This book focuses on the rapidly growing research field of imperfect competition, asymmetric information and other market imperfections in a macroeconomic context. It brings together leading researchers from the USA and Europe to examine the implications for macroeconomic policy of imperfections in output, labour and financial markets. The contributions include state-of-the-art research at the frontier of the discipline, as well as several general surveys and expository chapters which synthesize the large literature. This is the first volume of previously unpublished research papers to focus exclusively on this literature. It should be a valuable resource for graduate students and researchers in macroeconomics.

The new macroeconomics: imperfect markets and policy effectiveness

The new macroeconomics: imperfect markets and policy effectiveness

EDITED BY
HUW DAVID DIXON
University of York
and
NEIL RANKIN
University of Warwick

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

First published 1995

A catalogue record for this book is available from the British Library

Library of Congress cataloguing in publication data

The new macroeconomics: imperfect markets and policy effectiveness / edited by Huw Dixon and Neil Rankin.
p. cm.
'This book arose out of the seventeenth Summer Workshop held at Warwick University on 12–30 July 1993, with the same title as this volume' – CIP pref.
ISBN 0 521 47416 7. – ISBN 0 521 47947 9 (pbk.)
1. Competition, imperfect. 2. Economic policy.
3. Macroeconomics. I. Dixon, Huw. II. Rankin, Neil.
HB238.N485 1995
339 – dc20 94-37308 CIP

ISBN 0 521 47416 7 hardback
ISBN 0 521 47947 9 paperback

Transferred to digital printing 2004

SE

Contents

Contributors

Daron Acemoglu, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA

Torben Andersen, Institute of Economics, Universiteit van Aarhus, Universitetsparken, 8000 Aarhus C, Denmark

Jean-Pascal Bénassy, CEPREMAP, 142 Rue du Chevaleret, 75013 Paris, France

Russell Cooper, Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, USA

Claude d'Aspremont, CORE, Université Catholique de Louvain, 34 Voie du Roman Pays, 1348 Louvain-la-Neuve, Belgium

Huw David Dixon, Department of Economics, University of York, Heslington, York YO1 5DD, UK
E-mail: hdd1@york.ac.uk

Rodolphe Dos Santos Ferreira, BETA, Faculté des Sciences Economiques et de Gestion, Université Louis Pasteur, 67000 Strasbourg, France

George Evans, Department of Economics, University of Oregon, Eugene, OR 97403, USA

Jeff Frank, Royal Holloway, University of London, Egham, Surrey, TW20 0EX, UK

Louis-André Gérard-Varet, GREQAM, Centre de la Vieille Charité, 2 Rue de la Charité, 13002 Marseille, France

Brian Hillier, Department of Economics, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

Morten Hviid, Department of Economics, University of Warwick, Coventry CV4 7AL, UK

Andrew John, Department of Economics, 114 Rouss Hall, University of Virginia, Charlottesville, VA 22901, USA

Omar Licandro, FEDEA, Calle Jorge Juan 46, 28001 Madrid, Spain

Garey Ramey, Dept of Economics, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0508, USA

Neil Rankin, Department of Economics, University of Warwick, Coventry CV4 7AL, UK
E-mail: n.rankin@csv.warwick.ac.uk

Gilles Saint-Paul, DELTA, Ecole Normale Supérieure, 48 Boulevard Jourdan, 75014 Paris, France

Christian Schultz, Institute of Economics, University of Copenhagen, Studiestraede 6, DK-1455, Copenhagen-K, Denmark

Richard Startz, Department of Economics, University of Washington, Seattle, WA 98195, USA

Jon Strand, Department of Economics, University of Oslo, Box 1095, Blindern, 0317 Oslo, Norway

Alan Sutherland, Department of Economics, University of York, Heslington, York YO1 5DD, UK

Tim Worrall, Department of Economics, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, UK

Preface

This book arose out of the seventeenth Summer Workshop held at Warwick University on 12–30 July 1993, with the same title as this volume. We had felt for some time that there was a need for a conference to bring together some of the people involved in the study of various 'market imperfections' and their implications for macroeconomics. We had become aware of the rapidly burgeoning literature in the particular area of imperfect competition and macroeconomics in the course of writing our survey (Chapter 2 in this volume) over the period 1988–91. We wanted to organize a conference that would explore this along with other topics: financial market imperfections, bounded rationality, and menu costs, to name but a few. The idea coalesced with some encouragement and prodding from Marcus Miller, and our idea was generously funded by the Economic and Social Research Council of the UK, and by the Human Capital and Mobility Programme of the EU Commission.

The workshop was a great success, and we would like to thank all those who came for making it so. All of the contributors to this volume attended the workshop. The papers appearing here were either given at the workshop, or were 'commissioned' there specially for the volume, with the exception of two reprints (Chapters 2 and 4) which are included because they particularly complement the general theme. We would like to thank all the authors for their discipline and goodwill in enabling us to send the final manuscript off to the Press by July 1994, only twelve months after the workshop.

The Warwick conference was to be the first of several exploring the themes in this book. In January 1994, the CNRS and French Ministry of Finance organized a conference on 'Recent Developments in the Macroeconomics of Imperfect Competition', and in July 1994 a conference funded by the EU Human Capital and Mobility Programme (and others) was held at Carlos III University, Madrid, on 'Alternative Approaches to Macroeconomics'. These three conferences together have served to create a momentum and sense of common purpose amongst economists from both sides of the Atlantic working in these areas. We very much hope that this volume captures some of the excitement and atmosphere of these conferences (the proceedings of the Paris conference are to be published in a special edition of the *Annales d'Economie et de Statistique*).

Lastly, we would like to thank the various people who have helped us, both in organizing the workshop, and in putting together the book. At Warwick these include Marcus Miller, the workshop chairman, and Mandy Eaton, the invaluable secretary to many Warwick workshops; also the Economics Department chairman Nick Crafts, organizing committee member Jonathan Thomas, and members of the Economics Department secretariat who helped out during the three-week event. At York University, Marta Aloi and Michele Santoni very efficiently prepared the index. At Cambridge University Press we are very grateful to Patrick McCartan for his help and encouragement along the path to publication.

Acknowledgements

We would like to record our deepest thanks to the Economic and Social Research Council of the UK (grant no. H50126501293) and to the Human Capital and Mobility Programme of the EU Commission (contract no. ERB-CHEC-CT-92-0051), who funded in equal part the 1993 Warwick Summer Workshop out of which this book arises. We also warmly thank Oxford University Press for permission to reprint Dixon and Rankin's article from *Oxford Economic Papers*, and the University of Toronto Press for permission to reprint Cooper's article from the *Canadian Journal of Economics*.

