We can distinguish three different perspectives as to the appropriate fiscal strategy: Deficit Hawks, Deficit Doves and Deficit Owls (the distinction is credited to Stephanie Kelton, a professor at UMKC in the USA).

Deficit Hawks recommend that government strive to achieve fiscal balance, even though most recognise that it is hard to match exactly revenues and expenditures over the course of a year. Hence deviations from fiscal balance will occur, but government should always respond to imbalances. Thus, if a deficit occurs one year, government should try to run a surplus the following year - by cutting spending and raising taxes.

Deficit Doves believe government should aim to achieve fiscal balance over the course of a business cycle, but should run deficits in recessions and surpluses in expansions. Hence, a government should be willing to use its fiscal capacity as a counter-cyclical policy tool to offset swings of private sector spending. For example, Deficit Doves argued for deficits to stimulate the slumping economies of the major Western nations. In their view, the time to move towards fiscal balance would come only after a robust recovery had got underway, and tax revenues had started to increase.

Deficit Owls take an entirely different position, based on **Functional Finance** principles. For them, the fiscal outcome for a sovereign government is not a useful target for policy making. It is not **functional** in the sense of providing policy guidance. Rather, policy ought to target important economic goals such as full employment, price stability, poverty alleviation, environmental sustainability, and the overall standard of living. **The Deficit Owl perspective is the only one that is consistent with MMT.** Let us see why.

As we know from previous chapters, the sovereign government spends by issuing currency, which today is mostly through electronic entries on balance sheets. Taxes lead to debits of those entries. Logically, the spending must precede the taxing since accounts cannot be debited before they are credited.

MMT sees the government fiscal outcome as an *ex post* identity. At the end of a year, it will certainly be true that government spending over the year is equal to tax revenues plus net bonds issued plus net high powered money issued, as shown in Equation (14.1). This is simply an accounting identity that must hold. Equally important, MMT does not see  $T$ ,  $\Delta M$  and  $\Delta B$  as alternative methods of financing government spending, but rather as different parts of the process of conducting fiscal policy, as described in the previous chapter, and illustrated by the numerical example. Spending begins with crediting private bank accounts. The payment of taxes leads to private bank accounts being debited. Then, if government spending is greater than taxes there is a net crediting of reserve accounts at the central bank ( $\Delta M_h > 0$ ).

Normally the reserves created will be greater than what banks need to hold whether or not there are legal reserve requirements. Banks with excess reserves at the central bank will try to lend them in the interbank overnight lending market. However, the overall banking system has excess reserves, so there will be no demand for them.

The excess supply of reserves would cause the overnight interbank lending rate to fall. Once it has fallen below the target level (range), the central bank would respond by selling bonds, that is an open market sale. However, in normal periods central banks have a limited supply of government bonds - they can only sell bonds that they have previously bought. So in the presence of budget deficits, the central bank would need the treasury to create and sell more bonds - in the new issue market. Central banks and treasuries coordinate their operations to ensure that fiscal operations have minimal undesired impacts on banking system reserves. Hence bonds will be issued more-or-less in step with budget deficits in order to drain excess reserves from the banking system.

At the end of the year, we would find that government spending less taxes will be equal to the change in HPM, that is the change to banking system reserves plus the change to private sector holding of cash, and in addition, the change to private sector holding of government bonds. As noted, in normal times, the growth of banking system reserves is quite small. Thus, the deficit is usually approximately equal to  $\Delta B$ . The growth of cash held is likewise fairly small, and linked more-or-less closely to growth of national income.

On the other hand, let us presume that a sovereign government spent by crediting banks with reserves but chose to leave excess reserves in the banking system. This could happen, for example, if government adopted a zero interest rate target - in that case excess reserves would drive the overnight interbank lending rate to zero and government would not need to do anything. We would then see that  $G + iB - T = \Delta M_h$ . But the difference between this case and the more usual case discussed above, where  $G - T$  is approximately  $\Delta B$  has nothing to do with the way government chose to 'finance' its spending. In both cases, government spent by crediting bank accounts. The different outcome is due to the choice to either drain excess reserves or to leave excess reserves in the system. That depends on whether government wants to target a positive overnight interest rate, or a zero overnight interest rate. Most economists see that as a monetary policy decision, and not a fiscal policy decision. In the post-GFC period the level of bank reserves significantly increased in a number of developed economies, including the USA, UK and Australia. This reflected banks becoming more risk averse in these countries, but also the desire of central banks in the USA and UK to both maintain a low overnight rate, but also flatten the yield curve via quantitative easing (see Chapter 15).

We conclude that the **Government Fiscal Constraint** is neither a constraint nor does it present alternative ways of financing government spending. Rather it is an *ex post* accounting identity whose outcome is determined by decisions made by households, firms, financial institutions, the central bank, and even foreigners. Households, firms, and foreigners decide how much cash they want to hold. Banks (and the central bank through required reserve ratios in the USA), determine the holding of reserves. The central bank decides whether the overnight rate target will be above zero. All of those decisions go into determining the split between  $\Delta B$  and  $\Delta M_h$ .

That is not an *ex ante* decision of treasury to either **borrow** or **print money**. Indeed, treasury cannot decide *ex ante* what the fiscal outcome will be (fiscal balance, deficit, or surplus) since that depends on tax revenue generated over the course of the upcoming year, plus unplanned

spending linked to unforeseen events and the impact of automatic stabilisers. Equation (14.1) is useless for planning purposes, and has no explanatory power *ex post* in terms of explaining the composition of the right hand side.

### Functional finance

American economist Abba Lerner wrote two important articles in the 1940s. By a happy coincidence he was a professor at the University of Kansas City, which became the University of Missouri at Kansas City, one of the places where MMT has been developed. One of these articles proclaimed that “money is a creature of the state” (Lerner, 1947: 313). Obviously that is also the position of MMT as outlined in this textbook: the state chooses a money of account, imposes liabilities in that unit, and issues currency denominated in the same unit that are accepted in payment of taxes. All of this was understood by Lerner. In his article on functional finance, he calls it the new fiscal theory. He says that like any new theory it seems extremely simple and it is that simplicity that makes people suspicious.

Lerner wrote (1943: 39):

The central idea is that government fiscal policy, its spending and taxing, its borrowing and repayment of loans, its issue of new money, and its withdrawal of money, shall all be undertaken with an eye only to the results of these actions on the economy and not to any established traditional doctrine about what is sound or unsound.

He goes on to outline two principles of functional finance:

1. The first financial responsibility of the government (since nobody else can undertake that responsibility) is to keep the total rate of spending in the country on goods and services neither greater nor less than that rate which at the current prices would buy all the goods that it is possible to produce. (Lerner, 1943: 39)

When spending is too high, the government is to reduce spending and raise taxes; when spending is too low, the government should increase spending and lower taxes.

2. An interesting corollary is that taxing is never to be undertaken merely because the government needs to make money payments... Taxation should therefore be imposed only when it is desirable that the taxpayers shall have less money to spend. (Lerner, 1943: 40)

If the government is not to use taxes to ‘make money payments’, then how are these to be made? According to Lerner, the government should not turn to borrowing for the purposes of spending, because, “The second law of Functional Finance is that the government should borrow money only if it is desirable that the public should have less money and more government bonds” (Lerner, 1943:40). In other words, the purpose of taxes and bonds is not really to finance spending as each serves a different purpose (taxes remove excessive private income while bonds offer an interest-earning alternative to money). Instead, the government should meet its needs “by printing new money” (p.41) whenever the first and second principles of functional finance dictate that neither taxes nor bond sales are required. That is, as discussed above, the choice over whether to leave high powered money (mostly reserves) in the system or to drain it through bond sales depends on what we normally call a monetary policy decision (interest rate policy).

In summary, Lerner argued (1943: 41):

Functional Finance rejects completely the traditional doctrines of ‘sound finance’ and the principle of trying to balance the budget over a solar year or any other arbitrary period. In their place it prescribes: first, the adjustment of total spending (by everybody in the economy, including the government) in order to eliminate both unemployment and inflation, using government spending when total spending is too low and taxation when total spending is too high; second, the adjustment of public holdings of money and of government bonds, by government borrowing or debt repayment, in order to achieve the rate of interest which results in the most desirable level of investment; and third, the printing, hoarding or destruction of money as needed for carrying out the first two parts of the program.

He concluded that functional finance “is applicable to any society in which money is used as an important element in the economic mechanism” (Lerner, 1943: 50). In this textbook we want to narrow the application of functional finance somewhat to the sovereign government - which is consistent with the views Lerner advanced in his other great article that proclaimed “money is a creature of the state”.

In that piece, Lerner insisted that (1947: 313):

[W]hatever may have been the history of gold, at the present time, in a normally well-working economy, money is a creature of the state. Its general acceptability, which is its all-important attribute, stands or falls by its acceptability by the state.

Just how does the state demonstrate acceptability?

The modern state can make anything it chooses generally acceptable as money ... It is true that a simple declaration that such and such is money will not do, even if backed by the most convincing constitutional evidence of the state’s absolute sovereignty. But if the state is willing to accept the proposed money in payment of taxes and other obligations to itself the trick is done. Everyone who has obligations to the state will be willing to accept the pieces of paper with which he can settle the obligations, and all other people will be willing to accept these pieces of paper because they know that the taxpayers, etc., will accept them in turn (Lerner, 1947: 313).

This seems to be a clear exposition of what we now call the MMT, **taxes drive sovereign currency** view: even if it has not always been the case, it surely is now true and obvious that the state writes the ‘description’ of money when it denominates the tax liability in a money of account, and defines the ‘thing’ that ‘answers to the description’ when it decides what will be accepted at tax offices. The ‘thing’ which answers to the ‘description’ is widely accepted not because of sovereignty alone, not because of legal tender laws and not because it might have (or have had) gold backing, but because the state has the power to impose and enforce tax liabilities and because it has the right to choose ‘that which is necessary to pay taxes’.

This right, as emphasised by Keynes, “has been so claimed for some four thousand years at least” (Keynes, 1930: 3). While Keynes is no historian and while one might quibble over the exact number of years since states first claimed these rights, there can be no doubt that all modern states do have these rights. As Lerner said “Cigarette money and foreign money can come into wide use only when the normal money and the economy in general is in a state of chaos” (Lerner, 1947: 313).

One might only add that when the state is in crisis and loses legitimacy, and in particular loses its power to impose and enforce tax liabilities, ‘normal money’ will be in a ‘state of chaos’, leading, for example, to use of foreign currencies in private domestic transactions. In all other cases, it is state money, which is used, and state money that the state accepts in payment of taxes.

### 14.3 Fiscal Policy Debates: Crowding Out and (Hyper)Inflation

#### Crowding out?

Many pundits wrongly believe that when central banks issues treasury debt, it can **crowd out** private borrowing and hence spending. The idea is that there is a limited supply of private sector saving for which government **borrowing** and private sector borrowing compete. If government tries to borrow more, by issuing and selling more bonds, then the competition for finance would push up interest rates. Some private firms would decide not to borrow at the higher rates and hence investment would be lower and also durable consumption expenditure by households, some of which is financed by borrowing.

That, however, is incorrect. Government budget deficits generate non-government surpluses (flows) that accumulate to the non-government sector net acquisition of financial assets (a stock) - as we have learned in this and earlier chapters. Since there are more savings and greater financial wealth, it is not true that government is competing with private sector borrowers for a limited supply of savings flows to place government bonds into wealth portfolios that are fixed in size. Both savings and portfolios expand as government deficits grow.

Further, as we have seen, the central bank department of government sets the overnight interest rate. It can keep that low no matter how big the government deficits are. This is why, for example, that Japan has kept a near-zero overnight interest rate (and also very low rates on its long maturity treasury debt) ever since its economy collapsed, leading to the biggest fiscal deficits in the developed world (see the earlier discussion of Japan’s deficit, debt and interest rate outcomes in Chapter 2). Similarly, the Fed has kept US interest rates low since the Global Financial Crisis, in spite of budget deficits that rose to 10 percent of GDP. Finally, all through WWII the Fed also kept rates near zero as budget deficits reached 25 per cent of GDP. What all this means is that there is no reason to expect fiscal deficits to push up interest rates - since interest rates (at least at the short end of the maturity structure) are policy-determined.

For these and other reasons, the **crowding out** argument against fiscal deficits is not based on a coherent understanding of operational realities or of the empirical data. The notion that government bonds compete with a fixed supply of saving must be rejected. It would be closer to the truth to assume that the demand is perfectly elastic for sovereign government securities issued when government runs deficits.

Indeed, primary dealers finance their purchases of bonds at auction in the repo market, mostly using treasury bonds as collateral, while the newly issued bonds will likely serve as collateral for further credit creation in financial markets. Far from crowding out, bonds can actually enable more private credit creation than would occur in their absence.

This sequence of arguments is known as the loanable funds theory. In summary, adding the rule that the treasury must finance its operations in the open market, which we discussed in the previous chapter adds complexity compared to the relatively simpler operations that would be required to let deficit spending proceed in the absence of the two rules. Nevertheless, the nature

of these operations as described by the general case of a consolidated government balance sheet is more-or-less the same.

The generic case of net spending of the sovereign currency by treasury, which was illustrated by the numerical example, is a reasonable representation of the true nature of government debt operations. The sovereign currency issuer cannot run out of its own currency, and we can summarise its spending operation as a **keystroke** creation of high powered money when it spends as a simplification that is not too misleading.

### **Competing theories of inflation**

In Chapter 11, various theories of inflation which have underpinned major macroeconomic policy debates over the last 50 years or so were outlined and assessed. The analysis commenced with the Quantity Theory of Money and the Keynesian Phillips curve trade-off and then followed by the Expectations Augmented Phillips Curve, developed by Friedman and Phelps, and concluded with the hysteresis based theory of inflation, which restored a trade-off between inflation and labour underutilisation.

In Chapter 12, we argued that there was a fundamental policy choice between the adoption of an unemployment buffer stock to sanction the inflationary process and the adoption of a fixed money wage, employment buffer stock which is the basis of the JG proposal. We developed the argument that full employment was the primary policy goal and after consideration of other strategies which were designed to achieve full employment, concluded that a JG should be adopted.

We now again focus on the hostility of orthodoxy to the active, discretionary use of fiscal policy and specifically the MMT view that there are no financing constraints on the use of fiscal policy, given that treasury spends a fiat currency, which by definition is not backed by a valuable commodity, through keystrokes.

Many argue that it is the adoption of fiat money that causes inflation. If only the nation's money were tied to something with real value (like gold), that would allow money to retain its value so that prices would not rise. On the other hand, MMT argues that money's value is not, and never has been, determined by a commodity like gold. Rather, money is the unit of account in which debts and credits are denominated. We can think of money as entries on balance sheets. Critics of MMT react in horror at such a suggestion, because they believe that MMT's claim that government spends through keystrokes is a recipe for inflation, if not hyperinflation. The following section will allay those fears.

### **Inflation and sovereign fiscal policy**

It is important to recall from Chapter 11 that **inflation is defined as a persistent rise in prices at the aggregate level**. This occurred in the early 1970s following the **energy shock** which was set off by geo-political forces. Energy prices quadrupled over a short period of time, and most prices quickly rose because energy is used to produce and/or transport most goods and services. In the presence of strong trade unions in many countries of the developed world, who were seeking to restore their real wages, an inflationary process commenced, which was typically met by contractionary macroeconomic policy and the emergence of stagflation. By contrast, the introduction of the Goods and Services Tax in Australia in mid-2000, which was compensated

for by income tax cuts, following negotiation with the trade union movement, led to a once off increase in a wide range of prices, but no inflationary process ensued.

As Keynes argued, the economy needs some 'stickiness' of wages and prices in the money of account - or people might abandon money. That is what can happen in a hyperinflation, so with money's value falling quickly people try to find a substitute that can maintain its value. However, inflation in the US, in Australia and in most developed economies has been sufficiently low that the domestic currency has remained a useful money of account, and the domestic currency has been voluntarily held in spite of inflation. Economists have been hard-pressed to find significant negative economic effects from inflation at rates of under 40 per cent per annum. But clearly people do not like inflation when it rises to double digits, given the almost inevitable loss of real income, and policymakers usually react to double digit inflation by adopting austerity programs in an effort to reduce aggregate demand.

The question is whether austerity is the right policy. If an economy is operating beyond full employment, then by Lerner's first principle of functional finance, government needs to dampen demand - by reducing spending or raising taxes. There are instances in a number of countries over the past half century in which demand probably became excessive, raising production beyond the full employment level. Major wars are the typical trigger for demand pull inflation. But in most developed countries, aggregate demand has been insufficient to generate inflationary pressure since WWII. Instead, inflation has often been accompanied by substantial unemployment (stagflation). The **misery index** that adds the inflation and unemployment rates together, which is really based on adding together apples and oranges, has resonated with some voters, commentators and politicians since the late 1970s.

Many critics think that when advocates of MMT state that **government spends its currency into existence** and that logically it **spends first, then taxes** and that **for sovereign government, affordability is not a question** they are making policy recommendations rather than describing a modern monetary economy. All major economies went off the gold standard a very long time ago. MMT's advice about spending simply follows the recommendations arising from Lerner's functional finance principles, namely spend more if employment and incomes are too low, and provide more reserves if banks are short of reserves and driving up overnight interest rates and compromising monetary policy.

Insufficient spending and employment can be readily deduced by observing jobless workers who cannot find jobs and factories operating below full capacity, but primarily the former. These observations indicate that effective demand is too low, which means that government can either cut taxes or increase spending to raise demand. A sovereign government that issues its own currency can always afford to do this. The question is how best to stimulate demand.

Friedman famously used the metaphor of helicopters flying around dropping money - arguing that is how inflation is caused (Friedman, 1969). In reality, if government really did inject money that way, it would be a form of fiscal policy via **transfer spending** that is the somewhat arbitrary distribution of welfare or social security payments, rather than monetary policy (see the next chapter).

Textbook Keynesian fiscal policy is similar to this, and is often likened to **pump-priming** stimulus. This type of fiscal policy attempts to indiscriminately raise government spending and thus private income to encourage consumption. In depressed economic conditions this makes sense because it is likely that across all sectors of the economy there is sufficient slack. That



allows for increased production and employment everywhere, with little pressure on prices or wages.

However, typically this is not the best form of expansionary fiscal policy, designed to achieve full employment. The problem is that normally some sectors have lots of slack, while others face relatively tight conditions. This is particularly true across the labour force since the demand for educated and highly skilled workers is typically tighter than that for workers with less experience, education and training. Unemployment is never equally shared, and the most disadvantaged workers generally experience higher unemployment. Trying to **pump up** aggregate demand may well cause bottlenecks in some sectors that lead to workers negotiating for higher wages which, in turn, drives up prices in those sectors, even though there is substantial slack in other sectors. That means that inflation can be generated long before reaching full utilisation of plant and equipment and full employment of workers across most sectors of the economy. As inflation rises, government can lose its nerve and abandon stimulus before full employment is achieved. In fact, given the lags in implementing fiscal policy, which were highlighted by Friedman, a commitment to the achievement of full employment may be undermined by fiscal policy not operating as a counter-cyclical measure and thus destabilising the macro economy. Friedman (1953) argued that discretionary fiscal policy was potentially destabilising because there were long and variable time lags associated with: (i) recognition of the need for discretionary fiscal policy, and (ii) the design and implementation of an appropriate policy response and (iii) the economy responding to the policy measures that were adopted.

Further since pump-priming entails the payment of market wages, the policymakers have limited scope for imposing a counter-inflation sanction, while maintaining a commitment to full employment. This tends to lead to a **stop-go** pattern of policy making.

What is needed, instead, is targeted policy - policy that directs additional demand creation where it is most needed. This is not as hard as it might sound. Government does not need to keep tabs on every single sector of the economy to fine-tune its stimulus where it is needed. As we have discussed throughout the textbook, full employment is the most important policy goal for the economy, particularly if the UN Human Rights agenda is to be followed (see Chapter 1), so it is far more important to keep humans fully employed than it is to make sure that capacity is fully utilised.

In the chapter on the Job Guarantee (JG) proposal, we noted that government can direct its spending where it is most needed - directly to the unemployed - by offering a job to anyone who needs one. Since government spending will automatically increase when it is needed (that is, when the number of job seekers rises) and fall when it is not needed (when workers leave the program for alternative work), the policy operates counter cyclically and only provides the amount of stimulus needed. It is true that some factories and some retail outlets will still find themselves with excess capacity or with insufficient capacity. However, these will react to market signals by either cutting capacity or increasing it, as necessary. Clearly there are always winners and losers in a market based economy, but the private sector in total performs better when responding to market signals in a fully employed economy, which is achieved by putting into place the JG program.

The fear that 'fiat money' is not backed by a real commodity, such as gold, reflects a belief that if government can buy anything for sale merely by **keystrokes** then it will try to buy everything. The 2013 shutdown of the US government because the Republican party refused to raise the debt

limit so that the government could spend, would imply that real world politics does not work that way. There is ample evidence that, at least in nominally democratic countries, governments exercise too much restraint in their spending and taxation measures. Usually the elected representatives haggle over how much should be spent, and on what, and then the head of the administration (the President in the case of the USA) decides whether to approve a budget which then provides the spending authority. In the UK and Australia legislation needs to be passed by the Parliament before the spending and taxation measures are implemented.

The spending and tax legislation does not determine what *ex post* spending and tax revenues will be since these depend on spending by the non-government sector, due to the operation of automatic stabilisers. In a recession, more is spent on unemployment compensation; in an expansion, revenue might exceed projections. However, budgets are framed in a way that imposes quantitative spending rules on treasuries (see Chapter 12), rather than enabling governments to spend without constraint.

A government that decides to keep spending and raising the price it is willing to pay to purchase resources and output undoubtedly will cause high inflation. There is no substitute for good governance. An unaccountable government will not be constrained in its fiscal strategies either by the political process or by a gold standard. However, it is important to understand the cause of this inflationary process.

Monetarists are hostile to the creation of HPM to finance deficits, because they claim it is inflationary due to the Quantity Theory of Money (QTM). MMT advocates would first highlight institutional practice, namely that net treasury spending initially causes to an equal increase in HPM, which we have emphasised. Second, they would challenge the theory of inflation based on QTM, and argue that if a fiscal deficit gives rise to demand pull inflation, then the *ex post* composition of  $\Delta B + \Delta M$  in Equation (14.1) is irrelevant. Overall spending in the economy is the driver of the inflationary process.

In Chapter 2, we documented Japanese fiscal deficit and debt ratios, interest rates and inflation rates since at least 1990. Despite persistent deficits since the earlier 1990s, interest rates have remained low (as we discussed earlier in the Chapter) and annual inflation rates have only once exceeded 2 percent since 1996, and indeed in numerous years there has been deflation that is negative inflation. The Japanese unemployment rate has remained between 3.5 per cent and 5.5 percent since 1996. Thus it is not deficit spending and the way it is financed that impact on inflation, but rather total spending. With significant excess capacity over this period, there has been plenty of scope for stimulatory fiscal policy without promoting demand pull inflation.

Let us now look at experiences with much higher inflation rates than those usually encountered in developed nations. We are referring to rates so high that they do harm economies. We will see that extremely high inflation is unusual. Further, there appears to be no reason to believe that the sort of creeping inflation that is common will gradually rise to hyperinflationary rates.

Still, critics will point to specific examples - either to the pre-WWII Weimar Republic, or to the more recent experience in Zimbabwe, where it is claimed that government did spend without restraint, which destroyed the value of the currency. We now turn to the analysis of hyperinflation.

## Alternative explanations of hyperinflation

Many fear that if a government operates along MMT lines, then we are on the path to ruinous hyperinflation. Indeed, MMTers are commonly accused of promoting policy that would recreate the experiences of the Weimar Republic or Zimbabwe hyperinflations. These were supposedly caused by governments that resorted to **money printing** to finance burgeoning deficits - increasing the money supply at such a rapid pace that inflation accelerated to truly monumental rates.

In his classic 1956 paper Phillip Cagan defined hyperinflation as an inflation rate of 50 per cent or more per month. The most popular explanation of hyperinflation is the Monetarist quantity theory of money: the government engages in excessive money creation, causing prices to rise. However, as prices rise, the velocity of circulation increases, because no one wants to hold onto money very long as its value falls rapidly. Wage increases are demanded daily, so that goods and services may be purchased because tomorrow the same level of wages will purchase less than today. Even though the money supply grows as rapidly as government can print notes, it never keeps up with rising prices. The faster that prices rise, the higher velocity climbs and eventually workers demand hourly payment and run to the stores at lunchtime because by dinner time prices will be even higher.

Essentially, the above was Cagan's explanation for the fact that a simple version of the quantity theory did not fit the data: if prices rise so much faster than the money supply, how can we conclude that the hyperinflation is caused by 'too much money chasing too few goods'? To fit the facts, the quantity theory was revised to state that in a high inflation environment, the old quantity theory presumption that velocity is stable (which is necessary to maintain a link between money and prices) no longer holds.

So armed with the revised quantity theory, we can still claim that high inflation and hyperinflation result from too much money even though velocity rises, when money growth lags behind the inflation rate. Monetarists claim that government controls the money supply, and hence hyperinflation must be due to government policy. In addition, in hyperinflationary periods, the supply of government currency (paper notes) rises rapidly (with extra zeroes added). Finally, government runs deficits when it finds its tax revenue cannot keep up with its spending, so it is alleged to frantically print money to make up the difference - and that adds to the **too much money chasing too few goods**.