E·S·R·C
ECONOMIC
& SOCIAL
RESEARCH
COUNCIL

Introduction

Huw David Dixon and Neil Rankin

> A landing on the non-Walrasian continent has been made. Whatever further exploration may reveal, it has been a mind-expanding trip: we need never go back to $\frac{dp}{dt} = \alpha(D - S)$ and $q = \min(S, D)$
>
> E.S. Phelps and S.G. Winter, in Phelps *et al.* (1970, p.337)

There have always been many streams of thought in macroeconomics. In this volume we have brought together what seem to us to be several convergent streams which deserve the title 'The New Macroeconomics'. This title is chosen to be consonant with the terms 'New I.O.' and 'New International Trade Theory'. In each case, the adjective 'new' has referred to the transformation of an existing area of economics by a shift of approach, the introduction of new microeconomic theory. In the case of the New I.O. in the early 1980s, this was the first field of economics to apply (and develop) the then recent developments in game theory in the late 1970s. In the case of the New International Trade Theory, it was a shift away from the Walrasian paradigm of price-taking agents in competitive markets towards the Brave New World of imperfectly competitive firms in a strategic environment. It is indeed a Brave New World, if only because it is less familiar and more heterodox.

One of the main themes in the New Macroeconomics has been the shift towards a macroeconomics based on microfoundations with market imperfections of one kind or another. As in International Trade Theory, this has primarily led to a focus on imperfect competition itself, either in labour or product markets (or both). However, the emphasis has also been on integrating imperfections in financial and labour markets based on imperfect/asymmetric information in these markets. Most recently, there has been a growing interest in the issue of bounded rationality in macroeconomics (see for example Sargent, 1993, for an excellent summary). Whilst we would subsume all of the 'New Keynesian' literature under this umbrella, we have avoided the term here, and opted for a more general concept. The reason for this is that we do not believe that the New Macroeconomics is necessarily or inevitably 'Keynesian' in its flavour. Whilst the title of this volume indicates that we might have a prior that market imperfections may tend to lead to policy effectiveness, we certainly do not believe this to be a general truth. The functioning of imperfect markets is of necessity more diverse than 'perfect' markets, and hence can sustain a wide variety of macroeconomic properties.

This volume aims to provide a snapshot of what to us are some of the more interesting developments in the New Macroeconomics.[1] There are three sorts of studies in this volume. First, there are overviews and surveys of the area. These are in part I: our own survey (chapter 2) focuses on the key macroeconomic issues of the positive and welfare effects of monetary and fiscal policy; Bénassy's chapter 1 investigates whether imperfectly competitive macromodels have properties of a more 'classical' or 'Keynesian' nature; Dick Startz in chapter 3 provides a very lucid perspective on imperfect competition and specifically 'New Keynesian' economics. Secondly, there are three specially commissioned chapters written in 'handbook' style, concentrating on a clear exposition of the main ideas, suitable for graduate students and non-specialists in the area. These are Hillier and Worrall's chapter 12 on the macroeconomic implications of financial market imperfections, Alan Sutherland's chapter 16 on the macroeconomic implications of menu costs on the micro level, and Andrew John's chapter 6 on externalities. Thirdly, there are ten original research papers at the forefront of the discipline (chapters 4, 5, 7–11, 13–15). We thus hope that there is something in this volume for everyone, and that it will provide a useful resource for those interested in new developments in macroeconomics, from graduate students upwards.

The volume is divided into five parts. Part I contains the overviews and perspectives on the general field. The subsequent parts collect together the studies on specific market imperfections: goods market imperfections in Part II, labour market imperfections in Part III, financial market imperfections in Part IV, and lastly the themes of bounded rationality and nominal rigidities in Part V. In this Introduction, we will discuss the main ideas and a little of their history and timing. We are not historians of economic thought, but we believe that the discipline needs some historical perspective, and so provide a sketch of the ideas. This is very much a personal perspective, and should not be taken as a definitive piece of scholarship.

Imperfect competition and macroeconomics: a brief history

Whilst the idea that imperfect competition might be important for macroeconomics is as old as macroeconomics (dating back to Kalecki), the formalization of this idea had to wait until the late 1970s.[2] The microeconomics of a monetary Walrasian economy (by which we mean one in which prices clear markets so that supply equals demand) was well developed in the 1950s, most notably with Don Patinkin's *Money, Interest and Prices*, which laid the foundation for the Neoclassical synthesis.[3] Several writers (notably Clower, 1965 and Leijonhufvud, 1968 and including to some extent Patinkin himself) saw the foundations of macroeconomics as based on something different from the Walrasian equilibrium model.

The first full formalization of this 'non-market-clearing' approach in a macroeconomic context was Barro and Grossman (1971, 1976). They formulated a simple general equilibrium model which kept the Walrasian assumption of price-taking by firms and households, but assumed that prices were exogenously given at non-market-clearing levels (the Walrasian market-clearing outcome was a special case). It was Jean-Pascal Bénassy's Ph.D. thesis at Berkeley under Gerard Debreu (Bénassy, 1973),[4] which fully integrated this non-market-clearing approach to macroeconomics with traditional microeconomics. The advantage of this non-market-clearing approach was that it enabled us to understand the phenomena of effective demand, the multiplier and involuntary unemployment which were (and are) seen by some as central to understanding macroeconomic phenomena. This approach became known as the fix-price approach (after John Hicks' distinction between fix and flex price models), and was popularized by Edmond Malinvaud in his *The Theory of Unemployment Reconsidered* (1977).

The obvious problem with the fix-price approach is that it treats prices as exogenous. Apart from the analysis of the then centrally planned economies,[5] this assumption really only applies to a transitory state, a 'temporary equilibrium' (as is clear in Bénassy's work). The key question naturally arose of how you make prices endogenous. This is an issue for Walrasian as well as fix-price models. Whilst it is straightforward to treat prices as endogenous in a mathematical sense in a supply and demand model, it is very difficult to model it as the outcome of an economic process. Arrow (1959) pinpointed the problem in his paradox: the model of perfect competition is based on the assumption that all agents act as price-takers, yet it requires someone to make the prices, ensuring that prices adjust to bring supply and demand into balance. The microfoundations of perfect competition are an obscure and difficult subject.

One obvious solution to the problem of how to make prices endogenous is to introduce price-setting[6] agents into the macroeconomic system. Essentially, this means introducing an alternative equilibrium concept to the common Walrasian one. There had been some attempt to do this in a microeconomic general equilibrium setting (most importantly Negishi, 1961; Arrow and Hahn, 1971, pp.151–68; Marshak and Selten, 1974).

However, the integration of price-setting into a macroeconomic setting occurred in the mid-1970s with a series of studies (Bénassy, 1976, 1978; Grandmont and Laroque, 1976; Negishi, 1978, 1979). All of these studies adopted the 'subjective demand curve' approach. In essence, the firm had a conjectured demand curve, which was tied down only by the 'Bushaw–Clower condition' that it passed through the actual price–quantity pair. This literature never prospered: the dynamics of the models rested on the way firms updated their subjective demand curves, and hence were to some extent arbitrary. Hahn (1978) tried to tie down the subjective conjectures by some notion of 'rationality', but his solution produced a multiplicity of equilibria, and was shown to rest crucially on an arbitrary property of non-differentiability in conjectures (see the comment by Gale, 1978).

The breakthrough that was to pave the way for subsequent work was Oliver Hart's study (1982), which was first circulated in an earlier version (1980). In this, Hart introduced and made operational the notion of an *objective* demand curve in a general equilibrium setting (see Hart, (1985b) for a full discussion). The basic idea was to define the demand curve facing the agents (in Hart's case Cournot oligopolists and quantity-setting unions) in terms of the actual consumer demand under certain well specified assumptions about what was constant as agents varied their quantities (e.g. consumer incomes, prices in other markets). The term 'objective' is perhaps a little extreme, as there is an essential arbitrariness in the assumptions made to derive the demand curve. However, the degree of arbitrariness seemed much smaller than with subjective demand curves, and the concept very rapidly caught on.

Introducing imperfect competition into the macroeconomic model had powerful implications. Just as monopoly or trade unions restrict output and employment in microeconomic models, so they might in the macrocontext. However, one of the most important differences is in welfare analysis: imperfectly competitive equilibria are of their nature usually socially inefficient. Hence the monopolist will set its price in excess of marginal cost, and the union will restrict employment in order to raise wages above the (marginal) disutility of labour. This inefficiency stands in stark contrast to the Walrasian equilibrium, which is in general Pareto optimal (from the First Fundamental Theorem of Welfare Economics). This inefficiency can take the form of involuntary unemployment. Furthermore, the social inefficiency of equilibrium gives rise to the exciting *possibility* that if monetary or fiscal policy is able to raise output and employment, it may give rise to an increase in welfare. Some of these possibilities were realized in Oliver Hart's study, which thus had as part of its title '...with Keynesian Features', the Keynesian Features being a multiplier and involuntary unemployment.

The development after Hart's study was rapid. There are several different strands of analysis. We shall deal with them in turn (we do not mean to suggest any priority by our order). The first strand concerned monopolistic competition and menu costs – the so-called PAYM model after Parkin (1986), Akerlof and Yellen (1985a), and Mankiw (1985). In essence, these models used menu costs/bounded rationality to provide some nominal price (wage) rigidity using the fact that a price- (wage)-setting firm would set prices (wages) at an optimal level.[7] The second set of studies was concerned with the issue of the effect of imperfect competition on the size of the fiscal multiplier in an economy with a competitive labour market (Dixon, 1987; Mankiw, 1988; Startz, 1989[8]). These models all came up with the *profit-multiplier* relationship, that is the notion that imperfect competition in the output market leads to a higher multiplier because imperfect competition leads to higher profit margins, profits increase with output, and hence a feedback multiplier occurs. These results are, however, very sensitive to assumptions about technology and preferences (see Dixon and Lawler, 1993), and open to different interpretations (Dixon describes this multiplier as Walrasian since welfare is decreasing with output, whilst Mankiw and Startz interpret it as Keynesian).

The menu cost and multiplier studies had concentrated on imperfect competition in *output* markets only. It was Snower (1983), d'Aspremont *et al.* (1984), Bénassy (1987), Blanchard and Kiyotaki (1987)[9] and Dixon (1988) who developed the implications of imperfect competition in a *unionized* economy. In this setting, there is explicitly involuntary unemployment, and output increases can in principle generate welfare increases due to this. However, all of the studies with competitive or unionized labour markets agree that monetary policy is neutral (see Bénassy, 1987 for the most general statement here) without some feature overriding the underlying homogeneity of demand equations.

In parallel with the preceding developments, and largely separate from it, was Hall's work on imperfect competition in response to the Real Business Cycle (RBC) movement. Hall (1986, 1988) pointed out that fluctuations in the 'Solow residual' are not exclusively attributable to productivity once price exceeds marginal cost, thus weakening a key part of the empirical case in favour of RBC theory (the mainstream RBC literature was exclusively Walrasian in its assumptions – see Plosser, 1989, for a survey). However, it is only recently that imperfect competition has become common in RBC models (for example, Rotemberg and Woodford 1992; Horstein, 1993).

This is as far as we will go: the scene is set for our survey below to go into the details of some of these and subsequent developments in imperfect competition. The chapters in Part II of the volume, and Omar Licandro's chapter 7 in Part III are examples of the current state of play in this area.

Contract theory

The response of mainstream Keynesian macroeconomists to the New Classical Policy Ineffectiveness propositions of the mid-1970s was the theory of overlapping wage contracts. Gray (1976), Fischer (1977), Taylor (1979) and others put forward variants of this story. However, their models lacked serious microfoundations, and did not offer rigorous explanations of the form of contract, or how its level was arrived at. A second type of literature on contracts which developed slightly earlier was the 'implicit contract' approach initiated by Baily (1974) and Azariadis (1975). Implicit contract models aimed to explain the rigidity of *real* wages and hence the existence of an equilibrium level of unemployment. The argument was based on the assumption that workers are more risk-averse than firms, and so optimal risk sharing dictates that the firm insures the workers by keeping wages fixed over the business cycle. The difficulty with this from the macroeconomists' perspective was that unless there is some restriction on the form of contract or asymmetric information (on the latter, see Grossman and Hart, 1981), the employment level is first best, notwithstanding any wage rigidity. Furthermore, all of the early implicit contract models were microeconomic and essentially partial equilibrium.

Much work has been done on the theory of contracts since the initial studies, and the notion of implicit or explicit contracts is central to several chapters in this volume. Chapters 4 and 9 by Cooper and by Schultz make use of 'efficient' contracts in the labour market, and in Schultz's case the analysis rests on the characterization of self-enforcing contracts. Acemoglu (chapter 13) looks at the role of contracts in creating *nominal* rigidities, by considering agents' choices between nominal and real contracts motivated by the desire to hedge against price fluctuations. Contracts in finance markets (debt contracts) also lie behind the theories of the business cycle surveyed in Hillier and Worrall's chapter 12.