So MMT's critics argue that most of the blame for hyperinflation falls on money printing by governments to finance deficits. There are parallels here to the US, UK, and Japanese situation in 2012, when large fiscal deficits (plus quantitative easing) led to a significant increase in the stocks of reserves held by the banks. Consequently, the banks are alleged to have increased credit creation (lending) which led to an increase in the money supply and the price level. Let's take a look at an MMT response to these explanations of hyperinflation. We will make three points.

- As discussed above, when MMT says that government spends by **keystrokes**, this is a description, not a prescription. If critics were correct that government spending by **printing money** necessarily leads to high inflation or hyperinflation, then most developed nations would have at least high inflation, if not hyperinflation all the time because they all spend by keystrokes. Logically, all governments that issue their own currency have to spend it before they can collect it in taxes (or engage in bond sales) -

since no one else can create it. There is no alternative way for these governments to spend. Even if they promise to convert at a fixed exchange rate, they still spend by keystrokes. Yet there is no evidence of hyperinflation or high inflation in developed economies over the last 20 years. This suggests that to claim a causal relationship between **printing money** and hyperinflation is highly problematic.

- Particular cases such as the Weimar Republic or Zimbabwe need very careful scrutiny. Hyperinflations are caused by quite specific circumstances, although there are some shared characteristics of countries and monetary regimes that experience hyperinflation. While causes can be complex and varied, the Monetarist explanation sheds almost no light on the experience.
- Despite higher deficit ratios in the US, UK and Japan since the GFC, and higher stocks of bank reserves, which were also caused also by quantitative easing and the interest rate policies of the central banks, these countries have not endured hyperinflation or even high inflation. Also there is no evidence that this will occur in the foreseeable future.

Most critics of MMT and of so-called fiat money appear to imagine a past in which money was closely tied to a commodity like gold, which constrained the ability of both government and banks to create money **out of thin air**. The best example was the precious metal coin that supposedly gave a real value to government money, and forced government to actually get gold in order to spend. A strict gold standard with 100 per cent gold backing against paper notes (issued by government or banks) accomplished the same task.

These critics would also advocate the formalisation of constraints on fiscal policy, in the form of balanced budget amendments, debt limits, or for deficit doves a commitment to eventually slash deficit spending once the recovery gets underway. These already exist in the USA. In 2015 the UK Parliament passed a bill requiring the UK Treasury to achieve fiscal surpluses after 2019-2020, when GDP growth exceeded 1 percent, subject to the absence of spending shocks.

As we have argued, a floating exchange rate provides policy space that can be used by prudent governments with their own sovereign currencies to pursue domestic policy goals with a greater degree of freedom. Except for the losers of WWI (plus Poland and Russia, which were on the winning side but left the capitalist world), there are no cases of nominally democratic Western capitalist countries that have experienced hyperinflation in the past century. And if we limit our data set to those with floating currencies, there are no countries with exchange rate crises, either (see Chapter 16).

It is only countries with fixed exchange rates or other promises to deliver foreign currency or gold (such as debts in foreign currencies) that have hyperinflations and currency crises. This tends to result from the imprudent expansion of these IOUs relative to ability to actually deliver the foreign currency or gold. While it appears that the fixed exchange rate guarantees prudence, this not achieved in practice. The fixed exchange rate introduces exchange rate crises plus involuntary default as possibilities, and does not guarantee that government will be prudent.

When sovereign government promises to deliver foreign currency it actually exposes the nation to Weimar Republic hyperinflationary risks. This risk can be compounded if their banks are not necessarily prudent. We could for example look at the recent case of Ireland. While, prior to the GFC, the Irish government appeared to be the paragon of fiscal prudence, its banks lent excessively in foreign currency (the Euro). Further the government was active in promoting the

construction industry through subsidies and became over-reliant on stamp duty from sales of houses to generate tax revenue. When borrowers defaulted, the Irish government took on all the foreign currency debt - leading to a sovereign government debt crisis because it had adopted the Euro.

### Real world hyperinflations

High inflation and hyperinflation are rare events. In this section we look at historical examples of hyperinflation periods. Hyperinflations are caused by quite specific circumstances, although there are some shared characteristics of countries and monetary regimes that experience hyperinflation. The simple **printing money** to finance excessive fiscal deficits explanation sheds almost no light on the experience.

Today, the best known cases of hyperinflation occurred during the Weimar Republic and more recently in Zimbabwe. (Less well known but more spectacular was the Hungarian hyperinflation. The best analyses of these are at Mitchell (2010; 2011). We will not reproduce his analysis, but will summarise key points about the Weimar and Zimbabwe hyperinflations to assure readers that these were not simple cases of too much 'money printing' to finance government spending that was 'running amuck'.

The typical story about Weimar Germany is that the government began to freely print a fiat money with no gold standing behind it, without regard for the hyperinflationary consequences. The reality is more complex. First, we must understand that even in the early 20th century, most governments spent by issuing IOUs - although many were convertible on demand to UK Sterling or gold. Germany had lost WWI and suffered under the burden of impossibly large reparations payments that had to be made in gold, but it had very limited gold reserves. To make matters worse, much of its productive capacity had been destroyed or captured. Germany was supposed to export to earn the gold needed to make the payments demanded by the victors. Keynes (1920) wrote his first globally famous book, *The Economic Consequences of the Peace* in which he argued that Germany could not possibly pay the external debts denominated essentially in gold.

The nation's productive capacity was not even sufficient to satisfy domestic demand, and so was unable to export sufficient goods to pay huge reparations. The government believed that it was politically impossible to raise taxes to a sufficient level to match the value of resources needed for exports to pay the reparations. Instead it relied on issuing domestic currency in excess of taxes. This meant government competed with domestic demand for a limited supply of output - driving prices up. At the same time, Germany's domestic producers had to borrow abroad (in foreign currency) to buy imports that they needed. Rising prices plus foreign borrowing caused a depreciation of the domestic currency, which then increased the need to borrow (since foreign imports cost more in terms of domestic currency) and at the same time increased the cost of the reparations in terms of domestic currency.

While it is often claimed that the Weimar central bank contributed to the inflation by purchasing debt from the treasury, actually it operated much like central banks do today: it bought government debt from banks - offering them a higher earning asset in exchange for reserves. Fiscal deficits grew rapidly from the high inflation and then hyper-inflation as tax revenue could not keep pace with rising prices.

Finally in 1924 Germany adopted a new currency, and while it was not legal tender, it was designated acceptable for tax payment. The hyperinflation ended. It is evident that the major

supply constraints in its post-war plus the obligation to make substantial reparations were key factors in the Weimar's hyperinflation.

Let us turn to Zimbabwe. This country was going through tremendous social and political upheaval, with unemployment reaching 80 per cent of the workforce and its GDP had fallen by 40 per cent. This followed controversial land reforms that subdivided farms and led to the collapse of food production. Government had to rely on food imports and IMF lending, which provides another example of a country taking on external debts in a foreign currency. With food scarcity and government and the private sector competing for a much reduced supply, prices were pushed up.

This was also another case in which government could not have raised taxes, for both political and economic reasons. Again, to label this a simple Monetarist case of government **printing money** really sheds no light on Zimbabwe's problems, which were caused mostly by social unrest, the collapse of agriculture and hence major supply shortages, and heavy external debt.

## Conclusion

It is important to acknowledge that greater constraints on government spending (or a greater capacity to increase taxes) might have successfully prevented hyperinflation in these cases. However, when specific cases of hyperinflation are studied, it is evident that it is not a simple story of a government adopting a fiat money and printing and spending too much. There are many paths to hyperinflation, but there are common problems: social and political upheaval; civil war; the collapse of productive capacity that could be due to war; weak government; and foreign debt denominated in external currency or gold. In these circumstances we do observe rising budget deficits and (by identity) growing outstanding government IOUs. But we also find banks creating money to finance private spending that competes with government to drive up prices.

So it is likely that tighter fiscal policy would have helped to reduce inflationary pressures. This probably would not have reduced overall suffering, since a common cause of hyperinflation is some kind of supply constraint on output. But the solution to the problems does not require adoption of a gold standard. Rather, to tackle a problem of high inflation policy makers should try to slow down the indexing of wages and transfers, stabilise production, reduce demand relative to supply, and quell social unrest. When high inflation has persisted for some time, it also helps to adopt a new currency and to default on external debts.

In conclusion, there is a link between high (or hyper) inflation, budget deficits, and money supply, although it is not a simple Monetarist dynamic. A government always spends by **keystrokes** that credit accounts, and taxes (or sells bonds) by reverse keystrokes that debit accounts. Deficits mean government credited more to accounts than it debited, so that government IOUs have been net created in the form of high powered money (HPM that is, reserves plus cash) and treasuries (bonds and bills). As discussed above, in high or hyperinflation periods, taxes (debits to accounts) grow more slowly than government spending (credits to accounts) so we expect deficits to result, which means government IOUs outstanding (HPM plus treasuries) grow.

Matters are made worse if a high interest rate policy is pursued by the central bank. This is because government typically matches its rising deficit with new government bond (debt) issues and interest payments add to government spending. If the central bank reacts to growing deficits by raising interest rate targets, it helps to fuel growth of the deficit and also adds demand

stimulus to the economy in the form of interest payments by government, but at the same time deters investment, which is counterproductive.

#### 14.4 The Debt Sustainability Debate

The discussion of government deficits and debts usually turn to the **sustainability** of continuous deficit spending that adds to debt and possibly to the debt to GDP ratio as well. In this section we will examine these issues. However, we also will argue that the modelling exercise is fundamentally misguided for a sovereign currency issuing government, on two grounds. First, since **affordability** *per se* cannot be an issue for a sovereign government, neither can sustainability in the sense that government can always make payments as they come due, no matter how high they become. However if the debt to GDP ratio continuously grew, and interest payments on the debt grew faster than national income, while affordability cannot be an issue, the **crowding out** other types of important government spending would be a concern. Second, and equally important, is that the simple modelling framework of the growth process is flawed because it ignores the likelihood of changes to economic behaviour that would alter the dynamic relationship between the deficit and debt ratios.

Let us begin with a typical model used to evaluate the sustainability of deficit spending. At the outset we should again note that orthodox economists view the **Government Fiscal Constraint** as an *ex ante* planning instrument and that the financing of deficits by the issue of HPM is inflationary. These arguments have been thoroughly canvassed and rejected by MMT in previous sections of this Chapter. However to maintain consistency with this literature, all deficits lead to debt issue in the following model.

We measure the change in the debt to GDP ratio between period 0 and period 1,  $\Delta d$ , on the left hand side:

$$\begin{aligned}
 (14.2) \quad \Delta d &= D_1/Y_1 - D_0/Y_0 = (D_0(1+r) - S_1)/(Y_0(1+g)) - D_0(1+g)/(Y_0(1+g)) \\
 &= (D_0/Y_0)(r-g)/(1+g) - S_1/Y_1 \\
 &= d_0(r-g)/(1+g) - s_1
 \end{aligned}$$

where  $D_t$  and  $Y_t$  are respectively real debt and real GDP in period  $t$ , respectively. The lowercase symbols  $d$ ,  $s$  denote the debt and primary surplus to GDP ratios, respectively.

The outstanding debt in period 0 will grow at a real rate  $r$ , due to interest payments, (that is, to  $D_0(1+r)$ ), but will be offset (increased) by a government primary surplus (deficit) in period 1,  $S_1$ .

However the debt ratio in period 1 must take into account the level of real GDP in period one. We assume that real GDP grows to  $Y_0(1+g)$  in period 1. These key terms are usually in **real**, that is, inflation adjusted, terms but that really does not matter; we can keep it all in nominal terms since ‘deflation’ by the inflation rate merely reduces all terms by the inflation rate. The final term on the right hand side is simply the original debt ratio,  $D_0/Y_0$ . Careful algebraic manipulation, including the separation of the surplus term, yields the final expression for the change in the debt ratio. We assume for the moment that  $r$  and  $g$  are constant.

Thus, according to this analysis, two factors assume importance in determining whether the debt ratio increases from one period to the next, namely:

- The sign of the difference between the real interest rate on debt ( $r$ ) and the real growth of GDP ( $g$ ); and
- Whether or not the government running a primary surplus in period 1, that is,  $S_1 > 0$ .

We can say that if  $r > g$ , then the debt ratio would continue to rise if the annual primary fiscal balances are zero or in deficit. On the other hand, if the government runs a constant primary surplus ratio, say  $s^*$ , then there is a critical value of the debt ratio, say  $d^*$ , at which the debt ratio will remain constant, as long as  $r$ ,  $g$ , and  $s^*$  remain constant.  $d^*$  can be written as:

$$(14.3) \quad d^* = s^* (1+g)/(r - g)$$

But this debt ratio,  $d^*$  is unstable, so if, for example, the primary surplus ratio temporarily departed from its previous value ( $s^*$ ), the debt ratio departs from its magnitude,  $d^*$ , and does not revert to  $d^*$ , even if  $s$  is restored to its previous value.

On the other hand, say the treasury runs a persistent deficit, but the real growth rate exceed the real interest rate, so that  $g > r$ , then there is a constant debt ratio, which satisfies Equation (14.3), with both the numerator and denominator now negative, so  $d^* > 0$ . Further it can be shown that, in contrast to the previous scenario, this equilibrium debt ratio is stable. On the other hand, in the unlikely event of both a primary surplus and  $g > r$ , then the debt ratio steadily declines.

Orthodox economists draw on a model of the type shown in Equation (14.1). Typically they solve the model based on constant values of the key parameters, namely  $g$ ,  $r$ , and  $s$ , as we have done above. However, they fail to acknowledge that the growth rate of GDP,  $g$ , and the treasury's financial balance,  $s$  are interdependent in the presence of the **spending multiplier effect**. Further, while the central bank has a major role in setting the overnight rate, which provides the intercept for the yield curve, it also has the capacity to flatten the yield curve by being prepared to buy longer term debt and hence influence the interest rate on debt,  $r$ .

Through quantitative easing the USA, UK, Japan and more recently the Eurozone countries through the ECB have all been trying to stimulate private sector spending through lower long term rates. Thus the levels of  $g$ ,  $r$  and  $s$  are not predetermined. This will have a major impact on the dynamics of the relationship between deficit and debt ratios. We now examine a number of scenarios, which challenges the commonly held view that fiscal policy geared to the achievement of sustained full employment is likely to lead to an unsustainable debt ratio.

- Higher government deficit spending could increase the GDP growth rate, and if necessary, through the actions of the central bank, it can be pushed above the interest rate. This can fundamentally change the debt dynamics, as argued above, by reversing the rise of the debt ratio.
- Modest Inflation: this tends to increase tax revenues, through bracket creep, so that they grow faster than government spending, thus lowering deficits. Many would point to the tendency to generate 'negative' real interest rates, which makes it easier for real growth to exceed the real negative interest rate. Again, subject to the magnitude of the primary deficit ratio, the deficit debt dynamics would change.
- Stimulus: government can try to adjust its fiscal stance (decreasing taxes and raising spending which will increase its deficit in the absence of sufficient crowding in). Under plausible values of the expenditure multiplier and the initial debt ratio, the debt ratio will decline, due to GDP growing faster than total debt.



- The private sector may adjust its flows (spending and saving) in response to the government's fiscal stance. If government continually spends more than its income, it will be adding net wealth to the private sector; and its interest payments will add to private sector income. This is likely to lead to additional spending in the private sector, and a lower saving ratio, due to the **wealth effect**. Thus spending will rise relative to private sector income. If the private sector reduces its surpluses, this can only be achieved by lower government sector's deficits. Thus the likely result is that tax revenues and consumption will rise, the government's deficit will fall, and the private sector's surplus will fall. The explosive debt growth scenario is implausible and is based on the presumption that the non-government sector never changes its behaviour.
- Finally, and this is the most contentious point. Suppose none of the dynamics just discussed come into play, so the government's debt ratio rises on trend. Will a sovereign government be forced to miss an interest payment? In the USA Chairman Bernanke used to explain that all Fed spending to bail out Wall Street occurred by using keystrokes, or, electronic entries on balance sheets. There is no technical or operational limit to its ability to do that.

We can conclude that there is a fundamental difference between perpetual private sector deficit spending and perpetual sovereign government sector deficits: the first really is unsustainable while the second is not.

We have argued that persistent government budget deficits lead to increasing private wealth and possibly higher treasury debt ratios (Watts and Sharpe, 2013). However the sovereign currency issuing government can continue to make all payments as they come due, no matter how big the debt ratio becomes. The mere act of making those payments could lead to inflation. They could lead also to policy changes, such as lower interest rates and behavioural changes by the non-government sector, which are likely to cause changes in growth rates, and deficit and debt ratios. Hence, a rising treasury debt ratio is unlikely to last forever.

As Lerner argued, we need to take a functional approach to fiscal policy. Rather than worrying about deficit and debt ratios we should focus on what really matters: employment, growth, inflation, exchange rates, environmental sustainability, inequality, and other social and economic indicators of quality of life.

## 14.5 Conclusion: MMT and Fiscal Policy

On one level, the MMT approach is descriptive: it explains how a sovereign currency works in practice. When we talk about government spending by keystrokes and argue that the issuer of a sovereign currency cannot run out of them, that statement is descriptive and factual. Equally the statement that sovereign governments do not borrow their own currency is also descriptive and factual. Our classification of bond sales as part of monetary policy, to help the central bank hit its interest rate target, is also descriptive and factual. And, finally, when we argue that a floating exchange rate provides the most domestic policy space that is also descriptive and factual.

**Functional finance then provides a framework for prescriptive policy.** It says that sovereign government ought to operate fiscal and monetary policy to achieve full employment. In Lerner's view this is done by setting the government's net spending at the right level - spending more and taxing less when there is unemployment - and setting the interest rate at the right level. That isn't

very radical; it was adopted by post-war Keynesians, and also even by Milton Friedman (who had his own version of functional finance).

However, Lerner's initial proposal was formulated in an economic environment of low inflation. Indeed, there were concerns about a return to deflation, such as that suffered in the 1930s. Later, after inflation reared its ugly head during the 1960s, Lerner became quite concerned about price stability. He developed a policy proposal which argued for wage and price controls. However since the late 1970s the major developed economies have always relied on tight fiscal and monetary policy to fight inflation.

The problem is that governments had to abandon any pretence that they were pursuing full employment. Indeed, unemployment became a tool for achieving price stability. It became worse, with the conventional wisdom arguing that central banks ought to pursue only price stability, and with the use of fiscal policy downgraded altogether. Lerner's 'steering wheel' approach to policy was abandoned. The result has been typically high unemployment and substandard economic growth. In the US poverty and inequality have risen. Globally, growing unemployment has been a problem even during economic expansions.

In Chapter 12 we examined an alternative strategy to create jobs without sparking inflation - the JG approach. The JG program directly targets the unemployed to lift them out of poverty. On the other hand, using general tax cuts or spending increases tends to favour the already relatively well-off rather than job opportunities **trickling-down** to the unemployed and poor.

A sovereign currency needs an **anchor**, and by setting the basic wage in a JG program, the program itself becomes the anchor. As long as the program wage is held steady, and so long as there are employees in the program, an employer can recruit a new worker out of the program at a slightly higher wage.

Operating the economy at full employment and with a relatively stable wage in a buffer stock jobs program will help to stabilise not only consumption spending and household income, but it also helps to stabilise wages and therefore prices.

## References

Cagan, P. (1956). "The Monetary Dynamics of Hyperinflation". In Friedman, M. *Studies in the Quantity Theory of Money*, University of Chicago Press, Chicago.

Friedman, M. (1953) *Essays in Positive Economics*, University of Chicago Press, Chicago.

Friedman, M. (1969) "The Optimum Quantity of Money," in *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine Publishing Company.

Keynes, J.M. (1920). *The Economic Consequences of the Peace*, New York: Harcourt Brace, New York.

Keynes, J.M. (1930) [1976] *A Treatise on Money*, volumes I and II, Harcourt, Brace & Company, New York.

Lerner, A. (1943) "Functional Finance and the Federal Debt" *Social Research*, 10(1), 38-51.

Lerner, A.P. (1947) "Money as a Creature of the State", *The American Economic Review*, Papers and Proceedings, 37(2), 312-317.

Mitchell, W.F. (2010) “Modern Monetary Theory and inflation – Part 1”, available at: <http://bilbo.economicoutlook.net/blog/?p=10554>.

Mitchell, W.F. (2011) “Modern Monetary Theory and inflation – Part 2”, available at: <http://bilbo.economicoutlook.net/blog/?p=13035>.

Watts, M. J. and Sharpe, T. (2013) “The ‘Immutable’ Laws of Debt Dynamics”, *Journal of Post Keynesian Economics*, 36(1), 59-84.

# Chapter 15: Monetary Policy in Sovereign Nations

## Chapter Outline

- 15.1 Introduction
- 15.2 Modern Banking Operations
- 15.3 Interest Rate Targets versus Monetary Targets
  - Lender of last resort and financial stability
- 15.4 Liquidity Management
  - Introduction
  - Different interest rate setting arrangements
- 15.5 Implementation of Monetary Policy
  - Transmission mechanism
- 15.6 Unconventional Forms of Monetary Policy
  - Introduction
  - Quantitative easing (QE)
  - Negative interest rates
  - Conclusion
- 15.7 Monetary Policy in Practice
- 15.8 The Advantages and Disadvantages of Monetary Policy
- 15.9 Central Bank Independence
  - Introduction
  - Rationale for independence
- 15.10 Horizontal and Vertical Operations: An Integration

## **Learning Objectives**

1. Review arguments in favour of interest rate rather than monetary targets by central banks.
2. Have the capacity to analyse the nature of liquidity management under different interest rate setting arrangements.
3. Understand how changes in monetary policy are transmitted and impact on the macroeconomy.
4. Recognise that the independence of the central bank is somewhat circumscribed.

## 15.1 Introduction

In Chapter 13 we examined monetary and fiscal operations. The central bank handles the treasury's tax revenue and receipts from bond sales as well as treasury's payments. These fiscal operations necessarily impact on banking system reserves. Thus the central bank must work closely with treasury in an accommodative role to minimise fluctuations in bank reserves to ensure that the interest rate target is met. Through its examination of this coordination, MMT has made an important contribution to our understanding of these operational realities.

In Chapter 14, we explored the debates about the merits of treasury engaging in discretionary fiscal policy. The claims that, when 'financed' by monetary expansion, expansionary fiscal policy was inflationary and that financing by debt issue caused crowding out of private sector expenditure, are rejected by MMT. Also the claim that, like the private sector, treasury could be subject to adverse deficit debt dynamics, was rejected for a range of reasons including that a sovereign issuer of the currency can always pay off debt denominated in its own currency.

In this third chapter on macroeconomic policy, our first objective is to bring together earlier discussions of modern banking operations in the form of a short summary. We then return to the analysis of the switch by many central banks from money supply targeting to interest rate targeting. We will then consolidate our understanding of liquidity management – i.e. the central bank operations, which ensure that the target interbank rate is achieved, by the consideration of different institutional settings. We will investigate the conduct of monetary policy by sovereign governments, which over the last forty years has been the dominant strand of macroeconomic policy in developed economies. We will then examine and assess the adoption of unconventional forms of monetary policy by some central banks, when they had limited capacity to ease (conventional) monetary policy any further.

We briefly examine the unemployment and inflation outcomes for Australia, with reference to when inflation targeting was introduced in mid-1993 and then assess the advantages and disadvantages of adopting monetary policy, as the primary macroeconomic policy tool.

The nature of central bank independence is investigated and this is followed by an integration of the verticalist and horizontalist strands of argument.

## 15.2 Modern Banking Operations

Private IOUs (funds owing) are denominated in the domestic currency. Similarly, private issuers of IOUs also promise to accept their own liabilities. For example, if a household has a loan with its bank, it can always pay principle and interest on the loan by writing a cheque on its deposit account at the bank. In this case, the bank accepts its own IOU in payment, just as governments accept their own liabilities (currency) in payments on debts due to government (tax liabilities).

Indeed, modern banking systems operate a cheque clearing facility so that each bank accepts cheques drawn on all other banks in the country. This allows anyone with a debt due to any bank in the country to present a cheque drawn on any other bank in the country for payment of the debt. The cheque clearing facility then operates to settle accounts among the banks. Banks clear accounts using government IOUs, and for that reason either keep some currency on hand in their vaults, or more importantly maintain **reserve deposits** at the central bank.

As we noted in Chapter 6, all modern financial systems have developed payments systems, (of which the cheque clearing facility is a part) that ensure banks can get additional currency and reserves as necessary from other banks (called the interbank overnight market), or through borrowing them from the central bank which enables them to clear accounts among themselves and with their depositors.

Also, banks try to minimise their holdings of currency in their vaults, not only due to security considerations, but more importantly they would prefer to hold assets in the form of loans because they will receive interest from the borrowers. Thus they leverage their currency reserves, and hold a small fraction of their assets in the form of reserves. So long as only a small percentage of their depositors try to convert deposits to cash on any given day, this is not a problem. However, in the case of a bank run in which a large number of depositors try to convert on the same day, the bank will have to obtain currency from the central bank.

This can even lead to a **lender of last resort** action by the central bank that lends currency reserves to a bank facing a run. In such an intervention, the central bank lends its own IOUs to the banks in exchange for their IOU - the bank gets a reserve credit from the central bank (an asset for the bank) and the central bank holds the bank's IOU as an asset. When cash is withdrawn from the bank, its reserves at the central bank are debited, and the bank debits the account of the depositor who withdrew funds. The cash then held by the depositor is the central bank's liability, offset by the bank's liability to the central bank.

### 15.3 Interest Rate Targets versus Monetary Targets

For a long time after the 1960s, Monetarist thinking held sway, which focused attention on a central bank's purported ability to control the money supply. Many central banks, including those in the USA, UK, Canada, Germany and Australia, targeted a monetary aggregate that is a measure of the **money supply**. They did so because through the Quantity Theory of Money (QTM) the growth of the money supply in the long run was alleged to determine the inflation rate. While mainstream (orthodox) economists used to argue that the central bank could use quantitative constraints as a means to controlling the private creation of money (credit creation), most economists now recognise that the central bank can only set the overnight interest rate, which has only an indirect impact on the quantity of reserves and the quantity of privately created money.

Central banks around the world initially set monetary targets in the late 1970s and into the 1980s as the Monetarist ideas gained traction in policy-making circles. However, by the mid-1980s they had all discovered that they were unable to control the money supply and abandoned this major plank of Monetarist thinking. Central banks now pay little attention to the growth rates of monetary aggregate, realising that they can really only set the price (the interest rate) and that the state of the economy sets the money supply. Students may wish to review the discussion of the contrasting perspectives about the credit creation process and the **endogeneity of the money supply** in Chapter 10.

Despite central banks ceasing to target a monetary aggregate, monetary policy, in the form of interest rate targeting, has remained the primary arm of macroeconomic policy and is still designed to control inflation. However it is rarely explained as to why discretionary changes in monetary policy through changes in the target interbank rate can effectively target the inflation rate.

At best, interest rate changes may lead to changes in spending, which could impact on the inflationary process.

The primacy of monetary policy over fiscal policy reflects the propositions arising from the Expectations Augmented Phillips curve, which became the dominant theory of inflation in the early 1970s. Fiscal policy was rejected as a strategy to achieve a lower rate of unemployment, because of the inflationary consequences. Low and stable inflation became the main target of macroeconomic policy, as opposed to full employment, and reflected the belief that an economic environment of low and stable inflation was the most conducive to private sector spending and employment. Private sector planning would not be undermined by unexpected changes in the inflation rate. In addition, there was the presumption that an independent body charged with formulation of monetary policy would make better decisions.

Some central banks are subject to formal **inflation targets**. For example, the Bank of England is subject to a target inflation rate of 2 per cent per year. On the other hand, the Reserve Bank of Australia adopted a **targeting range** and is required to maintain the Australian inflation rate at between 2 and 3 per cent per annum. The US Federal Reserve is not subject to a formal inflation target, but in its latest statement about longer-run policy goals and its monetary policy strategy, the Federal Open Market Committee states that a CPI inflation rate target of 2 per cent is “most consistent over the longer run with the Federal Reserve’s statutory mandate.” (FOMC, 2016)

In addition to controlling inflation, central banks are also subject to other policy goals. For example in the USA, the FOMC is “firmly committed to fulfilling its statutory mandate from the Congress of promoting maximum employment, stable prices, and moderate long-term interest rates” and also states that “the median of FOMC participants’ estimates of the longer-run normal rate of unemployment was 4.9 per cent” (FOMC, 2016). The Bank of England must also “support the Government’s economic objectives including those for growth and employment” (Bank of England, 2016). In Australia, the RBA must assist in the “the maintenance of full employment in Australia”; and more generally “the economic prosperity and welfare of the people of Australia” (Reserve Bank Act, 1959).

A third policy objective of moderating asset price growth is sometimes mentioned, but not formally acknowledged.

Central banks in many developed economies set their interbank targets each month, which result from the deliberations of their independent committees in the light of the prevailing policy priorities. Since the demand for reserves by the banks is relatively inelastic, accommodative behaviour by the central bank is typically not required following an announcement of a change in the target rate.

### **Lender of last resort and financial stability**

In a crisis, an important role played by the central bank is to operate as a **lender of last resort**, providing reserves on demand to financial institutions. Originally, this was to stop a bank run, when depositors tried to exchange their deposits for cash. When the GFC impacted, banks could not refinance their positions in assets because their creditors demanded payment as liabilities matured. Central banks around the world had to step in to provide the refinancing and prevent the financial system from collapsing. Central banks have unlimited capacity to provide currency to the financial system on behalf of the government of the day. During the GFC, the Australian Government also provided a deposit guarantee to give bank customers reassurance that their deposits were safe.



## 15.4 Liquidity Management

### Introduction

In this section we will build on the analysis in the earlier chapters of liquidity management by central banks to ensure that their target interbank rate is achieved, but we will consider different institutional arrangements.

MMT shares the view that the central bank cannot control either the level of bank reserves or the money supply. Instead the central bank must accommodate the demand for reserves. Thus, the supply of reserves is best characterised as horizontal, at the central bank's target rate. That is the **endogenous money, horizontal reserve approach**, which was developed over the 1970s and 1980s by Moore and other post Keynesians (see Lavoie 1984; Moore 1988). Virtually all economists, from all schools of thought, now accept this is a correct representation of the operating procedures of modern central banks.

However, the arguments for a horizontal supply of reserves provided by the central bank were formulated without considering:

- The operation of fiscal policy; and
- When the central bank's target overnight interest rate was either near-zero or equal to the interest rate paid to the banks for their holdings of reserves.

Through the operation of stimulatory fiscal policy, bank reserves increase initially since providers of goods and services to the government are paid by their deposits being credited, which leads to the reserves of their banks being credited. When the interest rate paid by the central banks on reserves held by the banks (the support rate) is less than the target interbank rate, then the central bank would have to drain any excess reserves by selling government bonds (debt). Otherwise the market rate at which banks lend to and borrow from each other would fall below the target rate, due to the operation of market forces. In the case of a fiscal surplus, open market operations may need to be conducted to augment the stock of reserves. In Chapter 13, we provide a simplified outline of the process of liquidity management by the central bank in the context of the implementation of fiscal policy. Thus the essence of the argument that reserves are supplied elastically at the target rate is unchanged, when fiscal policy is considered.

According to the logic of the **mainstream textbook money multiplier model**, the central bank could increase the money supply by injecting reserves through an open market purchase, which would enable greater bank lending, given the assumption that credit creation increases mechanically in line with the requirements of the **Fractional Reserve System** or the (predictable) ratio of deposits to reserves chosen by the banks (see Chapter 10 for a simple exposition of this theory of credit creation). However, this fails to recognise that the added reserves in excess of the banks' desired reserves would immediately drive the interbank rate to zero or to a non-zero support rate, since reserve requirements would not change until the following accounting period. If the target and support rates were unequal, the achievement of a positive target interbank rate would force the central bank to sell securities to drain those excess reserves, which had just been added.

On the other hand, if the central bank wanted to reduce the money supply by taking reserves out of the system when there were no excess reserves, then some banks would have insufficient reserves. The central bank would have no choice but to add reserves back into the banking

system to keep the market (interbank) rate at its target level. In both cases, the level of bank reserves and the money supply would remain unchanged. Thus the central bank does not have the discretion to alter the supply of reserves to pursue an ostensible policy objective. An increase in the money supply, resulting from credit creation by banks will cause changes in the monetary base, when banks adjust their reserves. The causation is not from a change in reserves (the monetary base or HPM) driving credit creation.

### Different interest rate setting arrangements

In the aftermath of the Global Financial Crisis, the USA and Japan adopted a near zero interest rate target (see Table 15.1), so excess reserves could be left in the system. In the USA between 0 and 25 basis points (0.25 per cent) was paid on reserves by the Fed until December 2015. In that case, irrespective of the level of excess reserves held by the banks, the market rate remains within that range, since banks could get 25 basis points on excess reserves from the Fed so there is no point in lending the excess reserves in the Fed funds (interbank) market at a lower rate. In December 2015 the USA raised its rate by 25 basis points to 25 – 50 basis points. In 2009 the Bank of England set both the bank (target) rate and the support rate at 0.5 per cent, so again excess reserves could be left in the banking system, since no bank would lend reserves to another at a rate less than 0.5 per cent.

**Table 15.1 Target interbank rates for developed economies**

Country	Interest Rate (February 2016)	Previous Change to Rate
UK	0.5% from March 2009	1.0% from February 2009
USA	0.25% – 0.5% from December 2015	0.0% – 0.25% from December 2008
Japan	-0.10% – 0.00% from February 2016	0.00% - 0.10% from April 2010
Australia	2.00% from May 2015	2.25% from February 2015
ECB	0.05% from September 2014	0.15% from June 2014

Thus, in the circumstances of either a positive target rate equalling the support rate or a near zero target rate, the central bank can leave banks with excess reserves without compromising monetary policy, so that they are not forced to play an accommodative role. However, a shortage of reserves would drive the market rate above the target rate and would require a response from the central bank.

The key question then is whether the capacity of some central banks to exercise some discretion in respect of liquidity management means that the debate over whether reserves drive deposits and hence the money supply should be revisited. The answer is no, because a profitable loan requires a creditworthy borrower and an adequate interest rate differential between lending and borrowing rates for the bank to make an adequate return. The presence or absence of sufficient reserves has an incidental impact on this calculation.

The Bank of England has made concessions about the role of banks in the financial system, in that it now rejects the simplistic intermediation role of banks using deposits as a basis for credit creation (McLeay et al. 2014). However, it still appears to take the view that the monetary base

or HPM, that is the level of bank reserves plus the holdings of cash by the non-government sector, influences credit creation by the banks (see below).

## 15.5 Implementation of Monetary Policy

### Transmission mechanism

Here we examine how changes in monetary policy are alleged to impact on the macro-economy through the operation of a Transmission Mechanism. Typically the Yield Curve (YC) slopes upwards due to uncertainty about inflation rate and interest rate risk (see Chapter 10). Let us consider a cut in the target interbank rate.

Then the fall in the (target) interbank rate is likely to reduce medium- to long-term interest rates on similar assets through arbitrage and hence the YC is likely to shift downwards. Also there is an effect on other private sector short-term rates (for example, on bank deposits) and also long term rates on business and consumer loans, which in Australia have normally been adjusted in line with the change in the target rate. However if the change in the target rate is widely anticipated by the markets, it may be already factored in with respect to longer-term rates, so the movement of the YC, following the announcement of the change in the target interest rate, may be less clear-cut.

Investment is a component of aggregate expenditure, which is considered to be interest sensitive. New physical capital investment is usually financed using borrowed funds. These projects are only undertaken if the expected net profit rate exceeds the real interest rate on borrowed funds. Here the long term interest rate is relevant, since the construction phase of major investment projects may last many months, if not years, and revenue from the project will not be generated until production starts and sales are made. More projects would be expected to make a sufficient return to justify additional spending, following the interest rate cut. Long run expectations of an uncertain future, which underpin profit expectations, are quite volatile, however and reflect confidence. Thus there is no guarantee that a 25 or 50 basis point cut in borrowing rates would have a positive impact on investment spending. In addition, the proposed level of investment say in 2016 will be the outcome of detailed planning in prior months and is unlikely to be sensitive to modest interest rate changes in the short term.

Durable consumption spending (for example, cars, houses and white goods) is often financed by borrowing, so borrowing costs and hence the capacity to service the loan, are also relevant to these spending decisions. In an uncertain economic climate, the likelihood of job security in the years ahead will also be an important consideration in terms of the capacity to repay the loan.

A fall in the target rate lowers domestic interest rates in general, and reduces the interest rate differential between the country in question, say Australia, and other countries. International capital flows (sales of financial assets to and purchases from foreigners) are claimed to respond to interest rate differentials. Then there is likely to be a fall in capital inflow (that is lower sales of domestic assets to foreign investors relative to the purchase of foreign assets by domestic citizens) and so the local currency is likely to depreciate. This means that imports are more expensive when priced in the local currency and less attractive to buy and exports are cheaper when their prices are denominated in a foreign currency, so net exports may increase, adding to any rise in total spending associated with changes in investment and durable consumption expenditure.

While small changes in long-term interest rates (following corresponding changes in the target rate) may have little impact on spending, higher and higher long-term interest rates will certainly kill off interest rate sensitive domestic spending.

Thus, reliance on monetary policy to impact on aggregate expenditure and indirectly the inflationary process is highly problematic. We briefly review the performance of the Reserve Bank with respect to inflation and unemployment outcomes in Section 15.7.

## **15.6 Unconventional Forms of Monetary Policy**

### **Introduction**

In developed countries adversely affected by the GFC, including the US, UK, Japan and the Eurozone, official rates were frequently cut in the months after 2008, down to historically low levels. With the exception of the USA, which raised the Federal Funds rate to a range of 0.25 per cent to 0.5 per cent in December 2015, these historically low rates have persisted through to the beginning of 2016.

The period since the advent of the GFC highlights the fact that monetary policy is still considered to be the main arm of macroeconomic policy. In addition, monetary policy has not only been conducted to control inflation. Since the GFC, inflation has been low for the developed economies and indeed the prospect of deflation has been often canvassed, particularly in the Eurozone countries. The GFC rapidly manifested itself as a real crisis with the collapse of private sector spending. Fiscal stimulus measures were sanctioned by the Organisation for Economic Cooperation and Development (OECD) and the International Monetary Fund (IMF) but with the proviso, that such measures should be discontinued if adverse deficit-debt dynamics occurred. Stimulus measures were adopted in countries, including Australia, the UK and the USA, in 2009 but were discontinued by 2010. Thus there has been a reliance on monetary policy, but with rates close to zero in the major economies, there has been limited scope for further cuts in official rates. Consequently, these major developed economies have adopted unconventional forms of monetary policy.

### **Quantitative easing (QE)**

In the absence of a capacity to reduce the intercept of the yield curve (YC), since their target interbank rates have been so low, central banks in the UK Japan, the USA and most recently the ECB, have resorted to trying to flatten the YC, by so-called Quantitative Easing. In simple terms, this means that central banks in these countries have developed programs for buying both longer term treasury debt and, in some instances, private sector financial assets from the bank and non-bank private sector. The objective here is to boost demand for these financial assets, which increases their market prices and thus lowers their yields, thereby flattening the YC. In addition, the impact of the private sector selling financial assets to the central bank is that bank reserves increase, which contributes to the overall growth of bank reserves in these countries. The interest rate setting arrangements, post-GFC, in these countries, namely the equality of the target and support rates, enabled QE to be implemented.

The Bank of England identified a number of mechanisms by which QE is supposed to promote spending. First, the purchases of financial assets reduce short-term and longer-term rates and promote greater confidence. This is the signalling channel. Second, the purchases of long-term

securities by the central bank gives rise to capital gains for households who are assumed to consume part of their increase in wealth. This is referred to as the portfolio (re)balance channel. Third, the increased bank deposits and reserves created by asset purchases from non-bank institutions will increase the availability of bank credit, and so, banks may be more willing to lend (Joyce et al. 2012). This is the bank funding/lending channel.

MMT advocates would challenge whether the direct reduction of long-term rates via QE would be any more successful in stimulating these economies than the earlier cuts to official rates which indirectly reduced long-term rates. The main problem in many developed economies is sluggish growth and the consequent absence of large numbers of creditworthy firms seeking to borrow. Second, there may be some modest increases in spending from the wealth effects. Third, MMT would reject the bank funding/lending channel since it is premised on reserves driving deposits, in short the money multiplier mechanism.

### Negative interest rates

A second form of ‘unconventional’ Monetary Policy can be identified with a number of central banks, including the ECB (June 2014) and Japanese central bank (late January 2016) resorting to negative rates on deposits of the private banks (reserves). The ECB also commenced a QE program, which will continue until at least March 2017. The ECB rate is now set at minus 0.3 per cent (December 2015). It is designed to counter the impact of price deflation on real interest rates and encourage the private banks to engage in lending.

Japan cut its benchmark rate to minus 0.1 per cent. However, to reduce the impact on the earnings of financial institutions, the reserves held by financial institutions will be divided into three tiers to which a positive, zero and negative interest rate will be paid. Banks would have a greater incentive to economise on their holdings of reserves. In addition, the Bank of Japan also committed to continue its QE program by ongoing purchases of Japanese government bonds (JGBs) worth about of about 80 trillion yen per year. These two measures should reduce the intercept of the yield curve and “exert further downward pressure on interest rates across the entire yield curve” (Bank of Japan, 2016). The objective of this strategy is to “achieve the price stability target of 2 per cent at the earliest possible time”.

Mitchell (2016) is critical of the imposition of a negative target rate, as well as the ongoing adoption of QE. He notes that it is curious to impose a new public tax on the private sector, when the objective is to achieve a higher inflation rate, but acknowledges that this a cautious move compared to ECB’s imposition of a 0.3 per cent tax on all reserves. He reiterates the MMT argument that policies designed to increase bank reserves do not increase the incentive of banks to create credit. It can also be argued that the imposition of reduced interest rates is unlikely to promote increased bank lending in a depressed economic climate.

Mitchell (2016) quotes the Borio and Disyatat (2009:19) who argue that:

A striking recent illustration of the tenuous link between excess reserves and bank lending is the experience during the Bank of Japan’s ‘quantitative easing’ policy in 2001-2006. Despite significant expansions in excess reserve balances, and the associated increase in base money, during the zero-interest rate policy, lending in the Japanese banking system did not increase robustly.

Mitchell (2016) makes the point that:

Japanese banks are not expanding credit not as a result of their unwillingness to make loans or a lack of reserves. The reason for the slow credit is that businesses have sufficient capital stock to satisfy the demands of a very weak consumption sector and do not need to borrow.

He suggests that investment in Japan will remain sluggish (at around 13 per cent of GDP), until firms have more buoyant expectations of consumption expenditure and/or exports.

## **Conclusion**

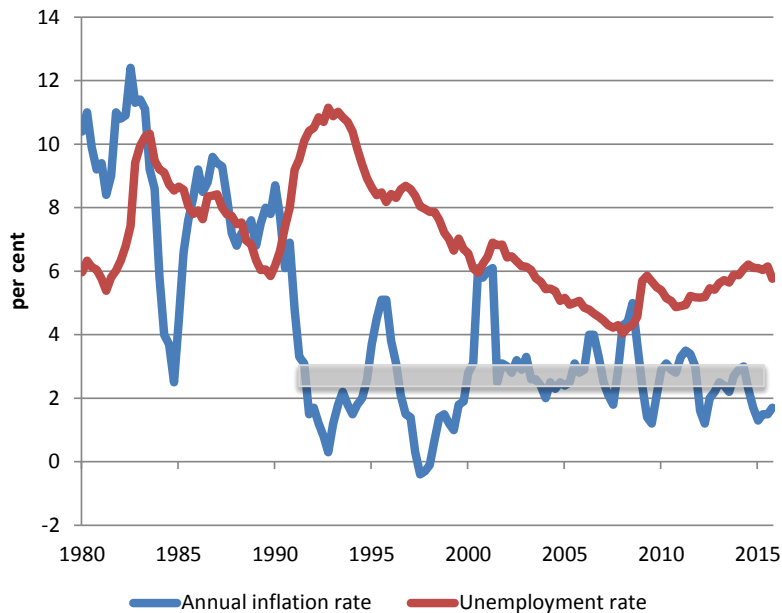
Consistent with neo-liberal thinking, governments have continued to rely on monetary policy as the primary macroeconomic policy tool since the GFC. It has largely failed to address the higher unemployment in the developed economies. As a consequence of this and the limited scope for further conventional monetary easing, central banks have resorted to unconventional monetary policy. MMT advocates argue that these measures would be ineffective, because marginal changes to interest rates, which are already at historically low levels, are most unlikely to impact on lending and spending in a depressed economic climate. The evidence so far supports this view.

## **15.7 Monetary Policy in Practice**

In this section we examine the performance of the Reserve Bank of Australia (RBA) with respect to inflation rate and unemployment rate outcomes (see Figure 15.1). In mid-1993 the 2 - 3 per cent inflation target for the RBA was introduced, which is shown by the shaded area on the graph. There was a fundamental change in the relationship between inflation and unemployment following the 1991 recession, which was analysed in Chapter 11. From about 2000, the inflation rate has exhibited a trend decline and was accompanied by falling unemployment until about 2008. From late 2010, the unemployment rate has steadily risen, albeit with some fluctuations, despite a series of reductions in the official (cash) rate, which culminated in the reduction from 2.25 per cent to 2 per cent in May 2015 (see Table 15.1).

Thus, at a time when stimulus measures were justified with the inflation rate below 2 per cent and unemployment rising, cuts to the cash rate have simply been ineffective, which is consistent with the thinking of MMT advocates. This issue is illustrated even more clearly if the performance of the peripheral Eurozone economies, such as Greece, Spain, Portugal and Ireland, is analysed. Despite persistent low inflation and periods of deflation (negative price level growth) and a low target interbank rate (currently 0.05 per cent) set by the European Central Bank, these countries are mired in high unemployment rates which still exceeded 9.1 per cent in the third-quarter of 2015 with Spain and Greece experiencing rates above 20 per cent.

**Figure 15.1 Unemployment rate and inflation rates for Australia, 1980 – 2015, per cent**



Source: RBA Statistics (CPI All Groups quarterly, annualised) Unemployed persons as percentage of labour force

## 15.8 The Advantages and Disadvantages of Monetary Policy

Monetarists and orthodox economists, in general, prefer to use monetary policy as the main arm of macroeconomic policy for a number of reasons. Monetary policy, which is implemented through interest rate setting, is claimed to be:

- Easy to implement (monthly) and flexible;
- Less subject to political interference; and
- More clearly understood by financial market traders.

In addition, if the inflation rate is subject to a target level, say 2 per cent or a range of say 2-3 per cent, then it provides an anchor for expectations of consumers and business about future inflation.

The disadvantages of monetary policy are claimed to be that it is:

- A blunt, indiscriminate policy, which is not guaranteed to work in a timely manner, whether it be stimulatory or contractionary;
- An inappropriate policy response, say to a cost-push inflation, driven by an external oil price shock from political uncertainty in the Middle East, or a widespread, severe drought;
- A single instrument, which in some macroeconomic circumstances, is attempting to affect three policy targets: inflation, GDP growth and asset price growth, which may be in conflict. The Tinbergen rule first outlined in his 1952 booklet, *On the Theory of Economic Policy*, was that consistent economic policy requires that the number of instruments is equal to the number of targets.

- Not regionally specific, so a housing boom in large cities may warrant an interest rate increase to reduce the growth of house prices, via higher (variable interest) mortgage repayments, but at the same time falling house prices and declining job opportunities in rural areas may justify a rate cut; and
- Often too tight if it is geared to a low inflation rate (or target range), which can impose major economic and social costs of higher unemployment, which are often underestimated.

## 15.9 Central Bank Independence

### Introduction

An important debate in the context of the institutional arrangements for the conduct of monetary and fiscal policy is the appropriateness of having an independent central bank within a modern monetary economy. If so, what should be the nature of the independence? We have already shown that the central bank must be accommodative in terms of its provision of reserves, unless either the target and support rates are equal or there is a near-zero target rate, when it does not have to drain excess reserves. However it is able to exercise some discretion in its setting of the target interbank rate, within the constraints of the policy objectives specified in legislation.

### Rationale for independence

Sovereign economies, such as Australia, New Zealand, Canada, and the UK, are not subject to legal restrictions on their central banks participating in the primary market for government debt, but in practice, central banks in those countries, except for Canada, have limited engagement in the primary market. This prohibition was written into US law from the founding of the Federal Reserve Bank in 1913.

Certainly there is a consensus amongst the orthodox economics profession as well as policymakers that central bank financing of budget deficits by buying treasury bills on the primary market should be forbidden. The UK Treasury, via the Debt Management Office, chooses to fully fund its financing requirement by selling debt in line with a preannounced schedule, which can be revised.

The following rationale is provided, which is representative of arguments made in support of this principle:

[T]he Government believes that the principles of transparency and predictability are best met by full funding of its financing requirement; and to avoid the perception that financial transactions of the public sector could affect monetary conditions, consistent with the institutional separation between monetary policy and debt management policy. (HM Treasury, 2012, p. 8)

In practice, the prohibition is easy to evade, which occurred during WWII in the US when budget deficits reach 25 per cent of GDP. Also, if a central bank is prepared to buy treasury bonds in the secondary market to peg an interest rate, then private banks will buy newly issued bonds and sell them to the central bank on the secondary market at a virtually guaranteed price. Since the purchases of bonds by the central bank supply the reserves needed by banks to buy bonds, a virtuous circle is created so that the treasury faces no financing constraint.



Since the Global Financial Crisis hit in 2007 these matters came to the fore in both the US and the Economic and Monetary Union (EMU) of the European Union. In the US, discussion of printing money to finance growing fiscal deficits was somewhat muted, in part because the Fed purportedly undertook quantitative easing to push banks to lend, rather than provide the treasury with cheap funding through the low interest rates. However the impact had been the same as during WWII with very low interest rates on government debt, despite a large portion of treasury debt ending up on the balance sheets of the Fed, via quantitative easing, while bank reserves correspondingly grew to historically high levels. The Fed also purchased and lent against private debt, as part of its QE program, which further added to excess reserves.