Efficiency wages

The theory of efficiency wages has its origins in the analysis of labour markets in Less Developed Countries (see, for example, Leibenstein, 1957; Stiglitz, 1976). The efficiency wage model belongs to the general class of monopsony labour market models, and provides an equilibrium derived under the assumption that firms set wages under conditions of asymmetric information, turnover costs and so on. The first application of the theory of efficiency wages to a macroeconomic context in developed countries came later: Solow (1979), Weiss (1980), Shapiro and Stiglitz (1984). The common thread of all of the efficiency wage stories is that the firm may choose to set a wage which results in unemployment (a queue for employment). The basic intuition for this is that there is a positive relationship between the wage offered and the 'quality' of labour (this 'quality' can be productivity, effort, or propensity to quit). Offering a lower wage may therefore lead to a lower 'quality', and this may mean that despite queues for jobs ('involuntary unemployment'), firms may not wish to lower the wage (and hence the quality) paid to workers.

The difference between this literature and the standard 'imperfect competition' work is that it has tended to focus almost entirely on the labour market and adopted a partial equilibrium framework. For a treatment that integrates efficiency wages with the temporary equilibrium approach, see Picard (1993, chs. 6–8), and for the explicit implications of efficiency wages for the business cycle see Danthine and Donaldson (1990) and Strand (1992a). Chapter 8 by Gilles Saint-Paul extends this literature by showing how an efficiency wage mechanism can affect not only the *level* and *variability* of unemployment, but also account for its *persistence.* Jon Strand's chapter 10 explores the issue of the *timing* of environmental policy intervention in the business cycle. Efficiency wage theory remains one of the main theories of unemployment, and its full potential in an explicitly general equilibrium macroeconomic framework is yet to be realized.

Credit market imperfections

That financial factors such as bankruptcies and indebtedness play a major role in economic fluctuations is a truism amongst non-economists. However, it was not until the 1980s that satisfactory formal models were developed to explore these factors (as opposed to the merely 'monetary' models). This is not to belittle the important contributions of many economists earlier in the century, most notably Irving Fisher (1933) and his work on debt deflation, but such work was not central to mainstream economics.[10] Indeed, much of the formal finance literature stressed the *irrelevance* of financial factors for real decisions such as investment: most notably the Modigliani–Miller Theorem (1958). The rapid development of the literature on asymmetric information in the 1980s opened up this field, as in the case of efficiency wages.

If lenders cannot costlessly observe the relevant characteristics of borrowers, banks might choose to set an interest rate which results in an excess demand for credit (Stiglitz and Weiss, 1981). This of course has an explicitly monetary flavour, and can be given a directly macroeconomic context: see, for example, Greenwald and Stiglitz (1993), Stiglitz (1992). In these circumstances the borrower's net worth becomes a key determinant of net investment and hence future net worth. This provides a key mechanism for the propagation of shocks through time. The non-linearities created by asymmetric information also make multiple equilibria and cycles likely, so that fluctuations can be self-sustaining.

Coordination failures

The concept of coordination failures is in some ways an old one: it is clearly present in Leijonhufvud's reappraisal of Keynes, for example (Leijonhufvud, 1968). However, the concept was first formalized by Cooper and John (1988),[11] using the concept of strategic complementarity developed by Bulow *et al.* (1985). The basic idea is simple: if the activities of agents are strategic complements, then the individual level of activity is an increasing function of the aggregate level of activity. If we add to this the feature of a 'spillover' or 'externality' (we will not enter the debate about how these words should be used!), then multiple Pareto-ranked equilibria are possible.[12] Although this idea was formalized in the late 1980s, it has proven a powerful tool for understanding previous work: in particular, the multiple equilibria in Diamond's coconut model of search (Diamond, 1982).[13] Andrew John's chapter 6 provides a very useful classification of types of externalities and the similarities and differences between externalities in models of imperfect competition and search models. d'Aspremont *et al.* in chapter 5 also explore the issue of coordination failure in the context of an overlapping-generations model of imperfect competition.

New Classical and New Keynesian economics

All of the strands we have identified above have been designated 'New Keynesian' by Greg Mankiw and David Romer (1991). The adjective 'New Keynesian' has in fact had a variety of interpretations (see Gordon, 1990; Ball *et al.*, 1988; Frank, 1986). Just as the New Classical school is based in a few US economics departments, so the 'New Keynesians' are based in a few departments, mostly on the east coast (Massachusetts in particular). However, as should be clear from the foregoing analysis, the 'Keynesian' concerns have been continuous in Europe and elsewhere, due to the lesser dominance of New Classical thought. In particular, due to obvious empirical and institutional differences, the study of imperfect competition (particularly in labour markets) has always had a higher prominence in European countries: the assumption of perfect competition seems more unpalatable and irrelevant. Just as in the 1970s Bénassy described his own work on temporary fix-price equilibria as 'neo-Keynesian', since it was bringing new microeconomic theory to traditional Keynesian concerns, the term 'New Keynesian' is a useful catch-all term for some of the developments we have discussed above.

Of course, the developments we are considering were occurring alongside the RBC phase of the New Classical school.[14] This shared the desire for microeconomic foundations, but found them in the intertemporal microeconomics of perfect competition: the macroeconomics of perfect markets. It is worth noting that the New Classical School has been primarily a US phenomenon, and a fresh-water one at that. Whilst it certainly caught on as the macroeconomic aspect of the free market ideology of some major governmental and international bodies, it was never popular amongst active academic researchers in Europe.

There was a clear contradistinction between New Classical and New Keynesian analysis in the 1980s. First, they used different microeconomic theories. The New Classical school put the emphasis on the competitive, the intertemporal and the more dynamic; the New Keynesian put the emphasis on the imperfectly competitive and more static. Second, they came up with different views about the working of the market: the New Classical economists saw fluctuations as efficient, resulting from the optimal response of consumers and firms to taste and technology shocks; the New Keynesians saw the level and size of fluctuations as resulting from market failure, and hence not efficient.

Towards a New Macroeconomics

If we put together these developments in macroeconomics during the 1980s, we believe that there is a clear movement towards the exploration of market imperfections and their macroeconomic implications. The seeds of this move were clearly sown sometime in the past: our whistlestop guide has not even had time to mention Phelps *et al.*'s influential *The Microeconomic Foundations of Macroeconomics* (1970) which has inspired many economists of differing persuasions. However, some time in the early to mid-1980s, economists (established researchers and Ph.D. students) started to explore some of these ideas in a formal and coherent way. The coherence of the New Macroeconomics stems from its integration of microeconomics into macroeconomic theory, and a willingness to try out new ideas.

The New Macroeconomics has a genuine transatlantic and indeed world-wide base. Insofar as macroeconomists are interested in the proper economic analysis of phenomena, rather than judging theory on its policy conclusions, the differences and classifications based on policy (Keynesian and Classical) will become increasingly blurred. To some extent this is already happening. RBC theory can be written with imperfectly competitive foundations (see, for example, Hairault and Portier, 1993), and competitive models can be written with boundedly rational agents (see Evans and Ramey's chapter 15 in this volume) or other imperfections. The recent explosion of interest in endogenous growth models is an obvious case in point. It has aspects of the 'New Classical' (the dynamic and intertemporal), and the 'New Keynesian' (market failure, externality, increasing returns, imperfect competition) approach. The leading researchers in this field have included economists who are (were?) New Classical (Robert Lucas and Robert Barro, for example), and more Keynesian economists (Larry Summers, for example). It seems to us that the boundary between the two schools of the 1980s is rapidly disappearing in the 1990s.

Ever since Keynes 'invented' macroeconomics as a discipline, there have always been attempts to integrate the macroeconomic with the microeconomic. This process has been accelerating over the last two decades, and we believe that the exploration of imperfect markets will lie at the heart of the enterprise in the years to come. This volume provides a snapshot of some of the more interesting work being done in this area. We very much hope that macroeconomic theorists will prove willing to explore this Brave New World in all its diversity. It has taken longer than anticipated by Edmund Phelps and Sydney Winter, but a quarter of a century after their prophecy, it is clear that the trip continues to expand minds.

Notes

1. The most obvious omissions are endogenous growth theory and the policy games literature. We felt that these were well covered elsewhere.
2. Although there are some exceptions: see, for example, Ball and Bodkin (1963) who introduced an imperfectly competitive labour demand curve into an otherwise standard aggregate supply function.
3. For a discussion of the tremendous impact of this book on the macroeconomics of the subsequent 25 years see Dixon (1994c).
4. Bénassy (1975, 1976 and 1978) all arose out of his thesis.
5. See John Bennett (1990) for an excellent exposition of the treatment of centrally planned economies using the fix-price approach. On the issue of transitional economies, see Bennett and Dixon (1993).
6. We use the term 'price-setting' here as a shorthand for wage- and price-setting.
7. Alan Sutherland (chapter 16) reviews the link between menu costs and nominal rigidity at the microeconomic level with *aggregate* price dynamics. As Caplin and Spulber (1987) showed, aggregation is a crucial issue here.
8. Note that the eventual publication dates are a trifle misleading due to refereeing and publication lags. Dixon's study was first published as Birkbeck College *Discussion Paper*, 186 in April 1986, and was based on lectures given at

Birkbeck in March 1985. Startz was circulating as a mimeo in 1985. Mankiw's study was an NBER *Discussion Paper* in 1987.

9. Blanchard and Kiyotaki arose out of Chapter 1 of Kiyotaki's Ph.D. Dissertation 'Macroeconomics of Monopolistic Competition' (Harvard University, May 1985) and a separate unpublished paper by Olivier Blanchard (who was the supervisor). It was Nobuhiro Kiyotaki's thesis which developed the application of the Dixit–Stiglitz model of monopolistic competition with CES preferences in a general equilibrium macroeconomic framework.
10. There is also the work by writers such as Paul Davidson (1972) and Hyman Minsky (1976), which can be broadly characterized as 'post-Keynesian' and, although largely informal, focused on these issues.
11. This was first written up as Cowles Foundation *Discussion Paper*, 745 (April 1985).
12. There has also been empirical work which supports the existence of multiple equilibria in the UK economy (Manning, 1990, 1992).
13. It also underlies Weitzman's notion of multiple underemployment equilibria (Weitzman, 1982), which inspired Solow (1986) and Pagano (1990).
14. In fact the model of the real business cycle with reasonably complete microfoundations was Kydland and Prescott (1982). Earlier studies were largely ad hoc and incomplete (although of course the seed of the idea goes back to Lucas and Rapping, 1969). The full intertemporal competitive general equilibrium macromodel was developed in the mid-1980s (for example, Prescott, 1986, and see Sargent, 1987 for a full exposition).

Part I

Overviews and perspectives

1 Classical and Keynesian features in macroeconomic models with imperfect competition

Jean-Pascal Bénassy

Introduction

Recent years have seen a rapidly growing development of macroeconomic models based on imperfect competition. A strong point of these models is that they are able to generate inefficient macroeconomic equilibria, obviously an important characteristic nowadays, while maintaining rigorous microfoundations. Indeed in these models both price and quantity decisions are made rationally by maximizing agents internal to the system, which thus differentiates them from Keynesian models, where the price formation process is *a priori* given, and also from classical (i.e. Walrasian) models, where the job of price-making is left to the implicit auctioneer.

Since for many years the macroeconomic debate has been dominated by the 'classical versus Keynesian' opposition, a question often posed by various authors, both inside and outside the domain, is whether these macroeconomic models with imperfect competition have more 'classical' or 'Keynesian' properties. The debate on this issue has sometimes become rather muddled and the purpose of this chapter is to give a few basic answers in a simple and expository way. This we shall do not by reviewing all contributions to the subject (there are already two excellent review articles, by Dixon and Rankin, 1994 and Silvestre, 1993), but by constructing a simple 'prototype' model with rigorous microfoundations, including notably rational expectations and objective demand curves, and examining how its various properties relate to those of Keynesian and classical models. Before that we shall make a very quick historical sketch of how these models developed in relation to the 'classical versus Keynesian' strands of literature.

A brief history

The initial results derived from macromodels with imperfect competition had a distinct Keynesian flavour, perhaps because the first models started from the desire to give rigorous microfoundations to models generating underemployment of resources. Negishi (1978) showed how under kinked demand curves some Keynesian-type equilibria could be supported as imperfect competition equilibria. Bénassy (1978) showed that non-Walrasian fix-price allocations could be generated as imperfect competition equilibria with explicit price-setters. It was shown in particular that generalized excess supply states of the Keynesian type, with all the ensuing inefficiency properties, would obtain if firms were setting the prices and workers the wages.

Policy considerations were brought in by Hart (1982), who constructed a Cournotian model with objective demand curves, which displayed 'Keynesian' responses to some policy experiments. These intriguing Keynesian results stirred much interest in the field, but soon after researchers began to realize that the most 'Keynesian' policy results were due to somewhat specific assumptions, and the next generation of studies showed that policy responses were of a much more 'classical' nature: Snower (1983) and Dixon (1987) showed that fiscal policies had crowding-out effects fairly similar to those arising in classical Walrasian models. Bénassy (1987), Blanchard and Kiyotaki (1987) and Dixon (1987) showed that money had the same neutrality properties as in Walrasian models. Although normative policies were seen to differ from classical ones (Bénassy 1991a, 1991b), we shall see below that this was not in a Keynesian manner.