While believers in the likelihood of hyperinflation argued that the Federal Reserve Bank was essentially printing money to finance the budget deficits, most other observers endorsed the Fed's notion that QE might allow it to 'push on a string' by spurring private banks to lend. This was thought to be desirable and certainly better than financing budget deficits to allow government spending to grow the economy.

The other case is in the EMU where the European Central Bank was presumed to be prohibited from buying debt of the member governments. By design, these governments were supposed to be disciplined by markets to keep their deficits and debt within Maastricht criteria, that is, a 3 per cent deficit to GDP ratio and a 60 per cent debt to GDP ratio. These plans did not come to fruition, as we learnt earlier. The ECB's balance sheet is even bigger than that of the US Federal Reserve Bank, through the purchase of government debt of EMU members, in the secondary markets (that is, after the debt has been issued and traded).

Central banks in most developed economies are answerable to the parliament and are obliged to provide detailed information about their operations and budget. As we noted earlier in the Chapter, most parliaments specify macroeconomic objectives that guide central bank policy, which often include reference to low inflation, high (full) employment, acceptable growth, and financial stability. Parliaments may not be prescriptive about the instruments to be used to achieve these objectives, including whether to use open market operations or the discount window.

There is the commonly held view that the 'independence' enjoyed by many central banks enables them to be insulated from political pressures from special interest groups, despite the committees or boards responsible for setting interest rates often being appointed by the political party in power. Consequently they are able to make decisions, which may not be popular but have long-term economic benefits. Finally, despite some mainstream economists believing that the central bank should exercise some direct control over taxation and spending decisions, which are implemented by treasury, in practice this does not occur.

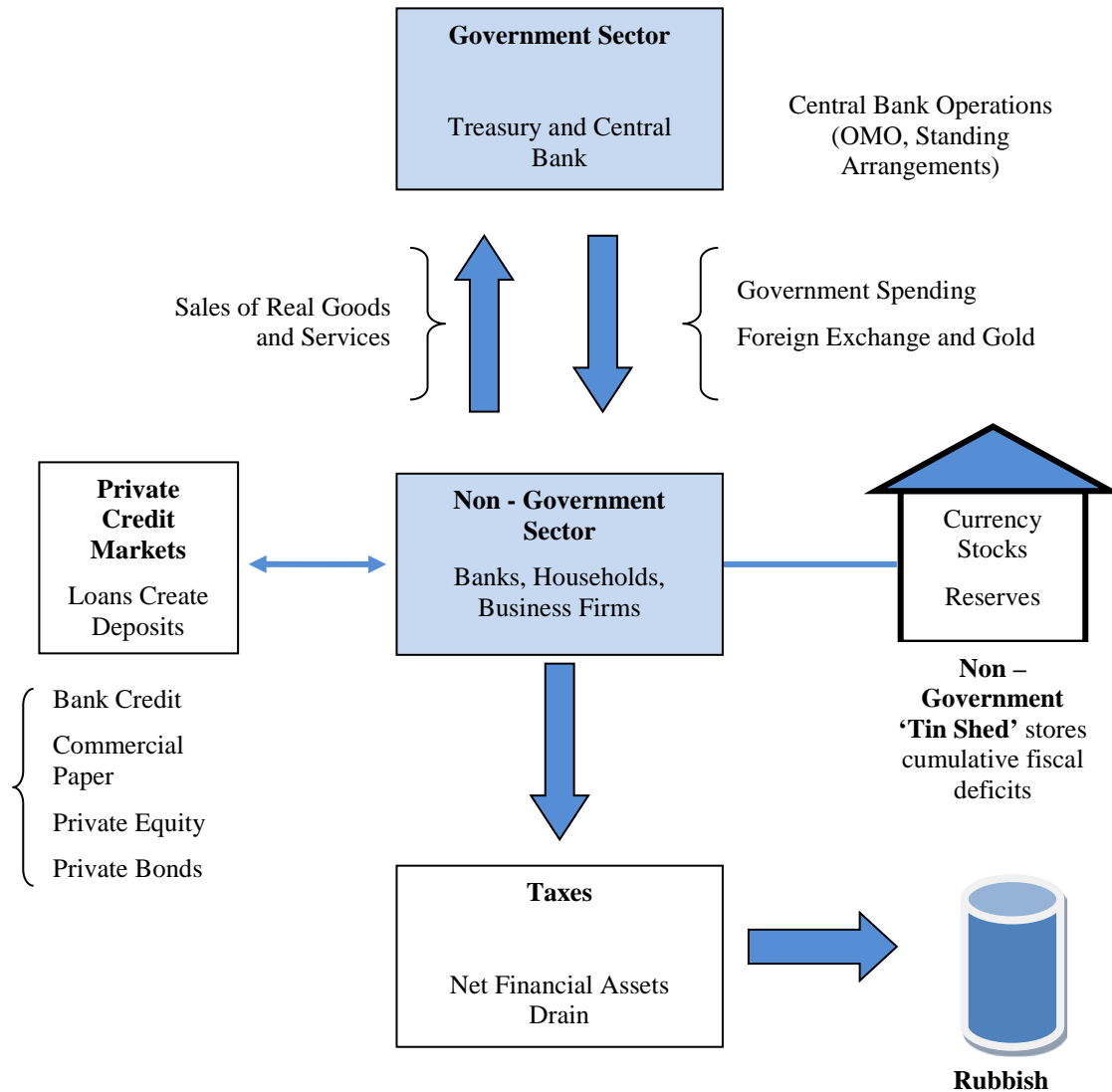
## **15.10 Horizontal and Vertical Operations: An Integration**

In some sense, the verticalists and the horizontalists have each captured elements of the money supply process. One can conceive of a vertical component of the money supply process that consists of the government supply of fiat money; money drops vertically to the private sector from government through government purchases of goods and services (and occasionally assets) as well as central bank purchases of assets (such as gold and foreign currency, and also through discounting of assets held by banks). Recall from our discussion above and in previous chapters

that the private sector is willing to accept government fiat money because the government has previously imposed tax liabilities on the private sector.

We will look at two figures 15.2 and 15.3 that outline the vertical and horizontal aspects of the supply process. Here is the first figure:

**Figure 15.2 Vertical and horizontal macroeconomic relations**



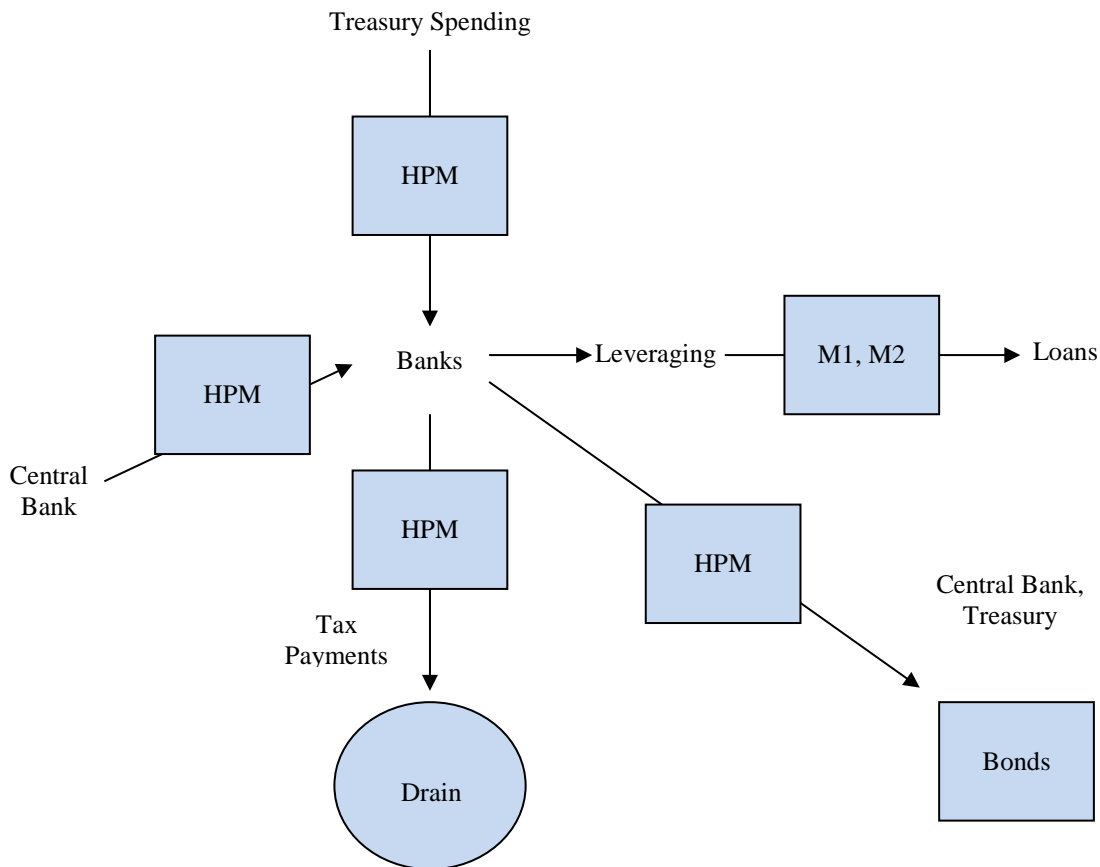
Tax payments (which discharge the liability) then drain fiat money, which can be pictured as a vertical movement from the private sector to government (and, hence, 'down the drain' as the money is simply wiped off the liability side of the central bank's balance sheet). The net difference between these two vertical flows (deficit spending) leads to accumulation of fiat money hoards (currency in the hands of the public plus bank reserves). The government can also offer to exchange interest-earning bonds for non-interest-earning cash and reserves.

We see that the government sector (treasury and central bank) injects 'currency' (broadly defined) into the economy. We can think of taxes as sending currency to the rubbish bin. What is

left (deficit spending) accumulates to stocks - the 'tin shed' - much of which is held by private institutions against the liabilities they issue.

We can think of the **bank-money-supply process as horizontal**; it can be thought of as a type of 'leveraging' of the hoarded vertical fiat money. Clearly, bank money is only one type of leveraging of the fiat money. A partial list of other types of leveraging would include commercial paper, private bonds, all types of bank liabilities, indeed all IOUs denominated in the fiat money of account. All of these private IOUs share three characteristics: they are denominated in the fiat money of account, they consist of long and short positions, and they are 'inside' debt such that the longs and shorts net to zero. A bank deposit can be thought of as a long position in fiat money, while the bank's borrowers have short positions, betting that they will be able to obtain money for delivery later.

**Figure 15.3 Horizontal and vertical components of the money supply**



A reduction of government spending can starve the non-government sector of funds, such that, those with outstanding loans to banks are not able to obtain sufficient money to make payments on loans. This is sometimes called a **short squeeze**. The borrowers, who might be workers, lose income because the cut in government spending causes unemployment to rise, or they might be business firms who experience a decline in revenue as sales fall. They cannot obtain the money through additional borrowing (an increase of the horizontal money supply). The squeeze arising from a cut in government spending could be eased if the 'savers', those with positive bank

deposit balances, were willing to increase their own spending or if others were willing to come into the market to take new short positions (lending to those squeezed). In that case, a funds shortage in the non-government sector could be relieved by operations in the horizontal section. But this is unlikely to occur when there is a slowdown in economic activity and the non-government sector adopts a cautious stance with respect to spending and lending.

The only reliable way a short squeeze can be relieved is if the government (via the vertical component) uses its capacity as the **net supplier of money**. If the government does not react to the short squeeze, the bank borrowers would be forced to try to sell assets, roll-over loans or try to obtain new loans. This can lead to a fall in asset prices, which could degenerate into a general debt deflation. Defaults also occur with increased frequency if the short squeeze persists. On the other hand, this can be avoided if the central bank enters as lender of last resort (discounting assets or buying assets held by the private sector in exchange for cash) or if the treasury increases its deficits.

If things become sufficiently bad, banks become insolvent, with asset values below the value of liabilities as borrowers start to default on their outstanding loans. If the depositors with long positions ‘liquidate’ (demand fiat money instead of bank deposits), banks are forced to the discount window to borrow reserves. Beyond some point, as bank balance sheets deteriorate, they will not have sufficient capital (net worth) to obtain discount window loans from the central bank, requiring the deposit insurer to step in to ‘resolve’ the bank. As prices fall, borrowers default, and banks fail and the private economy will almost certainly suffer a recession (or worse). However falling government tax receipts and perhaps rising government spending (through automatic stabilisers) increases the government deficit (and available net saving).

## References

Bank of England, (2016) “Monetary Policy Framework”, <http://www.bankofengland.co.uk/monetarypolicy/pages/framework/framework.aspx> (viewed 13 February)

Bank of Japan (2016) “Introduction of Quantitative and Qualitative Monetary Easing with a Negative Interest Rate” [https://www.boj.or.jp/en/announcements/release\\_2016/k160129a.pdf](https://www.boj.or.jp/en/announcements/release_2016/k160129a.pdf) (viewed on 21 February, 2016).

Borio C and Disyatat P (2009) “Unconventional monetary policies: an appraisal”, Bank for International Settlements Working Paper no. 292.

FOMC (2016) Statement on Longer-Run Goals and Monetary Policy Strategy, Adopted effective January 24, 2012; as amended effective January 26, 2016.

HM Treasury (2012) Debt and Reserves Management Report 2012–13, March 2012, available at: [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/191848/drmr\\_201314.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/191848/drmr_201314.pdf).

Joyce, M., Miles, D., Scott, A. and Vayanos, D, (2012) “Quantitative easing and unconventional monetary policy – an introduction” *The Economic Journal*, 122, F271-F288.

Lavoie, M. (1984) “The Endogenous Flow of Credit and the Post Keynesian Theory of Money”, *Journal of Economic Issues*, 18(3), 771-797.

McLeay, M., Radia, A., and Thomas, R. (2014) "Money Creation in the Modern Economy" Bank of England Quarterly Bulletin, 54 (1), 14–27.

Moore, B. (1988) Horizontalists and Verticalists: The Macroeconomics of Credit Money, Cambridge University Press, Cambridge.

Mitchell, W.F. (2016) The folly of negative interest rates on bank reserves (1 February) <http://bilbo.economicoutlook.net/blog/?p=32883>.

Reserve Bank Act (1959) Commonwealth of Australia, available at: <https://www.comlaw.gov.au/Series/C1959A00004>.

Tinbergen, J. (1952) On the Theory of Economic Policy, Amsterdam, North Holland.

# Chapter 16: Policy in an Open Economy: Exchange Rates, Balance of Payments and Competitiveness

## Chapter Outline

### 16.1 Introduction

### 16.2 The Balance of Payments

- The current account
- The capital account and financial account

### 16.3 Essential Concepts

- Nominal exchange rate ( $e$ )
- Change in  $e$  – appreciation and depreciation
- What determines the exchange rate?
- International competitiveness
- The real exchange rate

### 16.4 Aggregate Demand and the External Sector – Revisited

### 16.5 Trade in Goods and Services, Product Market Equilibrium and the Trade Balance

- National income equilibrium with trade
- The net exports function
- The impact on national income and net exports of a change in world income

### 16.6 Currency Crises

- The European exchange rate mechanism crisis – 1992
- The 1994 Mexican peso crisis
- The South East Asian debt crisis 1997

### 16.7 Capital Controls

### Appendix: Advanced Material

- An increase in world income leads to a rise in net exports
- The IMF Annual Report 1997

## **Learning Objectives**

1. Understand the components of the Balance of Payments and their inter-relationship.
2. Acknowledge the distinction between the nominal and real exchange rate.
3. Analyse the role of trade in the determination of equilibrium national income.
4. Recognise the common features of the three currency crises.

## 16.1 Introduction

In Chapter 7, we introduced trade into the income-expenditure model. The representation was simplified, in the sense that we assumed that exports were determined by the income levels prevailing in the rest of the world (that is, they were exogenous to the domestic economy) and that imports were a simple proportion of the national income of the home economy. This proportion was termed the marginal propensity to import.

In this Chapter, we extend our understanding of the way in which the economy behaves once it becomes open to the world. We will continue to consider the price level to be fixed, which means we are assuming that firms respond to an increase in aggregate demand by increasing real output. Later in the Chapter we will consider price level movements in the open economy context.

We will consider the concept of an exchange rate and examine how movements in exchange rates influence exports and imports and financial transactions between nations.

**For an economy as a whole, imports are real goods and services coming into the nation from abroad and as such represent a real benefit to residents. Conversely, exports are real goods and services that are sold to foreigners.**

**Exports represent a real cost to residents because they represent real resources (labour, capital and other productive inputs) that the residents are unable to utilise to produce goods and services for their own use.**

It is obvious that the only motivation for a nation to export, and incur the real costs involved in exporting goods and services abroad, is to gain foreign currencies, which, in turn, allow the nation to purchase other goods and services that it does not produce itself.

If imports exceed exports then a nation is able to enjoy a higher material living standard by consuming more goods and services than it produces for foreign consumption. We will consider how this conception of trade interacts with a flexible exchange rate.

You will already appreciate that the transactions between nations involve both real goods and services and financial flows. The financial transactions represent currency flows into and out of a nation and have significant implications for movements in the exchange rate and other macroeconomic aggregates, such as interest rates, the inflation rate and real GDP.

All transactions between a nation and the rest of the world are recorded in the Balance of Payments. We will initially examine the way the national statistician accounts for the external economy via the Balance of Payments, which is a framework that is closely related to the national accounts.

## 16.2 The Balance of Payments

Residents (households, firms, and governments) of every nation conduct economic transactions with residents of other nations and the record of all these transactions is recorded in the international accounts for each nation. The international accounts are made up of a number of component accounts (IMF, 2011: 7):

- The international investment position (IIP) which “shows at a point in time the value of: financial assets of residents of an economy that are claims on non-residents or are gold



bullion held as reserve assets; and the liabilities of residents of an economy to non-residents” (IMF, 2011: 7).

- The balance of payments, which is “a statistical statement that summarizes transactions between residents and non-residents during a period. It consists of the goods and services account, the primary income account, the secondary income account, the capital account, and the financial account” (IMF, 2011: 7).
- All “other changes in financial assets and liabilities accounts” (valuation changes etc) (IMF, 2011: 7).

The Balance of Payments and related accounts are compiled by national statistical agencies (such as the UK Office of National Statistics, the Australian Bureau of Statistics, the US Bureau of Economic Analysis) using an international standard set down in the International Monetary Fund’s Balance of Payments and International Investment Position Manual (BPM6) (IMF, 2011), augmented by the System of National Accounts (SNA) (United Nations Statistical Commission, 2009). While there are variations in terminology used by different nations the principles are universal.

The IMF Manual is seen as the “standard framework for statistics on the transactions and positions between an economy and the rest of the world” (IMF, 2011: 1).

The differentiating feature of these different accounts relates to “the nature of the economic resources provided and received” by the nation (IMF, 2011: 9).

Like any accounting framework, the Balance of Payments is based on a double-entry debit and credit system of record. Every transaction that is recorded has two equal and offsetting entries, each of which corresponds to the inflow and outflow of funds. We just record the net flows.

Table 16.1 shows how the Australian Bureau of Statistics presents the Balance of Payments data for Australia. Observe the heading structure: Current Account and Capital and Financial Account being the major sub-accounts of the Balance of Payments. Then within each of the sub-accounts are a number of other sub-headings, which record different elements of the transactions between Australia and the rest of the world.

We will briefly discuss the Current Account and the Capital and Financial Account.

**Table 16.1 Australian balance of payments, various years, current prices**

	2010-11	2011-12	2012-13
	\$m	\$m	\$m
<b>Current Account</b>	-34,384	-40,287	-47,654
<i>Goods and Services</i>	21,308	3,770	-10,487
Goods	28,228	13,807	1,415
Services	-6,920	-10,037	-11,902
 <i>Primary Income</i>	 -54,151	 -42,615	 -35,857
 <i>Secondary Income</i>	 -1,541	 -1,442	 -1,310
 <b>Capital and Financial Account</b>	 33,547	 38,903	 47,114
<i>Capital Account</i>	-556	-1,110	-1,114
Acquisition / disposal of non-produced, non-financial assets	-29	-28	-32
Capital Transfers	-527	-1,082	-1,082
 <i>Financial Account</i>	 34,103	 40,013	 48,228
Direct Investment	27,654	44,511	46,063
Portfolio Investment	31,465	44,287	30,225
Financial Derivatives	-9,271	-25,828	-8,545
Other Investment	-12,546	-17,050	-18,705
Reserve Assets	-3,199	-5,908	-811
 <b>Net Errors and Omissions</b>	 <b>837</b>	 <b>1,384</b>	 <b>540</b>

Source: Australian Bureau of Statistics, Cat No. 5302. Balance of Payments and International Investment Position, June, 2013.

## **The current account**

The current account records all current transactions between a nation's residents and non-residents in goods and services, primary income and secondary income.

The goods and services or balance of trade records "transactions in items that are outcomes of production activities" (IMF, 2011: 149) and reflect exchanges between the local economy and the rest of the world. The data is typically collected from information collected from exporters and importers by the nation's customs department.

Exports and imports of goods relate to movable or tangible goods, while services are considered to be all products other than tangible goods. Services include items such as banking and insurance, transport, and export education. While items bought by tourists while on holiday may be tangible, all such expenditure is recorded as services under the IMF conventions used.

Primary income (IMF, 2011: 183):

... represents the return that accrues to institutional units for their contribution to the production process or for the provision of financial assets and renting natural resources to other institutional units.

There are two categories of primary income accounted for:

- Income that is associated with the production process, for example, wages paid, taxes and subsidies on production. If a resident is paid for labour by a non-resident then primary income is deemed to have been earned and vice versa.
- Income that is associated with the ownership of financial assets, for example, dividends and interest.

These flows are accounted for in the primary account if they are current. You will appreciate that they impact on the measure of national income in the national accounts.

The secondary income account relates to current transfers between residents and non-residents, which do not add to national income, but rather, involve redistributions of income between nations. There is nothing of economic value that is exchanged in return for a secondary income transfer.

Some of the typical secondary income account transactions include personal transfers (remittances to or from overseas), charitable contributions, social benefits (such as, pension payments to or from abroad), and current taxes on income and wealth.

Economists are often focused on the current account because of the transactions it records are of direct relevance to the determination of national income. Our earlier discussions about the sectoral balances and the income-expenditure determination all explicitly considered the current account of the balance of payments.

Exports (injections) and imports (drains) are key components of aggregate demand.

## **The capital account and financial account**

While the current account of a nation tends to focus on transactions with the rest of the world, which impact on the measurement of national output and income, the capital account capital is the financial side of these transactions.

What would happen if a nation exported more than they imported? Ignoring the primary and secondary accounts for the moment, the net outflow of real goods and services would be accompanied by accumulating financial claims against the rest of the world. This is because the demand for the nation's currency to meet the payments necessary for the exports would exceed the supply of the currency to the foreign exchange market to facilitate the import expenditure.

How might this imbalance be resolved? There are a number of ways possible. A most obvious solution would be for foreigners to borrow funds from the domestic residents. This would lead to a net accumulation of foreign claims (assets) held by residents. This item would be recorded in the capital account as a debit because it enhances the capacity of non-residents to make transactions in the local economy.

Another solution would be for non-residents to draw down local bank balances, which means that net liabilities to non-residents would decline.

The capital account thus records the:

... credit and debit entries for non-produced nonfinancial assets and capital transfers between residents and non-residents. It records acquisitions and disposals of non-produced nonfinancial assets, such as land sold to embassies and sales of leases and licenses, as well as capital transfers, that is, the provision of resources for capital purposes by one party without anything of economic value being supplied as a direct return to that party. (IMF, 2011: 9).

The financial account is a balancing account, recording the "net acquisition and disposal of financial assets and liabilities" (IMF, 2011: 10).

If we add the current and capital account together then the result "represents the net lending (surplus) or net borrowing (deficit) by the economy with the rest of the world" (IMF, 2011: 10).

This difference is the net balance of the financial account, which details the funding of the net lending or borrowing from non-residents.

### 16.3 Essential Concepts

Before we consider a more complex income and expenditure model (to incorporate exchange rates) we need to understand the basic nomenclature.

The following essential concepts are used in open economy macroeconomics:

- Nominal exchange rates.
- Foreign exchange markets.
- Exchange rate determination mechanisms – fixed and flexible.
- Real or effective exchange rates, unit labour costs and competitiveness.

We will consider the history of exchange rate systems in a later section of this Chapter.

#### Nominal exchange rate ( $e$ )

The nominal exchange rate ( $e$ ) is the number of units of one currency that can be purchased with one unit of another currency. There are two ways in which we can quote a bi-lateral exchange

rate. Consider the relationship between the Australian dollar (\$A) and the United States dollar (\$US).

We might be interested in knowing the amount of Australian currency that is necessary to purchase one unit of the US currency (\$US1). In this case, the \$US is what we call the **reference currency** and the other currency is expressed in terms of how much of it is required to buy one unit of the reference currency. So  $\$A1.25 = \$US1$  means that it takes \$1.25 of Australian currency to buy one \$US.