As of now the common wisdom (although not a unanimously shared one), seems to be that standard imperfect competition models generate outcomes which display inefficiency properties of a 'Keynesian' nature, but react to policy in a more 'classical' way. If one wants to obtain less 'classical' results, one has to add other 'imperfections' than imperfect competition, such as imperfect information or costly price changes, to quote only two. Since the initial venture by Hart in this direction, many different models have been proposed. Because space is scarce and opinions as to which is the most relevant imperfection are highly divergent, we shall not deal at all with these issues, which are aptly surveyed in Dixon and Rankin (1994) and Silvestre (1993), and turn to the description of our simple prototype model and its properties, which will confirm and expand the 'common wisdom' briefly outlined above.

The model

In order to have a simple intertemporal structure, we shall study an overlapping-generations model with fiat money. Agents in the economy are households living two periods each and indexed by $i = 1, \ldots, n$, firms indexed by $j = 1, \ldots, n$, and the government.[1]

There are three types of goods: money which is the numéraire, medium of exchange and unique store of value, different types of labour, indexed by $i=1,\dots,n$, and consumption goods indexed by $j=1,\dots,n$. Household i is the only one to supply labour of type i, and sets its money wage w_i. Firm j is the only one to produce good j and sets its price p_j. We shall denote by P and W the price and wage vectors:

$$P=\{p_j|j=1,\dots,n\}$$
$$W=\{w_i|i=1,\dots,n\}$$

Firm j produces output y_j using quantities of labour ℓ_{ij}, $i=1,\dots,n$ under the production function:

$$y_j=F(\ell_j) \tag{1}$$

where F is strictly concave and ℓ_j, a scalar, is deduced from the ℓ_{ij}s via an aggregator function Λ:

$$\ell_j=\Lambda(\ell_{1j},\dots,\ell_{nj}). \tag{2}$$

We shall assume that Λ is symmetric and homogeneous of degree one in its arguments. Although all developments that follow will be valid with general aggregator functions (see the appendix) in order to simplify the exposition we shall use in the main text the traditional CES one:[2]

$$\Lambda(\ell_{1j},\dots,\ell_{nj})=n\left(\frac{1}{n}\sum_{i=1}^{n}\ell_{ij}^{\varepsilon-1/\varepsilon}\right)^{\varepsilon/(\varepsilon-1)}. \tag{3}$$

We may already note that to this aggregator function is naturally associated by duality theory an aggregate wage index w:

$$w=\left(\frac{1}{n}\sum_{i=1}^{n}w_i^{1-\varepsilon}\right)^{\varepsilon/(\varepsilon-1)}. \tag{4}$$

Firm j's objective is to maximize profits:

$$\pi_j=p_jy_j-\sum_{i=1}^{n}w_i\ell_{ij}.$$

Household i consumes quantities c_{ij} and c'_{ij} of good j during the first and second period of its life, and receives from the government an amount g_{ij} of good j in the first period. Also in the first period household i sets the wage w_i and works a total quantity ℓ_i given by:

$$\ell_i=\sum_{j=1}^{n}\ell_{ij}\leq\ell_0 \tag{5}$$

where ℓ_0 is each household's endowment of labour. Household i maximizes the utility function:

$$U(c_i, c_i', \ell_0 - \ell_i, g_i) \tag{6}$$

where c_i, c_i' and g_i are scalar indexes given by:

$$c_i = V(c_{i1}, \ldots, c_{in}) \tag{7}$$

$$c_i' = V(c_{i1}', \ldots, c_{in}') \tag{8}$$

$$g_i = V(g_{i1}, \ldots, g_{in}) \tag{9}$$

We assume that the function V is symmetric and homogeneous of degree one in its arguments. We may note that we use the same aggregator function for private and government spending so that our results will not depend, for example, on the difference between elasticities of the corresponding functions. Again for simplicity of exposition we shall use in the main text the traditional CES aggregator:

$$V(c_{i1}, \ldots, c_{in}) = n\left(\frac{1}{n}\sum_{j=1}^{n} c_{ij}^{(\eta-1)/\eta}\right)^{\eta/(\eta-1)} \tag{10}$$

to which is associated by duality the aggregate price index p:

$$p = \left(\frac{1}{n}\sum_{j=1}^{n} p_j^{1-\eta}\right)^{1/(1-\eta)}. \tag{11}$$

We shall assume that U is strictly concave and separable in (c_i, c_i'), $\ell_0 - \ell_i$ and g_i. We shall further assume that the isoutility curves in the (c_i, c_i') plane are homothetic and that the disutility of work becomes so high near ℓ_0 that constraint (5) is never binding. Household i has two budget constraints, one for each period of its life:

$$\sum_{j=1}^{n} p_j c_{ij} + m_i = w_i \ell_i + \pi_i - p\tau_i \tag{12}$$

$$\sum_{j=1}^{n} p_j' c_{ij}' = m_i \tag{13}$$

where m_i is the quantity of money transferred to the second period as savings, p_j' is the price of good j in this future period, τ_i is taxes paid to the government in real terms and π_i household i's profit income, equal to:

$$\pi_i = \frac{1}{n}\sum_{j=1}^{n} \pi_j. \tag{14}$$

The government purchases goods on the market and gives quantities g_{ij}, $j=1,\ldots,n$ to household i, allowing it to reach a satisfaction index g_i given by (9) above. It also taxes τ_i from household i, and we assume at this stage that these taxes are lump sum, in order not to add any distortion to the imperfect competition one.

Finally we shall denote by $\bar{m}_i$ the quantity of money that old household i owns at the outset of the period studied (which corresponds of course to its savings of the period just before).

Because the model so far is fully symmetric, we shall further assume:

$$g_i=g \qquad \tau_i=\tau \qquad \bar{m}_i=\bar{m} \qquad \forall i. \tag{15}$$

The imperfect competition equilibrium

As we indicated above, firm j sets price p_j, young household i sets wage w_i. Each does so taking all other prices and wages as given. The equilibrium is thus a Nash equilibrium in prices and wages. A central element in the construction of this equilibrium is the set of objective demand curves faced by price- and wage-setters, to which we now turn.