Alternatively,  $e$  can be defined as the amount of US dollars that are needed to buy one unit of Australian currency (\$A1). In this case, the \$A is the reference currency. So, in the example above, this is written as  $\$US0.80 = \$A1$ . Thus if it takes \$A1.25 to buy one \$US, then \$US0.80 is required to buy one \$A.

It is clear that the quotation under the first alternative with the US dollar as the reference currency is the inverse of the second alternative. But to understand exchange rate quotations you must know which is the reference currency. In this Chapter we use the second convention so  $e$  is the amount of foreign currency, which is required to buy one unit of the domestic currency. This is typically how the exchange rate is quoted in the Australian media.

**$e$  is the amount of \$US which is required to buy one unit of the domestic currency (\$A)**

#### Change in $e$ – appreciation and depreciation

Imagine that an Australian resident wishes to buy a product from a USA supplier who quotes the current US price as \$US36; and the \$US-\$A parity is currently at \$0.80. Then the equivalent Australian price is \$A45 (divide the foreign price by the nominal exchange rate). This situation is shown in the first row of Table 16.2.

What happens if the nominal exchange rate falls to 0.60 (as shown in the second row of Table 16.2)? This means that instead of 80 cents US being required to purchase one \$A only 60 cents US is required.

So, **a fall in  $e$  means that the \$A has depreciated** – \$1A (the reference currency) is worth less in terms of foreign currency. In the example shown in Table 16.2, this would mean that the price of the product from the USA would now be equal to \$A60 (\$US36 divided by 0.60).

Thus, even though the quoted US dollar price for the product remains unchanged, the local price equivalent is now higher when the nominal exchange rate depreciates.

The example shows that a depreciation of the \$A leads to:

- Foreign goods being more expensive in terms of their \$A prices, and, other things equal, this should lead to a **fall in the quantities of imports demanded**.
- The prices in \$US that foreigners that have to pay for our goods being lower, for given \$A prices of our goods. Other things equal, this should lead to a **rise in the quantities of exports demanded**.

Now, assume that the USA-Australian parity rises from 0.80 to \$1.00. This means that we now need \$1US to purchase one \$1A. So, given our exchange rate definition, **a rise in  $e$  means that the \$A has appreciated**.

In the example shown in Table 16.2, this would mean that the price of the product from US would now be equal to \$A36 (1 times \$US36).

The example shows that an appreciation of the \$A leads to:

- Cheaper foreign goods in terms of their \$A prices, and, other things being equal, this should lead to a **rise in the quantities of imports demanded**.
- Higher prices in \$US which foreigners will have to pay for our goods, for a given \$A prices of our goods. Other things being equal, this should lead to a **fall in the quantities of exports demanded**.

**Table 16.2 Comparison of international prices**

e \$US to \$UA1	Foreign Price \$US	Local Price Equivalent (\$A)	$\Delta$ \$A compared to Starting Value
0.80	36	45	Starting value
0.60	36	60	Depreciation
1.00	36	36	Appreciation

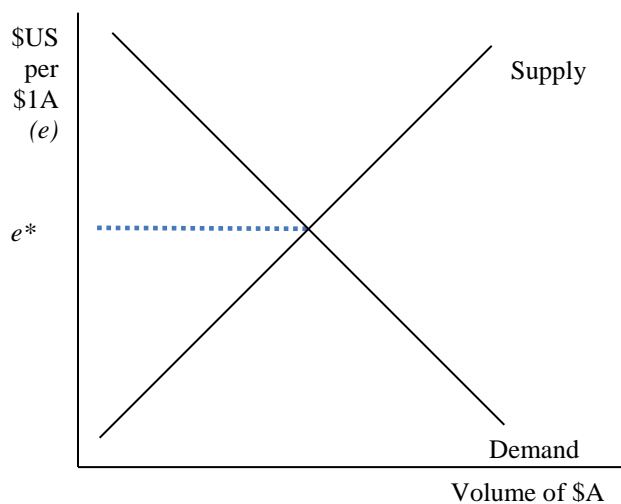
**What determines the exchange rate?**

Exchange rates are determined by the supply of and the demand for currencies in the world foreign exchange markets, which could be the local bank foreign currency desk or elsewhere, like a train station kiosk in a city where travellers meet.

Sometimes we refer to **foreign exchange** in jargon as **forex**. The supply of and demand for currencies are in turn linked to trade and capital flows between countries.

In Figure 16.1 we consider the foreign exchange market for the \$A and the \$US. In reality, many currencies are traded in the foreign exchange market.

**Figure 16.1 A simple bi-lateral foreign exchange market**



Consider the supply of Australian dollars to the foreign exchange market. When Australian residents buy foreign goods (imports), buy foreign assets or lend abroad, they need to purchase the relevant foreign currencies in which the transaction is denominated. To buy the currency they desire, they supply \$As in exchange.

On the demand side, when foreigners buy Australian goods and services (exports) and/or Australian financial assets they require \$A. They purchase them in the forex market by supplying their own currency in exchange.

Like any market-determined price,  $e$  is in equilibrium (at  $e^*$ ) when supply equals demand.

If there is an excess demand for \$A ( $D > S$ ) then there is pressure for the \$A to appreciate in price relative to other currencies. As noted, an appreciation means that one unit of the reference currency (\$A) buys more US dollars, that is  $e$  rises.

If there is an excess supply of \$A, the \$A depreciates and one unit of the reference currency (\$A) buys less US dollars, so  $e$  falls.

These changes in  $e$  resolve the supply and demand imbalance. In the case of a depreciation in the Australian dollar, the foreign price of Australian exports is now lower (less \$US required to purchase a given \$A priced good), and with export demand varying inversely with price (by assumption), the demand for exports and hence \$As rises.

Assuming a fixed import price in the foreign currency, the \$A price of imports has risen which reduces the quantity demanded.

While most currencies float freely against each other, at times the central bank will enter the foreign exchange markets as a buyer or seller of the local currency as a means of influencing the parity determined in that market. This is called **Official Intervention**.

What happens to the total \$A value of imports when the exchange rate depreciates depends upon what economists term the **price elasticity of demand**. Price elasticity is defined as the responsiveness in percentage terms of quantity to price changes.

When demand falls less in percentage terms than the price rises, we consider the demand for the commodity to be inelastic. Total revenue (or spending) will rise in this case.

When demand falls more in percentage terms than the price rises, we consider the demand for the commodity to be elastic. Total revenue (or spending) will fall in this case.

When price and quantity change by the same proportion the demand for the commodity has a unitary elasticity and total revenue (or spending) does not change.

Should the demand for imports be price inelastic (less than one), then the quantity demanded (volume) falls by a smaller percentage than the \$A rise in price and total import spending in \$A would increase.

However, if the price elasticity of demand for imports is greater than one, then the percentage decline in demand (volume) more than offsets the percentage gain in price and so total import spending in \$A falls. The demand and supply schedules for \$A that are shown in Figure 16.1 are consistent with both the demand for Australian exports and Australians' demand for imports being elastic.

The circumstances under which the trade balance unambiguously improves following a depreciation is referred to as the **Marshall-Lerner Condition**. It states that **net exports will improve following a depreciation as long as the sum of the export and import price elasticities exceeds unity**. You do not have to learn the proof underpinning this condition.

In summary, if the Marshall-Lerner condition is satisfied:

- An excess supply of \$A in the foreign exchange market leads to a depreciation ( $e$  falls) and a rise in net exports. This will reduce the excess supply of \$A in the foreign exchange market.
- An excess demand for \$A in the foreign exchange market leads to an appreciation ( $e$  rises) and a decline in net exports. This will reduce the excess demand for \$A in the foreign exchange market.

Another component of the current account is net income, which results from the foreign ownership of domestic assets and vice versa. We consider this component of the current account in more detail later in the Chapter.

This pattern of ownership of assets gives rise to a net flow of dividend and interest payments. If the net flow is positive then national income rises, other things being equal. If the net flow is negative then national income falls, other things being equal.

In Australia's case, the net income flows on the current account are negative. In this case, a depreciation in the \$A can lead to improved net income if the interest payments or dividends are denominated in a foreign currency. The gain through this part of the current account would supplement any gains that are made as a result of the impact of the depreciation on the trade balance.

For simplicity, we shall ignore the possible impact on net income and assume that, through the satisfaction of the Marshall-Lerner condition, a depreciation of the domestic currency not only improves the trade balance, but also the current account balance.

We can define three trade balance outcomes:

- The trade balance is in deficit if the local currency value of its exports is less than the local currency value of its import spending.
- The trade balance is in surplus if the local currency value of its exports is greater than the local currency value of its import spending.
- The trade balance is in balance if the local currency value of its exports is equal to the local currency value of its import spending.

Take Australia, as an example. A trade deficit for Australia means that increasing quantities of Australian dollars are being accumulated by non-Australian residents. In return, non-Australian residents have supplied goods and services (imports) to Australian residents.

Clearly, foreigners have allowed Australia to run a trade deficit because they preferred to accumulate financial assets denominated in Australian dollars. The alternative would have been to spend the Australian dollars they acquired through their exports to buy Australian goods and services (that is, to buy Australian imports).



Had foreigners used their export income, which is denominated in \$A to purchase other goods and services from Australia, then there would have been a trade balance.

A trade deficit thus means that foreigners are increasing their nominal savings (which in this case manifests as Australian dollar denominated financial assets).

### International competitiveness

In the previous section we learned that an appreciation (depreciation) of a nation's exchange rate leads to foreign goods becoming cheaper (dearer) in terms of the local currency, which should lead to a rise (fall) in the quantity of imports demanded, other things equal.

Further, an appreciation (depreciation) of a nation's exchange rate means that foreigners have to pay higher (lower) prices in their currency for locally-produced goods, which should lead to a fall (rise) in the quantity of exports demanded, other things equal.

These conclusions, however, only focus on one element of the competitiveness of a nation's goods and services in international trade – the nominal exchange rate,  $e$ .

But to really answer the question – are local goods and services becoming more or less competitive with respect to goods and services produced overseas, we have to relax the 'other things equal' assumption and consider the domestic and foreign inflation rates.

This leads us to define a new concept – the **real exchange rate** that depends on two factors:

- Movements in the nominal exchange rate,  $e$ ; and
- Relative inflation rates (domestic and foreign).

There are also non-price dimensions to competitiveness, including quality and reliability of supply, which are assumed to be constant at this stage.

We define the ratio of prices ( $P_w$ ) to domestic prices ( $P$ ) as  $P_w/P$ , which we call a relative price because it expresses the foreign price level relative to the domestic price level. We assumed that  $P_w/P$  was constant when we analysed movements in the nominal exchange rate in the previous section.

If the nominal exchange rate ( $e$ ) is fixed, then we can conclude:

- If  $P_w$  is rising faster than  $P$ , then local goods are becoming relatively cheaper than foreign goods; and
- If  $P_w$  is rising slower than  $P$ , then local goods are becoming relatively more expensive than foreign goods.

The inverse of the relative price ratio, namely ( $P/P_w$ ) measures the ratio of export prices to import prices and is known as the **terms of trade**.

### The real exchange rate

Movements in the nominal exchange rate and/or the relative price level ( $P_w/P$ ) provide information about movements in the relative trading competitiveness between nations. The real exchange rate measures the combined impact of these variables and is used to measure our competitiveness in international trade.

The real exchange rate ( $R$ ) is defined as:

$$(16.1) \quad R = (P_w/e)/P = (P_w/Pe)$$

where  $P$  is the domestic price level specified in local currency (say, \$A), and  $P_w$  is the foreign price level specified in foreign currency units (say \$US).

The real exchange rate is the ratio of prices of goods abroad measured in \$A ( $P_w/e$ ) to the \$A prices of goods at home ( $P$ ). So the real exchange rate,  $R$  adjusts the nominal exchange rate,  $e$  for the relative price levels.

To understand this better consider the following example. Assume that  $P = \$A12$  and  $P_w = \$US10$ , and  $e = 0.8$ . Remember a quotation of  $e = 0.8$  means that it takes \$0.80US to purchase one unit of Australian currency (that is, \$A1).

So  $P_w$  divided by  $e$  takes the foreign price expressed in foreign currency units and converts it into an equivalent Australian dollar price at the current exchange rate. The numerator and denominator are then in like units – in this case, Australian dollars – and so the movements are unambiguous.

In this case  $R = (\$US10/0.8)/12 = 1.042$ . The \$US10 translates into \$A12.50 and the US produced goods are more expensive than those in Australia by a ratio of 1.042, or 4.2 per cent higher.

A rise in the real exchange rate can occur if:

- The nominal  $e$  depreciates; and/or
- $P_w$  rises more than  $P$ , other things equal.

**We consider a rise in the real exchange rate to signal a nation (in the above example, Australia) has increased its international trade competitiveness and this should lead to an increase in local exports and reduce local imports.**

A fall in the real exchange rate can occur if:

- The nominal  $e$  appreciates; and/or
- $P_w$  rises less than  $P$ , other things equal.

We consider a fall in the real exchange rate to signal a nation's international trade competitiveness has fallen and this should lead to a fall in local exports and a rise in local imports.

In Chapter 8, we considered the factors that might impact on the price level of a nation. In particular, if prices are set on unit labour costs, then the way to decrease the price level relative to the rest of the world is to reduce unit labour costs faster than everywhere else or compress profit margins.

With constant profit margins, if the rate of growth in wages is faster than labour productivity growth, then unit labour costs rise and vice-versa. As we saw in Chapter 8, the real wage is a composite of the nominal wage determined in the labour market as a result of bargains between workers and employers and the price level, which is determined by firms in the goods and services market.

The problem is that if a nation attempts to improve its international competitiveness by cutting nominal wages in order to reduce real wages and in turn, unit labour costs, it not only undermines aggregate demand, but also may damage its productivity performance.

If, for example, workforce morale falls as a result of cuts to nominal wages, it is likely that industrial sabotage and absenteeism will rise, undermining labour productivity.

Further, overall business investment is likely to fall in response to the extended period of recession and wage cuts, which erodes future productivity growth. Thus there is no guarantee that this sort of strategy will lead to a significant fall in unit labour costs, and, if it were to be successful, there are likely to be adverse consequences for aggregate demand.

There is robust research evidence to support the notion that, by paying high wages and offering workers secure employment, firms reap the benefits of higher productivity and the nation sees improvements in its international competitiveness as a result.

## 16.4 Aggregate Demand and the External Sector – Revisited

In Chapter 7 we developed the income-expenditure framework to explain the how the level of real GDP (and real national income) are determined. The Chapter used the national accounting concept of GDP at constant rather than current prices as our measure of economic activity.

In our income-expenditure framework in Chapter 7 we expressed the flow of total expenditure in any period as the sum of the following sources of spending:

- Consumption by households or persons ( $C$ )
- Investment spending by firms ( $I$ )
- Government spending ( $G$ )
- Export spending by foreigners ( $X$ ) minus import spending by domestic residents ( $M$ ), which we denote as net exports ( $NX$ ) =  $(X - M)$ .

From Chapter 7, we know that the equilibrium level of real national income ( $Y$ ) is determined by aggregate demand (as long as prices remain unchanged), such that:

$$(16.2) \quad Y = E = C + I + G + (X - M)$$

In this Section, we will develop a more detailed account of net exports,  $(X - M)$ , to take into account the influence of the real exchange rate and international competitiveness, which we discussed in the last section.

In Chapter 7, our treatment of the determinants of net exports was deliberately very simple. We assumed that exports ( $X$ ) were given in any particular period and determined by national income in the rest of the world, which is beyond the influence of the local economy in question.

We also assumed that a nation imports a fixed proportion of every dollar of national income. We called that proportion the **marginal propensity to import ( $m$ )** and defined it as the extra import spending that occurs as a result of a dollar increase in national income.

In the previous section we learned that movements in the real exchange rate, which is a summary measure of international competitiveness, influences net exports.

We established that the higher is the value of the real exchange rate, the cheaper are locally-produced goods and services to foreign buyers, which means that they will purchase more of them. In other words, exports rise when the real exchange rate rises.

Further, the higher is the value of the real exchange rate, the more expensive are foreign-produced goods and services for local buyers, which means they will purchase less of them. In other words, imports fall when the real exchange rate rises.

There are many other factors in the real world that determine the demand for a nation's exports and the demand by residents for imports, which we abstract from here, in order to focus on the most significant determinants.

We are also abstracting from adjustment responses, which are common in international trade. So a rise in the real exchange rate might only influence future exports once existing export contracts, which tend to be multi-year, expire. In the following analysis we are simplifying by assuming that the response between movements in the real exchange rate and changes in the flows of export and import spending are within the current period.

Let us consider exports. We now assume that the level of exports in any period is determined by the real exchange rate ( $R$ ) and world income ( $Y_w$ ) and we write this in the following way:

$$(16.3) \quad X = \lambda Y_w + \theta_X R$$

which might appear at first inspection to be daunting but if you apply the techniques that we developed in the *Appendix (Methods, Tools and Techniques)*, you will grasp the meaning of this equation fairly easily.

In Equation (16.3) the Greek letters (coefficients) next to world income ( $Y_w$ ) and the real exchange rate ( $R$ ) measure how responsive export spending is to changes in these variables. The coefficient,  $\lambda$  measures how much a nation's export income rises as a result of a rise in world income. If you think about it,  $\lambda$  is, from the perspective of the rest of the world, its marginal propensity to import.

Similarly, the coefficient  $\theta_X$  measures the responsiveness of exports to changes in the real exchange rate. Remember we are simplifying by assuming this response is immediate and exhausted in the current period.

We specify  $\lambda$  and  $\theta_X$  to be positive. From our theoretical exegesis, we conjecture that when world income and the real exchange rate rises, we expect exports to increase. Thus both terms in Equation (16.3) have positive signs.

From the imports side, we assume that the level of imports that a nation purchases depends on both real national income ( $Y$ ) and the real exchange rate ( $R$ ). Thus:

$$(16.4) \quad M = mY - \theta_M R$$

The coefficient  $m$  is the marginal propensity to import and we know its value lies between 0 and 1. We conjecture that the impact of real exchange rate will be negative, which means that when the real exchange rate rises and the nation becomes more competitive and foreign goods become more expensive in local currency terms, import spending falls. Then with the coefficient  $\theta_M$  assumed to be positive, the responsiveness of imports to changes in the real exchange rate, takes the form  $-\theta_M R$ .

If we assume that domestic and foreign price levels are both constant, then movements in the real exchange rate,  $R$ , are mirrored in the nominal exchange rate,  $e$ . In other words, we could simply substitute the nominal exchange rate ( $e$ ) for the real exchange rate ( $R$ ) in Equations (16.3) and (16.4) without loss of understanding.

Thus Net Exports ( $NX$ ) depends on local real GDP, world real GDP and the real exchange rate, (see Equation 16.8), where the latter impact is the net result of the impact of the real exchange rate on exports and imports, respectively.

## 16.5 Trade in Goods and Services, Product Market Equilibrium and the Trade Balance

### National income equilibrium with trade

In this Section, we continue to assume that both  $P_w$  and  $P$  are constant which means that domestic and foreign firms respond to increases in real aggregate demand by increasing real output rather than prices.

Spending on domestic goods determines real output and income. Total spending on domestically produced goods and services is equal to total spending by domestic residents minus their spending on imports plus foreign demand for exports.

Referring back to Chapter 7 we have the following behavioural equations, which comprise our theory of aggregate demand:

$$(16.5) \text{ Consumption function} \quad C = C_0 + cY_d = C_0 + c(I - t)Y$$

$$(16.6) \text{ Investment function} \quad I = I_0 - bi$$

$$(16.7) \text{ Government spending} \quad G$$

$$(16.8) \text{ Net exports} \quad NX = \lambda Y_w - mY + \theta e$$

where  $\theta = \theta_X + \theta_M$ . Note that  $\theta$  is the net impact of changes in the real exchange rate, here expressed as the nominal exchange rate because we assume that  $P_w/P$  is constant.

We can substitute the individual behavioural equations into the equilibrium income Equation (16.2) such that:

$$(16.9) \quad Y = E = C_0 + c(I - t)Y + I_0 - bi + G + \lambda Y_w - mY + \theta e$$

which, if we refer back to the way we simplified Equation (7.15a) to get Equation (7.15c), we can write:

$$(16.10) \quad Y = (1/[1 - c(I - t) + m]) * [C_0 + I_0 - bi + G + \lambda Y_w + \theta e]$$

This expression for equilibrium national income tells us that real GDP ( $Y$ ) will be the sum of all the expenditure terms that do not directly depend on national income (those in the right-hand bracketed expression) times the multiplier (the first left-hand side term).

We can use this expression to study what happens to national income when one of the terms on the right-hand side of Equation (16.10) changes.

### The net exports function

Equation (16.8) represents net exports ( $NX$ ) in terms of world income ( $Y_w$ ), domestic national income ( $Y$ ) and the exchange rate ( $e$ ). The latter term was the net impact on aggregate spending of an exchange rate change taking into account the impacts on the individual components, namely exports and imports.

Figure 16.2 shows the Net Exports function expressed in terms of national income, so we assume that both the shift variables ( $Y_w$  and  $e$ ) are constant. At zero national income, there would be an overall trade surplus.

Then as national income rises the trade balance will be negatively sloped – that is, move from a surplus to a deficit. This is because as national income rises, imports rise via the marginal propensity to import ( $m$ ). At point  $Y_0$ , the net exports are balanced (exports equals imports).

To summarise, to the left of  $NX = 0$ , there is a trade surplus because, for a given level of exports, the lower level of income leads to a smaller expenditure on imports. To the right of  $NX = 0$ , there is a trade deficit, because at the higher level of income, imports are higher, relative to the fixed level of exports.

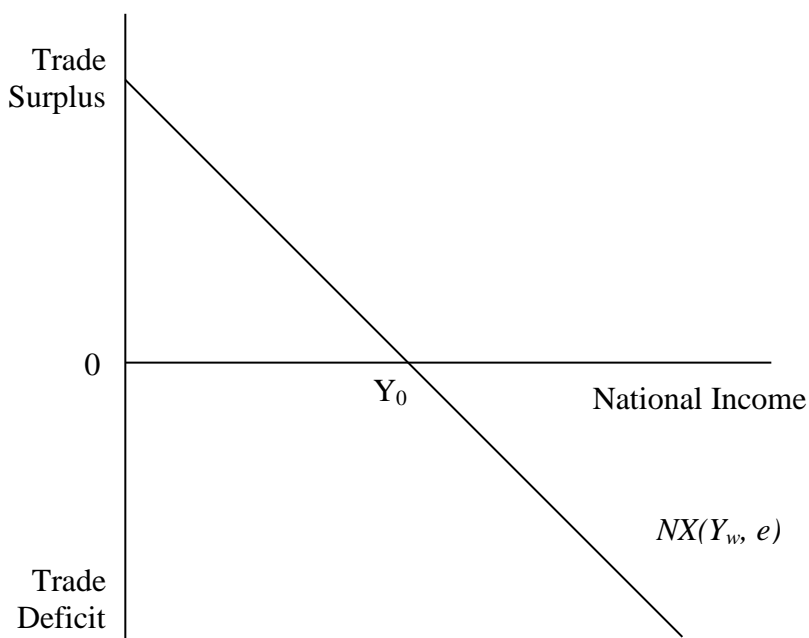
As noted, the Net Exports function is drawn against national income with the other variables in Equation (16.8),  $Y_w$  and  $e$ , being held constant. If either  $Y_w$  or  $e$  changed, the  $NX$  function would move up or down, depending on which variable changed and our assumptions about the parameters  $\lambda$  (for a change in world income) and  $\theta$  (for a change in the nominal exchange rate).

We have assumed that  $\lambda > \theta$  so that an increase in world income levels boosts our exports and so net exports rises (other things being equal). Further, in our discussion we assumed that the net impact on net exports of a change in the exchange rate, which is measured by  $\theta$  is positive. This means that exchange rate depreciation improves the trade balance.

In other words we conclude that:

- If world income rises (falls) the Net Exports function will move out (in) with an unchanged slope.
- If the exchange rate depreciates (appreciates) the Net Exports function will move out (in) with an unchanged slope.

**Figure 16.2 Net exports as a function of real national income**



The slope of the Net Exports function is determined by the marginal propensity to import. The larger is the marginal propensity to import the steeper the slope. This is because the larger is the marginal propensity to import,  $m$ , the greater is the leakage from the expenditure stream per extra dollar of national income generated and the more quickly the trade balance moves into deficit at each income level.

We can use this understanding and that provided by the equilibrium national income expression (Equation 16.10) to study what happens to national income when world income and/or the exchange rate change.

### **The impact on national income and net exports of a change in world income**

We have seen (from Equation 16.10) that the equilibrium level of national income (and real GDP) is dependent on the level of domestic autonomous expenditure ( $C_0 + I_0 + G$ ); the interest-rate sensitive component of investment (bi); the level of world income ( $Y_w$ ) and the real exchange rate (simplified to  $e$  because relative prices are fixed).

What happens to national income if either the level of world income or the exchange rate changes? In the analysis that follows we assume that the central bank supplies the monetary base at a constant discount rate, so that the interest rate charged by the private banks can be treated as constant.

Figure 16.3 reintroduces the Aggregate Demand function from Chapter 7. Recall Figure 7.5 which showed that a rise in autonomous spending would lead to the Aggregate Demand function shifting up in parallel fashion (the shift in the intercept being measured by the change in autonomous spending). It is straightforward to generalise that insight into the current context by noting that the autonomous expenditure components are just the right hand terms in Equation (16.8), which interact with the spending multiplier.

In Figure 16.3 the trade balance line is denoted as  $NX = 0$ . With the other determinants constant (world income and the exchange rate), we noted that there was one level of national income where the trade account would be balanced (where imports equal exports). We denoted that level of national income as  $Y_0$  in Figure 16.3.

We can translate that knowledge into Figure 16.3 by drawing an  $NX_0 = 0$  line which corresponds to national income level  $Y_0$ . All national income levels below  $Y_0$  will result in trade surpluses because imports will be lower than exports, other things equal. All national income levels above  $Y_0$  will generate trade deficits, because imports will be higher than exports, other things equal.

In our discussion of Figure 16.3, we also noted that the  $NX$  line would shift up if world income increased and down if world income decreased. In the context of Figure 16.3, this means that the  $NX = 0$  line will shift to the right if world income rises and to the left if world income falls.

The reason is simple. Start from national income level  $Y_0$ , which initially coincides with a trade balance  $NX_0 = 0$ . If world income rises, then at that national income level ( $Y_0$ ), exports will be higher than before and so the trade account would be surplus. Trade balance would require a boost to imports, which, in turn, would occur at higher levels of national income. We denote the new trade balance  $NX_1 = 0$  at  $Y_1$ .

But we also know that a rise in world income will lead to the rest of the world importing more goods and services from the domestic economy which means exports rise. As we learned in

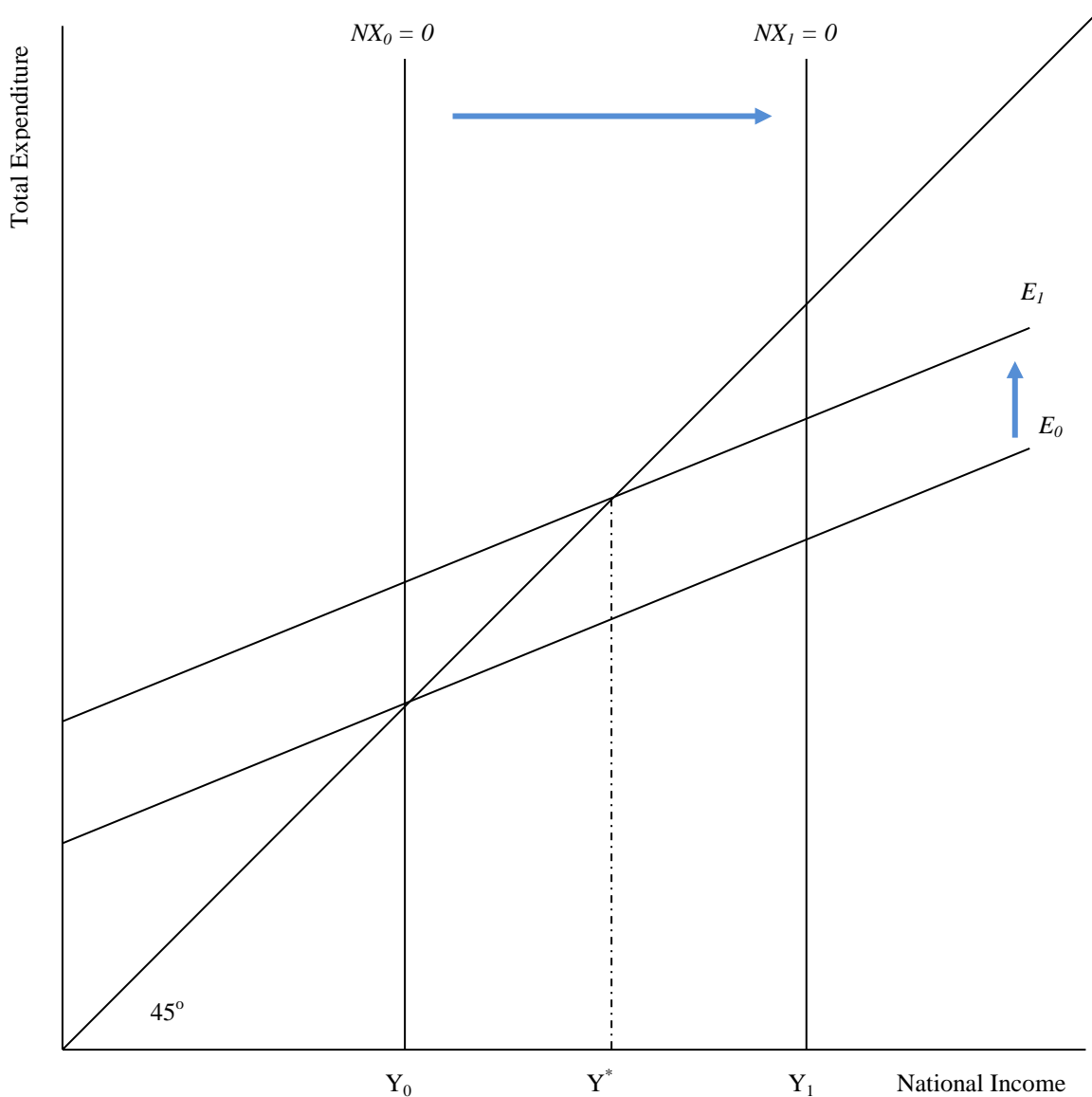
Chapter 12, if any of the expenditure components rise, then the aggregate demand function shifts upwards.

Figure 16.3 denotes an initial point  $Y_0$ , which just by coincidence is also a point of external balance. Aggregate demand is at  $E_0$ .

The rise in exports pushes the aggregate demand function upwards to  $E_1$  and the new national income equilibrium occurs at  $Y^*$ . At this point, real GDP is higher, national income is higher and we could show that employment would be higher and unemployment lower, once the cyclical labour supply adjustments that we studied in Chapter 9 were exhausted.

You will also note that the economy now has an external surplus, being left of the  $NX_1 = 0$  line (see Advanced Material for an explanation).

**Figure 16.3** Equilibrium national income with a change in world income





By way of summary, a rise in world income induces a rise in foreign purchases of the economy's exports, in the same way that the local economy's import demand will be stimulated if its national income rises.

This has three effects:

- The aggregate demand line ( $E_0$ ) shifts upwards by the initial injection of aggregate demand from exports ( $\Delta X$ ), giving rise to a new level of equilibrium national income  $Y^*$ . The increase in equilibrium national income is given by  $1/[1 - c(1 - t) + m]$  times  $\Delta X$ , given the constant interest rate. This is just a specific example of the general result that the aggregate demand schedule shifts in response to changes in autonomous spending;
- The  $NX = 0$  line shifts to the right from  $NX_0 = 0$  to  $NX_1 = 0$ .
- Imports rise too because the shift in the aggregate demand line (to  $E_1$ ) means that income levels are higher. But the shift in the  $NX = 0$  line is greater than the increase in equilibrium income. Thus  $NX$  rises, but by less than the rise in exports. This result holds even if domestic interest rates rise.

You should be able to work out what would happen if there were a world recession and world income levels fell. In this case, the local economy would experience a drop in aggregate demand because exports will be lower than previously and national income will fall.

The  $NX = 0$  line also shifts to the left and if the economy was, for simplicity, initially experiencing balanced trade, then the new equilibrium income level will be associated with a trade deficit (the fall in exports is greater than the fall in imports).

## 16.6 Currency Crises

Many developing countries have currency sovereignty, which means they can enforce tax liabilities in the currency that the government issues. It doesn't matter if other currencies are also in use in those countries, which is common. For example, the \$US will often be in use in a less developed country alongside the local currency and be preferred by residents in their trading activities.

But, typically, the residents still have to get local currency to pay their taxes. That means the government of issue has the capacity to spend in that currency.

The general principle thus remains – as long as there are real resources available for use in a less developed country, the government can purchase them using its currency power.

In particular, this concept of real fiscal space extends to the millions of people who are unemployed in less developed countries. Given there is no market demand for their services, the government in each country can easily purchase these services with the local currency without placing pressure on labour costs in the country.

The investment in these programs is measured by the real resources that they consume relative to not undertaking the initiative. These resources extend to imports and many less developed countries have to import food for basic subsistence.

Will this investment undermine the current account and introduce inflation as the current account depreciates due to a widening trade deficit? All open economies are susceptible to balance of payments fluctuations.

As we have learned in earlier Chapters, these fluctuations were problematic for nations running external deficits under the fixed exchange rate, convertible currency regimes (for example, the Bretton Woods scheme) because the government was forced to keep the domestic economy in a depressed state to keep the imports down, so that the central bank could maintain the parity without losing its foreign currency reserves.

For a flexible exchange rate economy, the exchange rate does the adjustment. **There is no consistent evidence that fiscal deficits create catastrophic exchange rate depreciations in flexible exchange rate countries.**

Any increase in domestic spending will push up import demand. In particular, growth in private capital formation is likely to be more import intensive because most less-developed countries import capital.

Well-targeted government spending can create domestic import-competing activity. For example, Job Guarantee workers could produce goods and services that a nation might normally import including processed food products.

Moreover, if there are systems in place to promote the skill development of the labour force, a fully employed economy is likely to attract foreign direct investment under a stable political system.

So while the current account might move into deficit as the economy grows, which means the nation is sacrificing less real resources away in return for real imports from abroad, the capital account would move into surplus. The overall net effect is not clear.

Finally, a depreciated currency stimulates local employment because imported goods become more expensive and exports become cheaper with the distributional impacts of these changes likely to be felt more by the middle and higher income groups than the poorer groups when luxury imported goods become more expensive.

There are likely to be once off changes in the exchange rate as the economy adjusts to the higher growth path and thus should not be a source of on-going inflationary pressure.

It is true that currency depreciation for a nation that is wholly dependent on imported food can be damaging. Note, this is not a balance of payments constraint as it is normally considered. It is a real resource constraint arising from the unequal distribution of resources across geographic space and the somewhat arbitrary lines that have been drawn across that space to delineate sovereign states.

In this context, a new multilateral institution should be created to replace both the World Bank and the IMF, which is charged with the responsibility to ensure that these highly disadvantaged nations can access essential real resources such as food and not be priced out of international markets due to exchange rate fluctuations that arise from trade deficits.

This agency should buy the local currency to ensure flexibility and the exchange rate does not price the population, in particular low-income households, out of food. This is a simple solution, which is preferable to forcing these nations to run austerity campaigns just to keep their exchange rate higher.

What about currency crises?

In the 1990s, there were three major currency crises in the world economy. First, the European currency crisis that followed the break-down of the Berlin Wall. Second, the Mexican currency crisis in 1992, when the peso plunged in value after capital flows, attracted by rising US interest rates, moved against Mexico. Third, there was the Asian debt crisis in 1997. We will briefly examine each episode and highlight the essential characteristics of the prevailing monetary system at the time of each, which precipitated the crisis.

Are there lessons we can learn from these relatively recent events?

### **The European exchange rate mechanism crisis – 1992**

In 1992, as the German government moved to unify the country after the breakdown of the Berlin Wall, the fiscal expansion required to improve the public infrastructure in the former East Germany, was accompanied by rising interest rates because the Bundesbank, (the German central bank) feared inflation.

The German mark appreciated significantly as a result of capital inflow attracted by the higher interest rates and net exports fell.

The problem was that the German mark was the benchmark currency in the European Exchange Rate Mechanism (ERM), against which other Western European nations fixed their currencies. This arrangement had been in place since March 1979.

Under any pegged currency arrangement, those nations pegging their currencies have to increase their own interest rates in line with increases in rates in the benchmark nation.

However, the other nations did not have a commensurate fiscal expansion to offset the damaging effects of the rising interest rates on their local economy. It became apparent that the commitment to the fixed exchange rate system (ERM) would mean rising unemployment in these nations and the associated political difficulties.

Players in the foreign exchange markets predicted that eventually the pegging nations would abandon the ERM and let their currencies depreciate against the German mark.

This led to the currencies of these nations being sold off in the foreign exchange markets, which immediately forced the nations to consider the fixed arrangements.

The impetus to the breakdown of the system, was the ‘short selling’ attack on the British pound by speculator George Soros. In foreign exchange markets, a speculator can contract to sell a currency at some future date for a predetermined price.

The contract, of-course, means that when the contract comes due, the speculator also has buy the other currency in the contract.

If the contracts are large enough they can thus have a significant impact on the value of the currency and lead other speculators to follow suit. George Soros short sold the British pound against the German mark.

Britain left the program in September 1992 after the Bank of England had spent over £6 billion selling foreign currencies in an attempt to maintain the currency within the agreed ERM limits as the speculative activity drove its price down. The British government was not prepared to increase its interest rates in line with Germany, which they considered would cause a major recession.

In this case, the speculators won, but only because these governments were intent on pegging their currencies but were not prepared to accept the monetary policy interdependence that came with the decision to peg.

### **The 1994 Mexican peso crisis**

The Mexican peso or Tequila crisis was termed “the first financial crisis of the twenty-first century” (Boughton, 2001) by the then Managing Director of the International Monetary Fund, Michael Camdessus.

The facts are well known although different economists emphasise different causes.

In the 1980s, Mexico endured a debt crisis, which led to misguided policy responses that essentially led to the 1994 currency crisis. The debt-crisis was the result of the large foreign-owned commercial banks taking on massive foreign-currency denominated floating interest rate loans. The funds were sourced from the petroleum exporting nations which were cashed up after the OPEC oil price rises in 1973-74 and later in the 1970s.

In the late 1970s, US interest rates were pushed up to deal with domestic inflation, and the private debt burden for Latin American nations became severe. The US recession in 1981 also damaged primary commodity export markets, a principal source of foreign exchange for the Latin American economies.

In 1982, Mexico announced that it could no longer service its external debts and lending to Latin American nations ceased, thwarting the normal refinancing of outstanding loans as they became due. Much of the debt was of a short-term nature.

The IMF became involved and as a price for the provision of bailout funds, the nations had to introduce sweeping free market reforms and severe fiscal austerity, consistent with the ideological position of the Fund. Widespread unemployment resulted and poverty rates rose sharply as state-owned industries were privatised, and tariff protection and welfare nets cut.

Growth returned in the early 1990s, and capital inflow, particularly from the growing US financial sector, boomed. The free market Mexican government, under guidance from the IMF also sought to make the economy as attractive as possible to the financial speculators.

The Bank of Mexico managed a fixed peso parity against the US dollar within a so-called ‘slippage regime’ between November 11, 1991 and December 21, 1994, where the rate was allowed to vary within certain daily bands.

The upshot was that the Mexican government stood ready to convert the peso into US dollars (and vice versa) at a fixed rate, which meant it always had to have sufficient reserves of US dollars to guarantee convertibility. Reliable convertibility was thought to be essential to establish the monetary credibility of the Mexican government and instil confidence into the international financial markets.

The reality was that all the risk was shifted from the foreign speculators to the Mexican government, a risk that ultimately the Mexican economy was unable to bear.

The domestic growth pushed out the current account deficit, which was being funded by the massive capital inflows. While the government, under pressure from the IMF, was running a small budget surplus, it was forced to keep issuing government debt to foreign creditors to provide financial instruments that would attract increasing capital inflow.

The capital inflows also led to an accelerating money supply and nominal demand growth began to outstrip the real capacity of the domestic economy, with rising inflation the result.

These developments meant that the peso should have depreciated but the Mexican government, under pressure from the IMF and the US government, maintained the peg even as speculators started to sell off the peso in favour of the US dollar. The Bank of Mexico stock of foreign exchange reserves (so-called 'hard' currency) started to run out and this led to further speculative attacks.

There was some political turmoil (an assassination in March 1994), which didn't help the deteriorating confidence in the Mexican economy.

The short-sighted commitment to the fixed parity by the Mexican government led them to further compound the crisis when they acceded to demands from large foreign investors (particularly Wall Street banks) to significantly increase the issuance of so-called Tesobono bonds, which were US dollar denominated, government debt instruments that insured the holder against any foreign exchange risk.

The foreign currency denominated bonds were increasingly substituted for peso denominated debt to keep the foreign investors happy. The foreign risk insurance explicit in the Tesobonos, lowered the price the government had to pay on this debt but also dramatically increased its own exposure to any depreciation in the peso.

Despite the tension within the government and between the government and central bank, the fixed parity was maintained even though it was obvious that the Government could no longer support the currency.

However, the IMF continued to argue that the Mexican policy settings were sound. At the Executive Board Meeting (November 30, 1994), the IMF was discussing the introduction of a new "short-term financing facility" for eligible nations to assist in the alleviation of short-term balance of payments pressure that might destabilise their currencies.

In the Minutes of that meeting (Minutes of the Executive Board of the IMF (EBM/94/104 – November 30, 1994), the IMF said that "all countries that have sound macroeconomic policies and do not have structural balance of payments problems would be eligible" (p14).

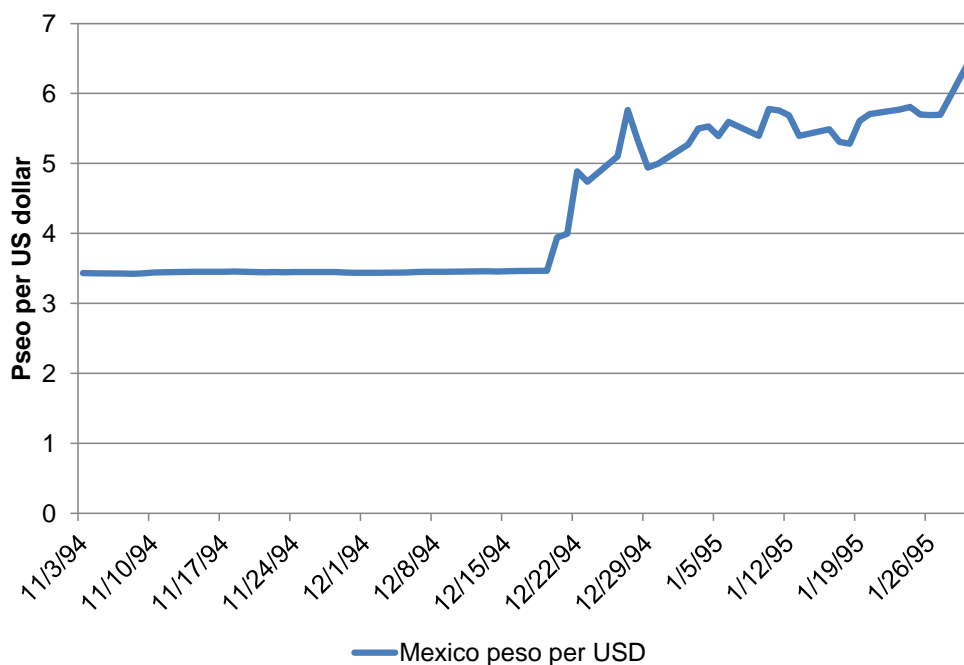
The Minutes reported that one member (Mr Kiekens from Belgium) noted that:

Countries with a genuine need for short-term Fund financing are few, If not very few ... Rare are those ... countries, which have a perfect track record and an external position that would otherwise be considered as viable and as having sound fundamentals. This is evident in that, of the three cases presented by the staff, only the Mexican case is a strong one. (IMF, 1994: 21)

However, less than a month later, the speculative outflow of pesos became too great and on December 22, 1994, the Mexican government floated the peso.

Figure 16.4 shows the evolution of the Mexican peso against the US dollar for the period November 1994 to the end of January 1995. On December 19, 1994 the peso was trading at 3.4662 per US dollar. Eight days later on December 27, 1994, the parity had risen to 5.7625 per US dollar, which represented a depreciation in the peso of 66.2 per cent.

**Figure 16.4 Mexican peso depreciation, December 1994**



The decision by the Mexican government to float the peso was forced on it because it ran out of the foreign reserves that were necessary for the central bank to maintain the peg against the US dollar because speculators were selling out the currency and driving its price down in world markets.

The short-term consequences of the depreciation were severe. The severity was linked, in part, to the government's tardiness in making the decision. The events occurred in a presidential election year and both the instability associated with the bitterly fought campaign (including the assassination of one of the leading candidates) and the reluctance of the incumbent who wanted to avoid the stigma of devaluation, meant that the peg remained in place despite massive outflows of funds.

The peg signified a sort of status for the Mexican government which had instilled a sense of confidence in the Mexican economy by becoming a member of the Organisation for Economic Cooperation and Development (OECD) and entering the NAFTA in early 1994.

They also liberalised credit and privatised the banking system, which exposed the nation to rapid capital inflow without a commensurate increase in the ability of its financial institutions to handle the risks involved in international finance.

Once they were forced to float, the response of the international markets was extreme. The investors (both foreign and Mexican), who had previously held out Mexico as the exemplar for Latin America to follow, sold off pesos in astonishing proportions in a space of a few days (December 20-22, 1993).

Given the desire to maintain the peg, the government left itself with two undesirable options as the capital outflow accelerated and the central bank foreign reserves rapidly declined.

First, they could have hiked interest rates to encourage investors to leave their funds in Mexico. The problem was that the required interest rate increases would have been so large that they would have plunged the economy into a major recession.

Second, they could have broadened the bands in which they allowed the peg to crawl which would have improved the current account a little and perhaps offset some of the mania that was creeping into foreign exchange markets about the likely depreciation of the peso.

But the consequences of that option would have inflated the debt servicing payments on foreign debt held by Mexican companies and the Government with the inevitable consequence of insolvency.

Eventually, a combination of IMF and US government assistance stabilised the financial system and placated the international investment community, which realised that the economic fundamentals were unchanged and had not justified the massive over-reaction.

In summary, the Mexican peso crisis teaches us some important macroeconomic policy lessons.

First, while a floating exchange rate may expose an economy to imported inflation in times of depreciation, the advantages in being able to stabilise domestic output and employment are significant. A nation, which pegs its currency, loses control of monetary policy and forces fiscal policy to play a passive role that becomes destructive when the currency depreciates significantly.

Second, while it is clear that a currency-issuing government can issue public debt, it is imperative that any liabilities it does issue are denominated in its own currency and assume no foreign exchange risk by way of indexing or insurance arrangements.

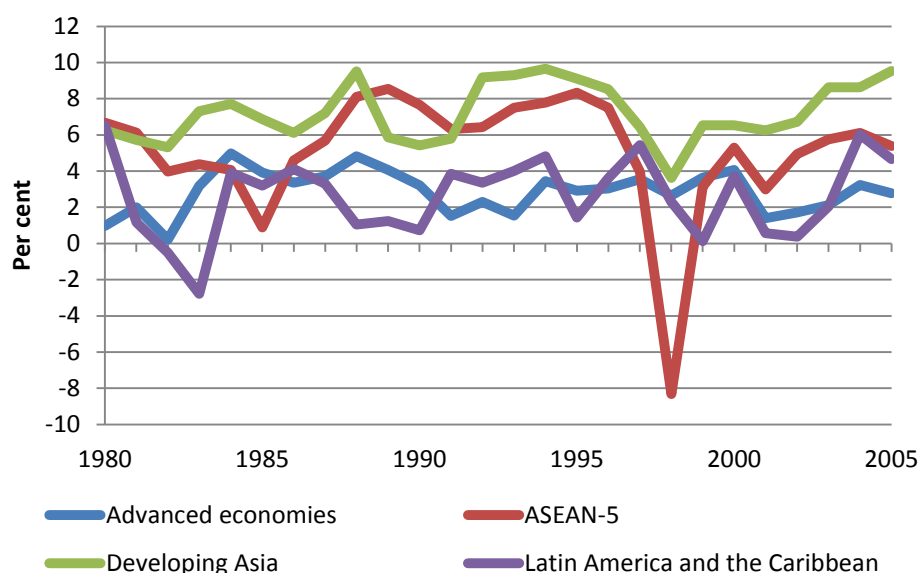
### **The South East Asian debt crisis 1997**

The last major currency crisis in the 1990s began in 1997 in the South East Asian nations of Thailand, Malaysia, the Philippines, Singapore, Indonesia and later, spread to the industrialised East Asian nation of South Korea.

In the two decades leading up to the crisis, the SE Asian nations had attracted large capital inflows and grew rapidly. The period of rapid growth, which started in the late 1980s was accompanied by high private saving ratios and strong investment. Further, inflation was low and the governments were largely running fiscal surpluses.

Figure 16.5 compares the annual real GDP growth rates for several blocks of nations including the ASEAN-5, which comprises: Indonesia, Malaysia, Philippines, Thailand, and Vietnam. You can consult the IMF World Economic Outlook database for full descriptions of the other groupings.

**Figure 16.5 Real GDP growth in South East Asia, 1980-2005, per cent per annum**



Source: IMF World Economic Outlook Database.

They were considered by the IMF and the World Bank to be models of sustainable development. The expression ‘The Asian Miracle’ was used to describe the rapid growth and rising living standards, particularly in the so-called four ‘Asian Tiger’ economies of Hong Kong, Taiwan, Singapore and South Korea.

The multilateral organisations mistakenly considered the rapid growth was the product of fiscal rectitude and free market dynamics, which they considered diverted resources to their highest value use and allowed these nations to be internationally competitive.

However, the reality was different. The Asian nations built their growth strategy, which began in the late 1960s (with Japan), on a mix of industrialisation, mercantilism and strong state-imposed industrial policies.