Objective demand curves

Deriving rigorously objective demand curves in such a setting obviously requires a general equilibrium argument (Bénassy, 1988, 1990). Calculations, which are carried out in the appendix, show that the objective demands for good j and labour i respectively are given by:

$$Y_j=\left(\frac{p_j}{p}\right)^{-\eta}\frac{1}{1-\gamma}\left[\frac{\bar{m}}{p}+g-\gamma\tau\right] \tag{16}$$

$$L_i=\left(\frac{w_i}{w}\right)^{-\varepsilon}\frac{1}{n}\sum_{j=1}^{n}F^{-1}(Y_j) \tag{17}$$

where $\gamma=\gamma(p'/p)$ is the propensity to consume out of current income and p' is tomorrow's price index. As an example, if the subutility in (c_i, c_i') is of the form $\alpha \log c_i + (1-\alpha) \log c_i'$ (which we shall use below), then $\gamma(p'/p)=\alpha$.

To make notation a little more compact, we shall denote functionally the above objective demand functions as:

$$Y_j=Y_j(P, W, \bar{m}, g, \tau, p') \tag{18}$$

$$L_i=L_i(P, W, \bar{m}, g, \tau, p'). \tag{19}$$

We should note for what follows that these functions are homogeneous of degree zero in P, W, $\bar{m}$ and p'.

Optimal plans

Consider first firm j. To determine its optimal plan, and notably the price p_j it will set, it will solve the following program (A_j):

$$\max p_j y_j - \sum_{i=1}^{n} w_i \ell_{ij} \quad \text{s.t.}$$

$$\left. \begin{aligned} & y_j = F(\ell_j) \\ & y_j \leq Y_j(P, W, \bar{m}, g, \tau, p'). \end{aligned} \right. \qquad (A_j)$$

We shall assume that this program has a unique solution, which thus yields the optimal price as:

$$p_j = \psi_j(P_{-j}, W, \bar{m}, g, \tau, p') \tag{20}$$

where $P_{-j} = \{p_k | k \neq j\}$.

Consider now young household i. Its optimal plan, and notably the wage w_i it will set, will be given by the following program (A_i):

$$\left. \begin{aligned} & U_i(c_i, c'_i, \ell_0 - \ell_i, g_i) \quad \text{s.t.} \\ & \sum_{j=1}^{n} p_j c_{ij} + \sum_{j=1}^{n} p'_j c'_{ij} = w_i \ell_i + \pi_i - p\tau \\ & \ell_i \leq L_i(P, W, \bar{m}, g, \tau, p') \end{aligned} \right. \qquad (A_i)$$

which, assuming again a unique solution, yields the optimal wage w_i:

$$w_i = \psi_i(W_{-i}, P, \bar{m}, g, \tau, p') \tag{21}$$

where $W_{-i} = \{w_k | k \neq i\}$.

Equilibrium

We can now define our imperfect competition equilibrium as a Nash equilibrium in prices and wages:

Definition: An equilibrium is characterized by prices and wages p_j^* and w_i^* such that:

$$w_i^* = \psi_i(W_{-i}^*, P^*, \bar{m}, g, \tau, p') \qquad i = 1, \ldots, n$$
$$p_j^* = \psi_j(P_{-j}^*, W^*, \bar{m}, g, \tau, p') \qquad j = 1, \ldots, n.$$

All quantities in this equilibrium are those corresponding to the fix-price equilibrium associated to P^* and W^*. Alternatively they are also given by the solutions to programs (A_i) and (A_j) on p.20, replacing P and W by their equilibrium values P^* and W^*.

Characterization and example

We shall assume that the equilibrium is unique. It is thus symmetric, in view of all the symmetry assumptions made. We shall have:

$$\ell_j=\ell \quad y_j=y \quad p_j=p \qquad\qquad \forall j$$
$$\ell_i=\ell \quad c_i=c \quad c_i'=c' \quad g_i=g \quad w_i=w \qquad \forall i$$
$$\ell_{ij}=\frac{\ell}{n} \quad c_{ij}=\frac{c}{n} \quad c_{ij}'=\frac{c'}{n} \quad g_{ij}=\frac{g}{n}. \qquad \forall i,j$$

Before studying the properties of our equilibrium, we shall derive a set of equations characterizing it, and give an example.

Characterizing the equilibrium

In order to derive the equations determining the imperfectly competitive equilibrium, we shall first use the optimality conditions corresponding to the above optimization programs of firms and households.

Consider first the program (A_j) of a representative firm j. At a symmetric point the Kuhn–Tucker conditions (recall that the objective demand curve has, assuming n is large, an elasticity of $-\eta$) yield:

$$\frac{w}{p}=\left(1-\frac{1}{\eta}\right)F'(\ell) \tag{22}$$

and the production function:

$$y=F(\ell). \tag{23}$$

Consider similarly the program (A_i) of a young representative household. At the symmetric equilibrium, calling λ the marginal utility of income, the Kuhn–Tucker conditions yield:

$$\frac{\partial U}{\partial c}=\lambda p \qquad \frac{\partial U}{\partial c'}=\lambda p' \tag{24}$$

$$\frac{\partial U}{\partial(\ell_0-\ell)}=\lambda w\left(1-\frac{1}{\varepsilon}\right) \tag{25}$$

and the budget constraint of this young household is written:

$$pc+p'c'=w\ell+\pi-p\tau=p(y-\tau) \tag{26}$$

We finally have the physical balance equation on the goods market:

$$c + c' + g = y \tag{27}$$

and the budget constraint of the representative old household:

$$pc' = \bar{m} \tag{28}$$

(22)–(28) describe the equilibrium. Before moving to the various properties of this equilibrium, we shall give a simple illustrative example.

An example

We shall now fully compute the equilibrium for the following Cobb–Douglas utility function:

$$U = \alpha \log c + (1 - \alpha) \log c' + \beta \log (\ell_0 - \ell) + v(g). \tag{29}$$

Solving first (24)–(26) we obtain the following relation characterizing the quantity of labour supplied by the young household:

$$\frac{w}{p}(\ell_0 - \ell) = \frac{\varepsilon}{\varepsilon - 1} \cdot \beta(y - \tau) \tag{30}$$

which together with (22) and (23) allows us to compute the equilibrium quantity of labour ℓ:

$$F'(\ell)(\ell_0 - \ell) = \frac{\varepsilon}{\varepsilon - 1} \cdot \frac{\eta}{\eta - 1} \beta[F(\ell) - \tau]. \tag{31}$$

Once ℓ is known, all other values are easily deduced from it:

$$y = F(\ell) \tag{32}$$

$$\frac{w}{p} = \frac{\eta - 1}{\eta} F'(\ell) \tag{33}$$

$$c = \alpha(y - \tau) \tag{34}$$

$$c' = (1 - \alpha)y + \alpha\tau - g \tag{35}$$

$$p = \frac{\bar{m}}{(1 - \alpha)y + \alpha\tau - g}. \tag{36}$$

Keynesian inefficiencies

Quite evidently the equilibrium obtained above is not a Pareto optimum, but we shall now further see that the nature of the allocation and its inefficiency properties look very much like those encountered in traditional Keynesian equilibria.