In nations such as South Korea, the state played a major role in the development process and defied the advice offered by free market economists at the IMF and the World Bank with respect to their development strategy. The latter group considered trade-led growth would only come if a nation exploited its comparative advantage.

However, the South Korean government selected and supported several key sectors to be their growth engines, despite none of them having any relative resource advantage (for example, chemicals). The textiles sector in Korea had indicated that a chemicals industry would support its own development.

The governments in fact, interfered with the ‘market’ in many ways. They provided credit at below market prices to targeted sectors. Substantial tax breaks were given to firms to increase profits and investment. Protection was provided to local firms against import competition. The state invested heavily in public research and development and shared the results with industry.

By the early 1990s, capital resources were shifting from the Tigers to China and India as a result of the cheaper labour resources available. The growth of China and India challenged the export



supremacy of the other Asian nations such as Malaysia, Taiwan, South Korea, Singapore, and Thailand, which had led the Asian growth phase in the late 1980s. The shifting investment and rising export strength of China in particular, reduced the growth rates in the Tiger nations. Several other shocks occurred in this period, which undermined the miracle.

First, the Chinese renminbi and the Japanese yen were devalued. Second, the US Federal Reserve increased interest rates, which pushed up the value of the US dollar and placed strain on the currencies which were pegged to it. Third, the fall in export earnings was exacerbated by the large fall in semiconductor prices (falling 36 per cent between 1993 and 1999).

The growth phase was also accompanied by a boom in real estate prices, which was fuelled by significant short-term foreign currency loans, increasing the risk exposure of the private sector reliant on export incomes to service the debts.

The crisis proper began in Thailand in July 1997. The Thai baht was pegged to the US dollar, a practice that was common among the Asian economies. Its real estate sector had pushed the nation's foreign debt beyond sustainable limits and speculative capital outflows, motivated by the fear of losses if the currency fell in value, put pressure on the exchange rate.

In the face of these pressures, the central bank was unable to maintain the peg as it ran short of the required foreign currency reserves. Once the government floated the baht on July 2, 1997, its value fell by more than 50 per cent as international investors dumped it on the foreign exchange rate and created a massive excess supply.

The collapse of the currency effectively rendered the nation bankrupt, given the large volumes of foreign-currency denominated debt held by the private sector. There was a significant fall in the local stock market and several major financial institutions were bankrupted.

These events exposed the dangers of maintaining currency pegs, which required central banks to have sufficient foreign currency reserves to maintain the agreed parities. This made all currencies in the region susceptible to speculative attacks.

While the structure of the Thai economy was very different to the Tigers in the East, speculators considered that all currencies were in danger. This belief became a self-fulfilling prophecy and by August 1997, speculative attacks on the currencies of Indonesia, Malaysia and the Philippines led to declines in their exchange rates.

The crisis spread in September to Hong Kong, Singapore, and Taiwan and in November 1997, the capital outflow from South Korea forced it to devalue.

The banks that had extended short-term loans to these nations, refused to roll-over the debt and an instant credit crunch was created.

It is clear that the stronger advanced nations such as the US, Japan and the EU could have intervened and facilitated enough liquidity to stop the capital outflow, which had resulted from the panic. Not only should their central banks have provided credit lines to the central banks in the Asian nations, but the advanced governments should have brokered roll-over arrangements with the private banks to stop the panic.

Instead, the main response from the advanced nations came through the IMF, which intervened first in Thailand in July 1997.

The Asian financial crisis exposed the deficiencies in the International Monetary Fund (IMF). In the *Advanced Material Box – The IMF Annual Report 1997* we see that the IMF considered the Asian economies both advanced (South Korea) and less developed (for example, Indonesia and Thailand) to be growing strongly on the back of extensive deregulation of their financial systems.

While there was some recognition that capital inflow was very strong and perhaps volatile, the IMF failed to correctly assess the vulnerability that their policy prescriptions (liberalisation etc) had created. The world found out just two weeks after this report was published.

By the end of 1997, the IMF was harshly criticising the Asian governments that earlier in the year they had been praising.

It is now accepted that a series of policy blunders in the IMF response deepened and spread the crisis.

In return for bailout funds, the IMF insisted that the nations under speculative attack in the currency markets introduce sharp increases in interest rates and substantial fiscal contraction. The IMF applied their 'one-size-fits-all' approach that was their standard response when budget deficits were significant and inflation accelerating.

This approach is questionable at the best of times, but certainly was inapplicable to the Asian economies, which were running budget surpluses and had stable inflation rates.

In its most simple form, the crisis was the result of excessive financial liberalisation that promoted massive capital inflow (and commensurate liabilities) between 1993 and 1996. The liabilities tended to be short-term but the funds were used for long-term investments (for example, real estate speculation).

When export growth slowed, the capital inflows started reversing very quickly. But the IMF demands ensured that the crisis moved out of the foreign exchange markets and became a full-blown economic recession.

As capital outflows accelerated with the worsening economic conditions, the IMF insisted that interest rates be pushed up further and fiscal contraction deepened.

As part of the Indonesian bail-out plan, the IMF forced the government to close 16 insolvent banks claiming that this would restore confidence in the remaining banks. The result was the opposite and the panicked withdrawals of funds undermined the solvency of many of the private banks.

The Indonesian central bank injected funds into these banks (equivalent to 5 per cent of GDP) to save them, which had the effect of exacerbating the collapsing rupiah and was at odds with the IMF's insistence that interest rates had to rise sharply.

In summary, the Asian financial crisis was the result of a lack of regulation on capital flows combined with the currency pegs.

In the case of the latter, these were interpreted by financial markets as the government insuring them against foreign exchange risk and so there was a lack of private foreign exchange hedging of the borrowing.

Once the currencies collapsed and floated, these unhedged positions quickly led to bankruptcy.

## 16.7 Capital Controls

The history of financial crises indicates that large-scale financial speculation can undermine a nation's real economy relatively quickly if the government attempts to peg their currency to another or the economy has significant foreign-currency denominated debt exposure (private or public).

While the international community could agree that certain forms of speculative activity would be considered illegal, in lieu of that, the nation under attack has to defend its own prosperity.

One such suggestion is to introduce capital controls, which limit the size and flexibility of international financial flows.

In September 1998, during the Asian debt crisis, the Malaysian government introduced capital controls after the currency had appreciated significantly and the central bank had pushed interest rates up towards 18 per cent and undermined the viability of many local businesses.

Capital controls are policies that restrict the free movement of capital, either in terms of inflows or outflows.

There are broadly two types of capital controls used:

- Administrative or direct controls, which impose limits or bans on capital flows.
- Market-based controls, which impose extra costs on capital flows which reduce the incentives to shift funds across national borders.

A government might, for example, place limits on foreign exchange transactions, international bank transactions, or bank withdrawals. Restrictions on movements of precious metals such as gold might also be considered.

The aim is to limit the scope of speculative flows (in or out) to manipulate the exchange rate and strain the central bank's foreign exchange reserves.

Capital controls allow the central bank to run an autonomous monetary policy and the treasury to use fiscal policy to manage domestic demand in the interests of the nation.

### Appendix: Advanced Material

#### An increase in world income leads to a rise in net exports

Prior to the rise in world income, the level of real GDP,  $Y_0$  is consistent with a balance of trade,  $NX_0 = 0$ .

Assuming no change in the exchange rate, we can write this as:

$$(16.11) \quad X_0 = M_0 + mY_0$$

where the  $M_0$  is a constant quantity of imports at the current constant exchange rate.

If we denote  $\Delta X = \lambda \Delta Y_w$ , which is the increase in exports resulting from the rise in world income, then national income  $Y_1$  at which net exports are zero, satisfies:

$$(16.12) \quad X_0 + \Delta X = M_0 + mY_1$$

If we subtract Equation (16.11) from (16.12) we get:

$$(16.13) \quad \Delta X = mY_1 - mY_0 = m\Delta Y'$$

So the change in the level of real GDP, such that net exports are again zero is given by:

$$(16.14) \quad \Delta Y' = \Delta X/m$$

On the other hand, the increase in equilibrium real GDP (national income) resulting from an autonomous increase in exports (driven by the rise in world income) is given by (from Equation 16.10):

$$(16.15) \quad \Delta Y = 1/[1 - c(1 - t) + m] * \Delta X$$

This magnitude is less than  $\Delta X/m$  because  $[1 - c(1 - t) + m] > m$ .

Thus at the new national equilibrium,  $Y^*$ , net exports are positive – that is a trade surplus.

### The IMF Annual Report 1997

The IMF Annual Report 1997 was published on April 30, 1997, just a few weeks before the collapse in the Thai baht triggered the region's crisis.

The Report noted (page 26):

Directors remarked upon the speed with which net capital flows to emerging markets had recovered from the disturbances associated with the Mexican financial crisis in early 1995. There was general agreement that investors had become more aware of the risks associated with investing in emerging markets. Notably, there had been a greater reliance on a wider range of macroeconomic, financial, and banking soundness indicators in assessing both economic conditions and investment opportunities in developing markets.

Chapter 5 of the Report summarised the assessments the IMF made from its annual consultations with the member countries:

#### South Korea:

- “Since 1994, the current account deficit had widened sharply ... [but] ... capital inflows, growing in response to continued relaxation of controls on foreign equity investment and on access to trade financing, had comfortably financed this deficit” (page 59).
- Its expanding privatisation program was praised (page 59).
- The IMF Directors “welcomed Korea’s continued impressive macroeconomic performance ... noted the challenge of ensuring that the speed of structural reforms in the financial sector and the capital account ... praised the authorities for their enviable fiscal record and suggested ... maintaining a strong budgetary position ... welcomed the recent acceleration of capital account liberalization ... considered that rapid and complete liberalization offered many benefits at Korea’s stage of economic development.” (pages 59-60)

#### Indonesia:

- The IMF considered “further substantial reforms, including financial sector reforms and the development of a strong capital market, were essential for maintaining rapid, sustained growth” (page 80).

- “Further deregulation and opening of the economy to world markets were the key to maintaining competitiveness” (page 81).

Thailand:

- The IMF “Directors strongly praised Thailand’s remarkable economic performance and the authorities’ consistent record of sound macroeconomic policies.” (page 91).
- “The stability of the baht had served the Thai economy well in the past, but Directors recommended a greater degree of exchange rate flexibility to improve monetary autonomy ...” (page 91).
- “Directors noted, economic fundamentals remained generally very strong, characterized by high saving and investment, a public sector surplus, strong export growth in recent years, and manageable debt and debt-service returns.” (pages 91-92).

## References

Boughton, J. M. (2001) “Was Suez in 1956 the First Financial Crisis of the Twenty-First Century?”, *Finance and Development*, September 2001, Vol. 38 No.3, available at: <http://www.imf.org/external/pubs/ft/fandd/2001/09/boughton.htm>.

International Monetary Fund (1994) Minutes of the Executive Board of the IMF (EBM/94/104 – November 30, 1994).

International Monetary Fund (1997) Annual Report of the Executive Board for the Financial Year Ended April 30, 1997, Washington, D.C., available at: <http://www.imf.org/external/pubs/ft/ar/97/97imf.htm>.

International Monetary Fund (2011) Balance of Payments and International Investment Position Manual (BPM6), Washington: Available from <http://www.imf.org/external/pubs/ft/bop/2007/pdf/bpm6.pdf> (BPM6).

United Nations Statistical Commission (2009) System of National Accounts 2008, New York.

# Appendix: Methods, Tools and Techniques

## Chapter Outline

- A.1 Overview
- A.2 Basic Rules of Algebra
  - Rule: addition and subtraction
  - Rule: multiplication and division
  - Model solutions
- A.3 A Simple Macroeconomic Model
- A.4 Graphical Depiction of a Macroeconomic Model
- A.5 Power Series Algebra and the Expenditure Multiplier
- A.6 Index Numbers
- A.7 Annual Average Growth Rates
- A.8 Textbook Policy Regarding Formalism

## A.1 Overview

In macroeconomics we often deploy symbols to represent real world variables of interest, such as real GDP, consumption, and investment. In this case, while the symbols can have an abstract meaning (for example,  $Y$  is real GDP) they will also usually have, a quantitative analogue (for example, in the September-quarter 2013, real GDP in Australia was estimated to be \$A387,031 million).

In Chapter 1, the concept of a model was introduced. A macroeconomic model expresses our theoretical conjectures about the relationships between the main macroeconomic variables such as employment, output and inflation.

A model is a generalisation about the way the system functions or behaves. It could easily be a narrative statement such as – a household will consume a proportion of their income after tax (disposable income). That theoretical statement might then be examined for its empirical relevance but will also stimulate further theoretical work trying to provide an explanation for that conjectured behaviour.

In economics, like other disciplines that use models, the narrative statement might be simplified with some mathematical statement involving symbols. In this context, the models will be represented by a number of equations (which could vary from one equation to hundreds or even thousands of equations), which describe relationships between the variables of interest. Thus mathematics is a form of shorthand in terms of concisely representing relationships between variables. We can then apply the basic rules of algebra to conduct our analysis.

As we mentioned above, we typically use letters (such as  $Y$ ) to denote different macroeconomic variables. A variable can take on different values in different time periods. The correspondence between the shorthand symbol and the variable is not always intuitive but we maintain the same conventions throughout this textbook.

Greek symbols (such as  $\alpha$ ) are often used to denote parameters of the model, which contribute to the formalisation of the relationships between variables. In the first instance these parameters are usually assumed to be constant over time.

So  $Y$  is often used to denote real GDP or National Income (but it can also be used to denote total output).  $C$  is usually used to denote final household consumption and  $I$  total private investment.  $X$  is typically used to denote exports and  $M$  imports although in some cases  $M$  is used to denote the stock of Money. In this text,  $M$  is exclusively used to denote imports.

There are several types of variables used in macroeconomic models:

- Endogenous or dependent variable – its value is determined by the solution to the model and is thus dependent on the values of other variables.
- Exogenous or autonomous variable – its value is given in advance of solving the model.
- Independent variable – its value determines (wholly or partly) the value of some related dependent variable.

As noted, variables are related by way of equations, which express the structure of the macroeconomic model. Usually a variable that we seek to explain is written on the left-hand side of the equals sign ( $=$ ) and is then expressed in terms of some other variables on the right-hand

side of the equals sign, which we consider are influential in explaining the value and movement of the left-hand side variable of interest.

The relationship between the variables on the right- and left-hand side of the equation is described in terms of some coefficients (or parameters).

For example,  $y = 2x$  is an equation which says that variable  $y$  is equal to 2 times variable  $x$ . The equals sign (=) tells us that the left-side of the equals sign is of the same magnitude as the right-side (that is, an equation has equal left and right sides).

You solve an equation by substituting values for the unknowns. The number 2 in the equation is called a coefficient which is an estimate of the way in which  $y$  is related to  $x$ .

So if  $x = 1$ , then we can solve for the value of  $y = 2$  as a result of this equation.

A coefficient can also be called a parameter – which is a given in a model and might be estimated using econometric analysis (regression) or assumed by intuition, so that the coefficient's value is strictly unknown.

For example, we might have written the above equation as  $y = bx$ , where  $b$  is the unknown coefficient. You will note that we would be unable to 'solve' for the value of  $y$  in this instance even if we knew the value of  $x$ . In the case above where we said  $x = 1$ , then all we could say that  $y = b$ . We would thus need to know the numerical value of  $b$  before we could fully solve for  $y$ .

There are several types of equations that are used in macroeconomic models:

- Identity equation – is an expression that is true by definition (usually relating to an accounting statement). For example, we will see that GDP is equal to the sum of the expenditure components, which is true as a result of the way we set up the national accounts and define the expenditure components. In mathematics, the symbol  $\equiv$  is used to denote an identity relationship.
- Behavioural equation – captures the hypotheses we form about how a particular variable is determined. These equations thus represent our conjectures (or theory) about how the economy works and obviously different theories will have different behavioural equations in their system of equations (that is, the economic model).
- Equilibrium equation – is an expression that captures a relationship between variables that defines a state of rest.

While the above example ( $y = 2x$ ) was easy to solve once we knew the value of  $x$ , sometimes it is useful to have models where we cannot solve for numerical values of the unknown variables of interest but we can simplify the equations to show the structure of the model in terms of what is important to advance our understanding of the relationship between our aggregates.

## A.2 Basic Rules of Algebra

You will need to learn some basic algebraic rules that are used to manipulate equations and solve for unknowns of interest. Often you will need to rearrange a given equation in order to determine the solution for the variable of interest.



### Rule: addition and subtraction

In an equation  $y = x$ , then we know the equivalent expression is  $y \pm z = x \pm z$ . So, for example,  $y = x$  is equivalent to  $y + 2 = x + 2$ .

In general, what we add to or subtract from one side of the equation, we have to add to or subtract from the other side to maintain the equality.

We can also substitute an expression from one equation into another and maintain equivalence.

For example, we might have  $y=2x$ , and  $x=6z$ . In this case, it is equivalent to write  $y=2(6z)=12z$ .

### Rule: multiplication and division

In an equation  $y = x$ , then we know the equivalent expression is  $3y = 3x$  or  $y/3 = x/3$ . If we multiply or divide the left-hand side of the equation by a variable (or more complex algebraic expression) then we have to multiply or divide the right-hand side of the equation by the same variable (or expression). Dividing by zero is not allowed, however, and multiplying by zero is not very helpful!

### Model solutions

In a system of equations, the values of some variables are unknown and are only revealed when we 'solve' the model for unknowns.

So if we have these two equations, which comprise a 'system':

$$(A.1) \quad y = 2x$$

$$(A.2) \quad x = 4$$

Then  $x$  is a pre-determined variable (with the value of 4) and is thus exogenous. You do not know the value of  $y$  in advance and you have to solve the equations to reveal its value – so it is endogenous. To solve this system we substitute the value of  $x$  in Equation (A.2) into Equation (A.1) so we get:

$$y = 2 \text{ times } 4$$

$$y = 8$$

More generally the solution of a system of structural equations entails expressing each of the endogenous variables, say  $y_1, y_2, \dots, y_n$  as functions of the exogenous variables  $x_1, x_2, \dots, x_m$ , so that there are  $n+m$  'solution' equations, (including the values of the  $x$  variables, like Equation (A.2) above).

Then the system of equations can be solved to generate  $n$  reduced form equations which express each  $y_i$  ( $i=1,2, \dots, n$ ) in terms of the  $x$  variables ( $x_1, x_2, \dots, x_m$ ). The known values of the exogenous variables ( $x$ s) can be substituted into each of these reduced form equations which yield solutions for the endogenous variables, say  $y_1^*, y_2^*, \dots, y_n^*$ . There are constraints on the initial system of (structural) equations, for there to be a unique solution for the unknown exogenous  $x$ s and endogenous  $y$ s, which include that the number of equations equals the number of unknowns – in this case  $n+m$ .

In modelling an economic system it becomes very complicated as to which variables can be considered endogenous and which are truly exogenous. At the extreme, everything might be

considered endogenous and then things get mathematically complex. There is a body of econometric theory which explores the problem of identifying values of coefficients in empirical work, which is well beyond this textbook.

### A.3 A Simple Macroeconomic Model

An example of an identity (which is true by definition) is the famous national income identity depicting aggregate demand and output, which we consider in Chapter 4:

$$(A.3) \quad Y \equiv C + I + G + X - M$$

We have used the identity sign (three parallel horizontal lines) instead of the equals sign (two parallel horizontal lines). That distinguishes Equation (A.3) from a behavioural equation which is always expressed using an equals sign (=).

This identity is also an equilibrium condition in the simple national income model but it provides no information about how the right-hand side variables behave, that is what factors influence them. To advance that understanding we form theories about the determinants of these variables.

These theories are expressed in behavioural equations. An example of a behavioural equation is the simple consumption function:

$$(A.4) \quad C = C_0 + cY_d$$

which says that final household consumption ( $C$ ) is equal to a constant ( $C_0$ ) plus a proportion ( $c$ ) of final disposable income ( $Y_d$ ). The constant component ( $C_0$ ) is the consumption that occurs if there is no income and can be considered to be dis-saving.

Note that subscripts are often used to add information to a variable. So we append a subscript  $d$  to our income symbol  $Y$  to qualify it and denote disposable income (total income after taxes).

We also use subscripts to denote time periods when we are considering a variable over time. So  $Y_t$  indicates we are considering the value of  $Y$  at time period  $t$ . Similarly,  $Y_{t-1}$  refers to the value of  $Y$  at time period  $t-1$ , where the lag (the  $-1$ ) depends of the periodicity of the data. If we were using quarterly data, then  $t-1$  would be last quarter and so on.

In macroeconomics, some behavioural coefficients are considered important and are given special attention. So the coefficient  $c$  in the consumption function is called the marginal propensity to consume (MPC) and denotes the extra consumption per dollar of extra disposable income. So if  $c = 0.8$  we know that for every extra dollar of disposable income that the economy produces, 80 cents will be consumed.

The MPC is intrinsically related to the marginal propensity to save (MPS) which is the amount of every extra dollar generated that is saved (after households decide on their consumption). So the  $MPS = 1 - MPC$  by definition.

The importance of MPC is that is one of the key determinants of the expenditure multiplier (more about which later). We will consider this in Chapter 7 when we discuss the expenditure multiplier.

We have already introduced the distinction between an exogenous (pre-determined or given) variable and an endogenous variables (which are determined by the solution to the system of equations).

An exogenous variable is known in advance of ‘solving’ the system of equation. We take its value as given or pre-determined. We might say, by way of simplification, that government spending ( $G$ ) in the above model is equal to \$100 billion which means that its value is known and not determined within the model.

The two equations (the identity and behavioural equation) noted above form a macroeconomic system. This is, of course is a very simple system. For the sake of exposition, we might assume the economy is closed which means there are no exports or imports. In that case, the national income identity becomes  $Y = C + I + G$ .

We also assume there is no taxation in the model, so that disposable income is equal to total income ( $Y$ ).

The model is now:

$$(A.5) \quad Y = C + I + G$$

$$(A.6) \quad C = C_0 + cY$$

For simplicity, we will assume that  $I$  and  $G$  are exogenous and their values are known in advance. The remaining two variables  $Y$  and  $C$  are endogenous and their values are dependent on the solution to the model.

How do we solve for  $Y$ ?

We substitute (A.6) for  $C$  in (A.5) such that:

$$(A.7) \quad Y = C_0 + cY + I + G$$

We can now re-arrange this (noting we have  $Y$  on both sides) by subtracting  $cY$  from both sides as per our algebraic rules. This gives:

$$(A.8) \quad Y - cY = C_0 + I + G$$

You will note that there are only predetermined variables (knowns) on the right-hand side now.

We can re-arrange this further:

$$(A.9) \quad Y(I - c) = C_0 + I + G$$

This uses the rule relating to factorisation where  $Y$  is a common factor on the left-hand side. Note that  $Y(I - c) = Y - cY$ .

We can then invoke the rule whereby we divide both sides by  $(I - c)$  to isolate (or solve) for  $Y$  on the left-hand side. Thus:

$$(A.10) \quad Y = [C_0 + I + G]/(I - c)$$

So in words, the equilibrium value of  $Y$  is a function of the autonomous variables in the model  $C_0 + I + G$ .

We call Equation (A.10) the reduced-form solution of the model, where there are only exogenous or predetermined variables on the right side and an unknown variable on the left-hand side of the equation.

In a macroeconomic model, all the endogenous variables can be expressed in reduced-form. So in the example above, our solution for consumption would be:

$$(A.11) \quad C = C_0 + cY = C_0 + c[C_0 + I + G]/(1 - c) = [C_0 + c(I + G)]/(1 - c)$$

Make sure you can derive the steps that are needed to get this solution.

The reduced-form of the system allows us to conduct sensitivity analysis, which involves changing the values of the exogenous variables or the coefficients (in this case, the MPC) and analysing the impact on the endogenous variables,  $C$  and  $Y$ , in the model.

As an example, what would be the impact of an expansion in government spending  $G$  on national income  $Y$ ? Note that we assume the other exogenous variables are unchanged.

From Equation (A.10), we know that when  $G = G_0$ :

$$(A.10a) \quad Y_0 = [C_0 + I]/(1 - c) + [G_0]/(1 - c).$$

Note, to make the manipulation clear, we have separated  $G$  from  $I$  and  $C_0$ .

If  $G$  rises from  $G_0$  to  $G_1$ , then:

$$(A.10b) \quad Y_1 = [C_0 + I]/(1 - c) + [G_1]/(1 - c)$$

So the change in  $Y = (Y_1 - Y_0)$  is the difference between (A.10b) and (A.10a):

$$(A.10c) \quad (Y_1 - Y_0) = [G_1 - G_0]/(1 - c).$$

To simplify our notation, we will usually denote the change in a variable using the Greek symbol  $\Delta$ . So Equation (A.10c) would be written as:

$$(A.10d) \quad \Delta Y = \Delta G/(1 - c)$$

where the time span that  $\Delta$  denotes is indicated by the context.

We can then express Equation (A.10d) as:

$$(A.10e) \quad \Delta Y/\Delta G = 1/(1 - c)$$

by dividing both sides of the equation by  $\Delta G$ . The right hand side of Equation (A.10e) is known as a **multiplier** because it tells us the magnitude of the change in  $Y$  for a unit change in  $G$ . We will examine multipliers in more detail in Chapter 7, where we include taxation and exports and imports, so the economy is open.

## A.4 Graphical Depiction of a Macroeconomic Model

We will also express our theories in graphical terms, which are an alternative to mathematical representation. Here are three ways to express the same theoretical idea.

1. Household consumption rises proportionately with disposable income but the proportion is less than one.

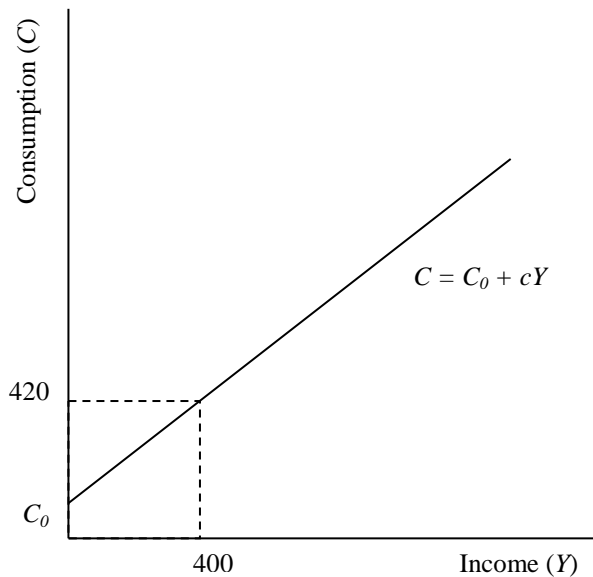
2.  $C = C_0 + cY_d$ , where  $0 < c < 1$  and  $C_0$  is a constant (fixed value). The less than signs ( $<$ ) tells us that the MPC lies between the value of 0 and 1, that is, it is positive but less than 1.

For now, we assume taxes are zero which means that national income ( $Y$ ) and national disposable income ( $Y_d$ ) are equal.

3. Graphical form:

Figure A.1 depicts a general consumption function.

**Figure A.1 Consumption function**



If  $C_0 = 100$ , and  $c = 0.8$ , and  $Y = 400$ , we can solve the equation  $C = C_0 + cY$  by inserting the known values of the parameters and explanatory variable (in this case, income) into the equation and solving it. Thus:

$$(A.11) \quad C = C_0 + cY = 100 + 0.8 * 400 = 420$$

You can also see that by tracing a vertical line from where disposable income equals 400 up to the graph line and then tracing across the vertical axis we derive the value of consumption by where that line crosses the vertical axis.

The slope of the line is the marginal propensity to consume ( $c$ ). How do we derive a slope of a line and what does it mean? In Chapter 7 we will deal with applications of the slope of a line when we study the principle of the expenditure multiplier.

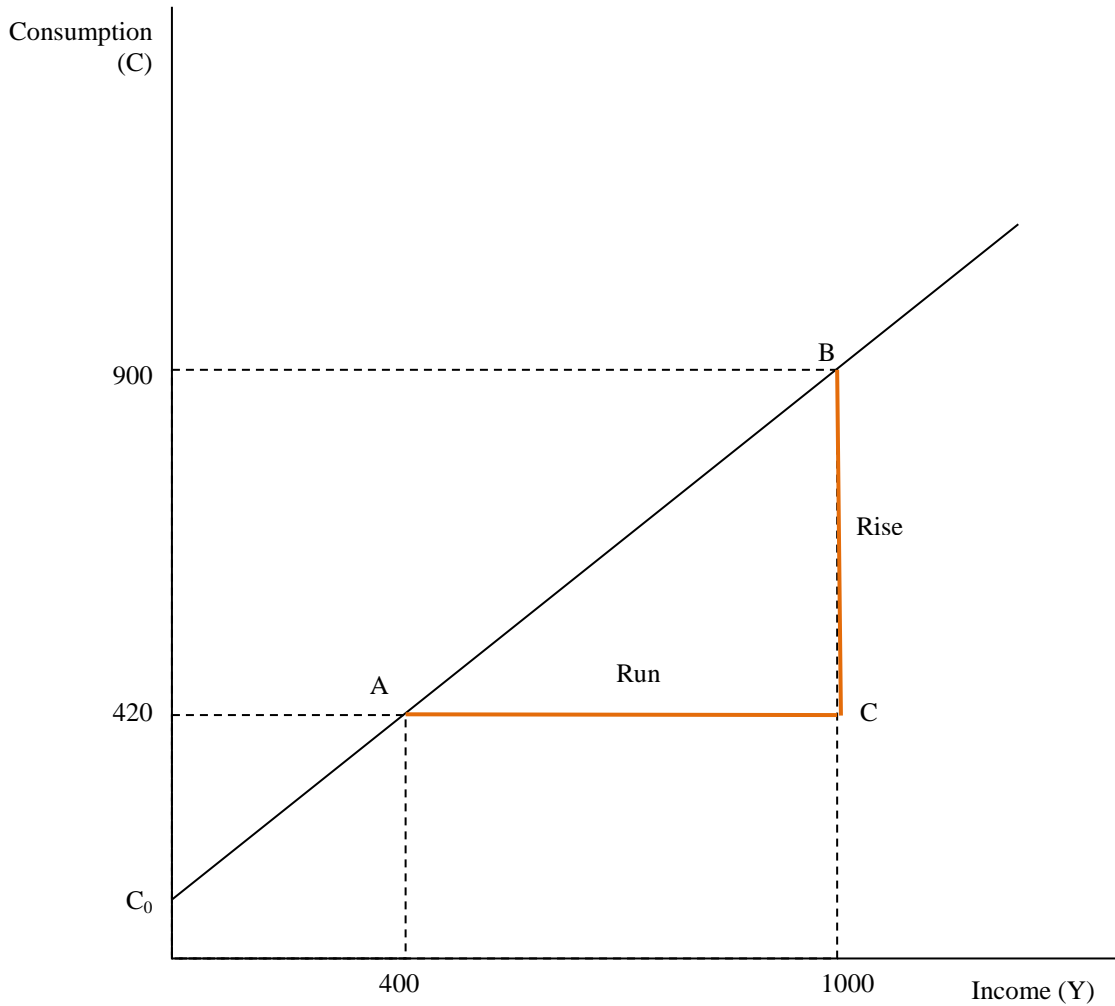
To advance our understanding, assume that national income rose to 1000. We know that consumption would rise to 900.

Check that you can solve the equation  $C = C_0 + cY = 100 + 0.8 \times 1000 = 900$

Consider Figure A.2 which shows the two combinations of  $Y$  and  $C$  that we have now determined. The first point denoted A, represents the combination  $(Y_1, C_1) = (400, 420)$ .

The second point, denoted B, represents the combination  $(Y_2, C_2) = (1000, 900)$ .

**Figure A.2 Slope of consumption function**



We represent the slope of the line that intersects A and B using the following formula:

$$(A.12) \quad c = \frac{C_2 - C_1}{Y_2 - Y_1}$$

You can see this is also the ratio of the segments  $BC/AC$  or in words ‘Rise over Run’ (the change in  $C$  divided by the change in  $Y$ ). This rule generalises to any linear function. Note that, depending on the context, Rise could be fall (in which case the slope is negative).

You can check the slope of our consumption function by noting that  $BC = 480$  and  $AC = 600$  so  $BC/AC = 0.8$ , which is the marginal propensity to consume ( $c$ ).

We noted earlier that we usually denote the change in a variable using the Greek symbol  $\Delta$ . In that context the slope of the consumption function would be  $c = \Delta C / \Delta Y$  and that slope would be constant at all points on the (linear) function.

The graphical approach to determining the slope of the line is confined to linear equations. If the function is non-linear (for example, a curve), the slope formula we have derived only provides us with the average slope between two points.

A more general approach is required when the relationship of interest is non-linear (for example, a quadratic or cubic function) and the slope is constantly varying.

In this situation, differential calculus is used and the slope of the function at some specific point is defined by the **derivative** of the function.

For the function,  $y = f(x)$ , the derivative is typically represented by the expression  $dy/dx$  or  $f'(x)$ .

Note that when deploying differential calculus, we are assuming (using the example above) that  $\Delta Y$  is infinitesimally small (that is, a value approaching zero). Typically, we will be concerned with changes in variables that are larger than this and so the tools of calculus will be of limited use to us in this textbook. Thus, we will use calculus sparingly in the textbook but it is useful to understand the basic concept of a derivative.

The slope of  $y = f(x)$  is  $\Delta y/\Delta x$ . We can see that  $\Delta y = f(x + \Delta x) - f(x)$ . Now what would happen if  $\Delta x$  was close to zero?

Take a specific example of a non-linear function,  $y = x^2$ . We say that  $x$  is raised to the power of two (or squared). You can see that in a linear function such as  $y = 2x$ ,  $x$  is raised to the power of one since  $x^1 = x$ .

$$\text{If } f(x) = x^2, f(x + \Delta x) = (x + \Delta x)^2 = x^2 + 2x \Delta x + (\Delta x)^2$$

Now if we substitute the above into the slope formula we get:

$$\begin{aligned} \frac{\Delta y}{\Delta x} &= \frac{f(x+\Delta x)-f(x)}{\Delta x} \\ \text{(A.13)} \quad &= \frac{x^2+2x\Delta x+(\Delta x)^2-x^2}{\Delta x} && \text{(cancels out } x^2 \text{ and } -x^2) \\ &= \frac{2x\Delta x+(\Delta x)^2}{\Delta x} && \text{(divide by } \Delta x) \\ &= 2x + \Delta x \end{aligned}$$

The derivative of a function  $y = f(x)$ , namely  $dy/dx$ , is the limiting value of  $\frac{\Delta y}{\Delta x}$  as  $\Delta x \rightarrow 0$ .

So when  $y = x^2$ , its derivative,  $dy/dx = 2x$ , since  $\Delta x \rightarrow 0$ .

What does this mean? It means that if  $y = x^2$ , the rate of change at any point is  $2x$ . If  $x = 4$ , then the rate of change in  $y$  is 8. If  $x = 5$ , the rate of change in  $y$  is 10.

More generally if  $y = x^n$ , it can be readily shown that  $dy/dx = nx^{n-1}$ . Thus if  $y = x^3$ ,  $dy/dx = 3x^2$ , which can be evaluated at any value of  $x$ .

For the consumption function  $C = C_0 + cY$ , the general formula for a derivative would yield  $dC/dY = cY^{1-1} = cY^0 = c$ , which is consistent with what we have already found. Note that  $Y^0 = 1$ .

## A.5 Power Series Algebra and The Expenditure Multiplier

We have introduced the concept of the expenditure multiplier, which we consider further in Chapter 7. It allows us to calculate the total change in national income following a change in one of the autonomous components of aggregate demand, such as government spending or private investment.

Multipliers play a central role in discussions of the economic impact of policy interventions. The multiplier depends on the magnitude of the initial injection of expenditure plus the induced expenditure that follows. We will learn that when, for example, the government increases its spending, national income rises by that amount, which, in turn, induces further consumption expenditure.

Thus if \$1 was injected into the economy, through additional spending, total income would initially rise by \$1. If the marginal propensity to consume was 0.8, then this initial rise in income would induce a rise in consumption of  $0.8 \times \$1$  or 80 cents in period 1. This initial \$0.80 rise in induced spending would further induce a rise in income of \$0.80 which would induce additional consumption in period 2 of  $0.8 \times 0.8$  or 64 cents and so on. (Note that this is a simple, mathematical exposition of what in the real world would be a complex process of adjustment of the economy to an increase of government spending).

The multiplier is thus related to the slope of the function and the algebraic notion of a geometric or power series.

Consider a \$1 increase in government spending with  $c$  representing the marginal propensity to consume and  $n$  being the number of periods. Then we could write the multiplier over  $n$  periods as  $k(n)$ , where:

$$(A.14) \quad k(n) = 1 + c + c^2 + c^3 + c^4 + \dots + c^n$$

The right-hand expression is called a power series.

Note that:

$$(A.15) \quad ck(n) = c + c^2 + c^3 + c^4 + c^5 + \dots + c^n + c^{n+1}$$

If we subtract (A.15) from (A.14) we get:

$$(A.16) \quad k(n) - ck(n) = k(n)(1 - c) = 1 - c^{n+1}$$

$$(A.17) \quad k(n) = (1 - c^{n+1})/(1 - c)$$

Then if we consider  $n \rightarrow \infty$  and denote the summation by  $k^*$

$$(A.18) \quad k^* = 1/(1 - c)$$

We interpret (A.18) as showing the size of the multiplier when  $n$  is a large number, so we are considering the overall impact of the multiplier process, rather than considering it over a finite number of periods,  $n$ . The value of  $c^{n+1}$  tends to zero as  $n$  gets large since  $0 < c < 1$ . That allows us to derive (A.18) from (A.17) and show that the multiplier,  $k^*$ , is  $1/(1 - c)$ , where  $c$  is the marginal propensity to consume.

## A.6 Index Numbers

An index number allows the comparison between the values of a variable over time. For example in Chapter 4, the construction of a Consumer Price Index was described. However if two or more variables are expressed in index number form, then straightforward comparisons of their respective rates of change (growth rates) can be made.

The creation of an index number requires a starting point (the base period or value) which is usually set at 100. Each observation is then expressed as a percentage of the base year.



For example, consider the data in Table A.1, which shows full-time, part-time and total employment for Australia from 2000 to 2012. The data is reported in units of thousands.

Visual inspection of the data can provide information as to the evolution of employment over this time period in Australia.

But what if we wanted to know whether full-time employment had grown faster or slower than part-time employment since 2000? This is where index numbers are useful.

**Table A.1 Employment in Australia, 2000-12, thousands**

	<b>Full-time Employment 000s</b>	<b>Part-Time Employment 000s</b>	<b>Total Employment 000s</b>
2000	6,614.6	2,375.1	8,989.7
2001	6,597.5	2,492.4	9,089.9
2002	6,648.7	2,622.0	9,270.7
2003	6,772.7	2,712.2	9,485.0
2004	6,930.4	2,730.9	9,661.3
2005	7,148.7	2,849.4	9,998.1
2006	7,326.9	2,930.2	10,257.1
2007	7,583.7	2,993.3	10,577.0
2008	7,782.5	3,093.1	10,875.6
2009	7,724.3	3,229.5	10,953.7
2010	7,852.5	3,337.6	11,190.0
2011	8,014.5	3,376.1	11,390.6
2012	8,098.3	3,417.1	11,515.4

Source: Australian Bureau of Statistics, Labour Force.

To convert this data into index numbers we define the base as the year 2000 and call that value 100. So for full-time employment in 2000, we would calculate  $100 \times (6,614.6 / 6,614.6) = 100$  and so on for the other variables of interest. In other words, we divide each value of full-time employment for 2000 through to 2012 by its value in 2000 (6,614.6) and then multiply by 100.

The index number for full-time employment in 2001 would thus be  $100 \times (6,597.5 / 6,614.6) = 99.7$  and so on.

Table A.2 shows the index numbers corresponding to the three employment time series. Note that while the raw employment data is expressed in units of thousands, the index numbers are unit free. Thus defining a common base year for the three series (i.e. values in 2000 are set at 100) means that comparisons between them can be readily made, even though their corresponding absolute values differ quite significantly.

We can see that the index number for full-time employment rose from 100 in 2000 to 122.4 in 2012, a 22.4 per cent increase, while the part-time employment index number rose from 100 in 2000 to 143.9 in 2012, a 43.9 per cent rise, almost twice as large an increase as full-time employment.

**Table A2      Employment indices for Australia, 2000-12**

	Full-time Employment 000s	Part-Time Employment 000s	Total Employment 000s	Full-Time Employment 2000=100	Part-Time Employment 2000=100	Total Employment 2000=100
2000	6,614.6	2,375.1	8,989.7	100.0	100.0	100.0
2001	6,597.5	2,492.4	9,089.9	99.7	104.9	101.1
2002	6,648.7	2,622.0	9,270.7	100.5	110.4	103.1
2003	6,772.7	2,712.2	9,485.0	102.4	114.2	105.5
2004	6,930.4	2,730.9	9,661.3	104.8	115.0	107.5
2005	7,148.7	2,849.4	9,998.1	108.1	120.0	111.2
2006	7,326.9	2,930.2	10,257.1	110.8	123.4	114.1
2007	7,583.7	2,993.3	10,577.0	114.7	126.0	117.7
2008	7,782.5	3,093.1	10,875.6	117.7	130.2	121.0
2009	7,724.3	3,229.5	10,953.7	116.8	136.0	121.8
2010	7,852.5	3,337.6	11,190.0	118.7	140.5	124.5
2011	8,014.5	3,376.1	11,390.6	121.2	142.1	126.7
2012	8,098.3	3,417.1	11,515.4	122.4	143.9	128.1

Note that for any pair of index number observations we can also compute percentage changes. For example, what was the growth of full-time employment in 2008 and 2009?

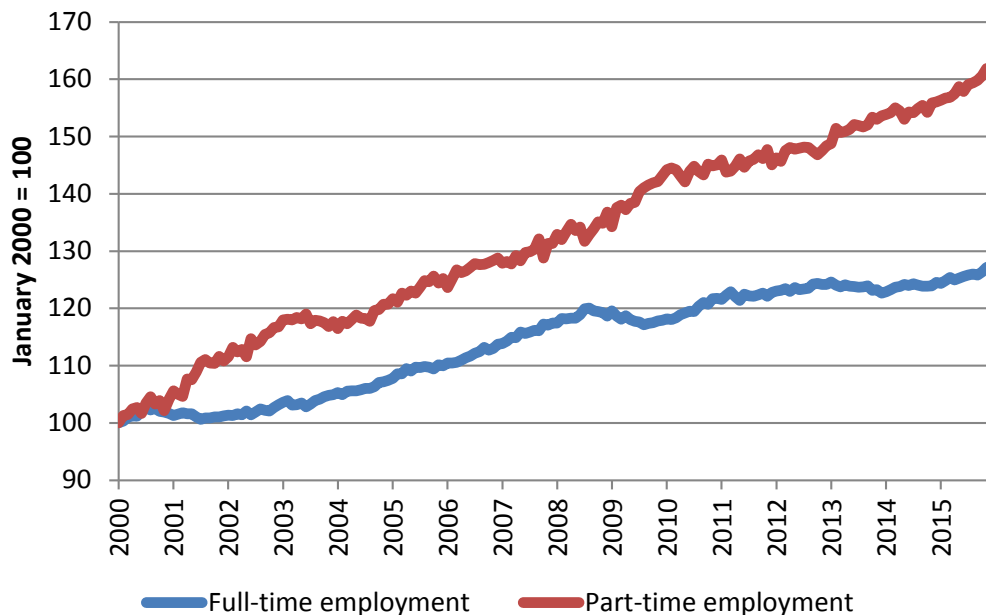
We know that full-time employment was 7,782.5 in 2008 and had fallen to 7,724.3 in 2009. We could calculate a simple percentage  $100 \times (7,724.3 - 7,782.5) / 7,782.5 = -0.75$  per cent.

We could find the same result using the index numbers 117.7 in 2008 and 116.8 in 2009 and computing  $100 \times (116.8 - 117.7) / 117.7 = -0.76$  per cent. (The small difference here is accounted for by rounding up the index numbers to one figure after the decimal point. If 4 significant figures are used, the index numbers for 2008 and 2009 would be 1.1766 and 1.1678, respectively yielding a fall of 0.747 per cent which rounds up to 0.75 per cent.)

It is also useful to graph index numbers if we are interested in a visual comparison of the behaviour of different variables.

For example, the visualisation easily allows one to see that in 2008, when the financial crisis emerged, full-time employment in Australia contracted while part-time employment accelerated. This observation would motivate a researcher to investigate the labour market processes that underpinned this outcome.

**Figure A.3 Employment index numbers, Australia, 2000-2012, 2000=100**



As mentioned above, index numbers also allow us to compare the evolution of two or more different variables over time when the underlying units of measurement of each variable are different.

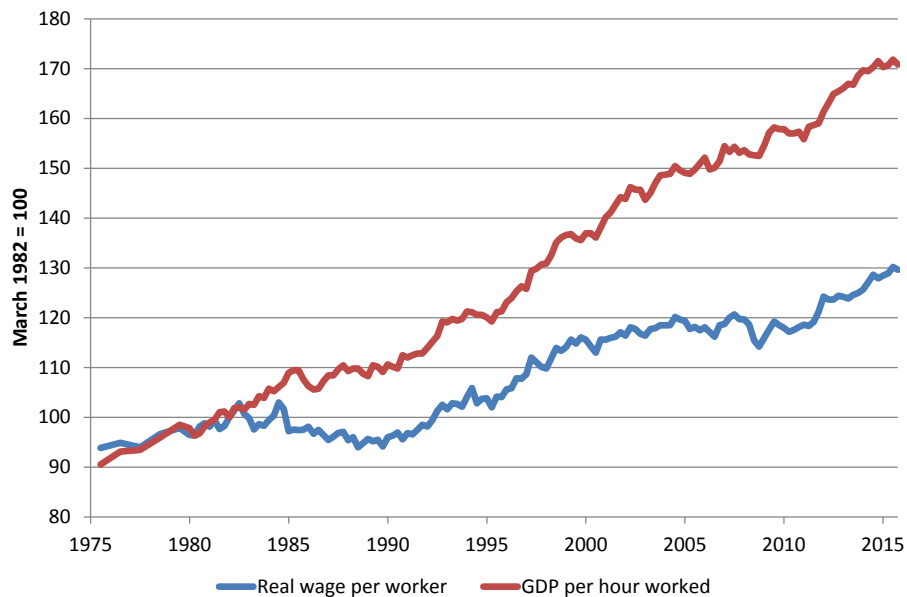
For example, for decades real wages in most nations grew in line with labour productivity. The latter created the space for the former to grow without invoking inflationary pressures. So economists are often interested in examining the relationship between the two variables over time.

Figure A4 shows the relationship between real wages and productivity growth in Australia from 1978 to 2012 in index number form where the base period is March 1978. The underlying data is quarterly.

Productivity is measured in terms of units of real GDP per hour worked while real wages are computed as the nominal wage series (\$ per hour) deflated by the consumer price index.

By converting the different series in common index numbers we are readily able to see the comparative behaviour of these related time series variables.

**Figure A.4 Real wages and productivity, 1978- 2012, March-quarter 1978=100**



Source: Australian Bureau of Statistics, National Account.

## A.7 Annual Average Growth Rates

In Chapter 7 we compute the percentage change in the CPI from one period to the next. We were essentially solving the equation:

$$\text{CPI}_t = \text{CPI}_{t-1}(1 + r^*/100)$$

where  $r^*$  is the rate of change of the CPI, expressed as a percentage.

Economists often are required to compute how fast the economy (or some other aggregate) is growing on average over a number of years. We are calculating an average compound growth rate.

In 1960, real GDP in Australia was \$249,083 million and it had grown to \$1,508,267 million by 2012. To calculate the average annual compound growth rate over this 52-year period, we need to utilise some simple algebra and deploy the notion of a compound growth rate.

We can write:

$$(A.19) \quad Y_t = Y_0 (1 + r)^t$$

where  $Y_t$  is real GDP in 2012;  $Y_0$  is real GDP in 1960;  $r$  is the average compound growth rate, expressed as a decimal, i.e.  $r = r^*/100$ ; and  $t$  is the number of periods that we are compounding over, in this case 52 years.

The task is to solve for the unknown  $r$ :

$$(A.20) \quad \begin{aligned} (Y_t/Y_0) &= (1 + r)^t \\ \ln(Y_t/Y_0) &= t \ln(1 + r) \\ \frac{\ln(Y_t/Y_0)}{t} &= \ln(1 + r) \end{aligned}$$

$$\exp\left[\frac{\ln(Y_t/Y_0)}{t}\right] = 1+r$$

$$(A.21) \quad r = \exp\left[\frac{\ln(Y_t/Y_0)}{t}\right] - 1$$

Table A.3 shows the calculation steps and you can paste the data into a spreadsheet and see if you can derive the same results by replicating the solution steps. The calculations show that the annual average compound growth real GDP growth rate for Australia between 1960 and 2012 was 3.52 per cent.

You can use this formula for any period or data frequency (for example, month, quarter, year) by substituting the appropriate information into the calculation.

**Table A.3 Compound growth rate calculations**

Real GDP 1960 \$m	$Y_1$	249083
Real GDP 2012 \$m	$Y_0$	1508267
Time periods (years)	$T$	52
GDP2012/GDP1960	$= (1 + r)t$	6.0553
Take logs	$= t \ln(1 + r)$	1.8009
Divide 52	$= \ln(1 + r)$	0.0346
Antilog	$= 1 + r$	1.0352
$r$		0.0352
$r^*$ per cent		3.5240

## A.8 Textbook Policy Regarding Formalism

We recognise that different students have different ways in which they learn and accumulate knowledge. Some prefer the mathematical approach while others prefer the graphical approach. Others still learn better through reading the written word, even though that form of communication is prone to interpretative issues.

In that regard, all the essential material in the text will be presented in all three ways (sometimes the mathematics will appear in the Appendix of the relevant chapter sometimes within the main body of the text).

We always aim to promote understanding and believe that a student is entitled to learn in the way that best suits their own proclivities.

However, it is also the case that professional economists use a variety of methods including numerical, graphical, algebraic and narrative in their work and we believe it is important to expose students to the broad range of methods employed in the real world.