The first common point is that we indeed observe at our equilibrium a potential excess supply of both goods and labour. (22) shows that marginal cost is strictly below price for every firm, and thus that firms would be willing to produce and sell more at the equilibrium price and wage, provided the demand was forthcoming. Similarly (25) shows that the households would be willing to sell more labour at the given price and wage, if there were extra demand for it. We are thus, in terms of the terminology of fix-price equilibria, in the general excess supply zone.

Secondly, (16) and (17), which yield the levels of output and employment for given prices and wages, are extremely similar to those of a traditional Keynesian fix-price–fix-wage model. In fact (16) and (17) are a multisector generalization of the traditional one-sector Keynesian equations. Let us indeed take all prices equal to p, all wages equal to w. We obtain immediately:

$$y_j = \frac{1}{1-\gamma}\left[\frac{\bar{m}}{p} + g - \gamma\tau\right] = y \qquad \forall j$$

$$\ell_i = F^{-1}(y): \qquad \forall i$$

a most traditional 'Keynesian multiplier' formula.

We shall finally see that our equilibrium has a strong inefficiency property which is characteristic of multiplier equilibria (see, for example, Bénassy, 1978, 1990), namely that it is possible to find additional transactions which, at the given prices and wages, will increase all firms' profits and all consumers' utilities.

To be more precise, let us assume that all young households work an extra amount $d\ell$, equally shared between all firms. The extra production is shared equally between all young households so that each one sees its current consumption index increase by:

$$dc = dy = F'(\ell)d\ell. \qquad (37)$$

Considering first the representative firm, we see that, using (22), its profits in real terms will increase by:

$$d(\pi/p) = \frac{F'(\ell)d\ell}{\eta} > 0. \qquad (38)$$

Consider now the representative young household. The net increment in its utility is:

$$dU = \frac{\partial U}{\partial c}\cdot dc + \frac{\partial U}{\partial \ell}\cdot d\ell$$

which, using (22), (24) and (25) yields:

$$dU=\left[1-\left(1-\frac{1}{\eta}\right)\left(1-\frac{1}{\varepsilon}\right)\right]\frac{\partial U}{\partial c}F'(\ell)d\ell>0. \tag{39}$$

(38) and (39) show that the increment in activity leads clearly to a Pareto improvement.

All the above characterizations point to the same direction: at our equilibrium activity is blocked at too low a level, and it would be desirable to implement policies which do increase this level of activity. The traditional Keynesian prescription would be to use expansionary demand policies, such as monetary or fiscal expansions. (16) and (17) show us that, if prices and wages remained fixed, these expansionary policies would indeed be successful in increasing output and employment. But – and this is where resemblance with Keynesian theory stops – government policies will bring about price and wage changes which will completely change their impact. We shall now turn to this.

The impact of government policies

We shall now study the impact of two traditional Keynesian expansionary policies, monetary and fiscal policies, and show that, because of the price and wage movements which they induce, they will have 'classical' effects quite similar to those which would occur in the corresponding Walrasian model. One may get a quick intuitive understanding of such results by looking at (22)–(28) defining the equilibrium, and noticing that the corresponding Walrasian equilibrium would be defined by exactly the same equations, with ε and η both infinite. The similarity of the first order conditions explains why policy responses will be similar.

The neutrality of monetary policy

We shall now consider a first type of expansionary policy, a proportional expansion of the money stock which is multiplied by a quantity $\mu>1$. This is implemented here by endowing all old households with a quantity of money $\mu\bar{m}$ instead of $\bar{m}$. Although the analysis of this case may seem fully trivial at first sight in view of the homogeneity properties of the various functions, one must realize that all equilibrium values in the current period depend not only on the current government policy parameters $\bar{m}$, g and τ, but also on p', the future level of prices, and therefore on all future policy actions as well. To keep things simple at this stage, we will assume that the government will maintain constant fiscal policy parameters g and τ through time, and that the economy settles in a stationary state with constant real variables and inflation. In that case we have the following relation between p and p':

$$\frac{p'}{p}=\frac{c'+g-\tau}{c'}. \tag{40}$$

Combining (22)–(28) and (40), we find that an expansion of $\bar{m}$ by a factor μ will multiply p, w and p' by the same factor μ, leaving all quantities unchanged. Money is thus neutral, as it would be in the corresponding Walrasian model.

Fiscal policy and crowding-out

We shall now study the effects of other traditional Keynesian policies, i.e. government spending g and taxes τ. In order to avoid complexities arising when the current equilibrium depends on future prices, we shall discuss the example on p.22, where the current equilibrium depends only on current policies.

Although (31)–(36) allow us to deal with the unbalanced budget case as well, we shall concentrate here on balanced budget policies $g=\tau$, which have been the most studied in the literature. Let us recall (31), giving the equilibrium level of employment:

$$F'(\ell)(\ell_0-\ell)=\frac{\varepsilon}{\varepsilon-1}\cdot\frac{\eta}{\eta-1}\,\beta[F(\ell)-\tau]. \tag{31}$$

Taking $\tau=g$ and differentiating it we obtain:

$$\frac{\partial y}{\partial g}<1 \tag{41}$$

$$\frac{\partial y}{\partial g}>0. \tag{42}$$

(41) indicates that the balanced budget multiplier is smaller than one, and therefore that there is crowding-out of private consumption, just as in Walrasian models.

(42) has been the source of much confusion, leading some authors to believe that they had found there some underpinnings to the 'Keynesian cross'-multiplier (see, for example, Mankiw, 1988). Clearly the mechanism at work here has nothing to do with a Keynesian demand multiplier, but goes through the labour supply behaviour of the household: paying taxes to finance government spending makes the household poorer, and since leisure is a normal good here, the income effect will naturally lead the

household, other things being equal, to work more, thus increasing activity. We should note that this effect would also be present in the Walrasian model and is thus fully 'classical', as was pointed out by Dixon (1987).

We should at this point also mention that, whereas the 'crowding-out' result (41) is fairly robust, the output expansion result (42) is much more fragile, and depends in particular very much on the method of taxation, as was shown notably by Molana and Moutos (1992). Indeed let us assume, using the same model as on p.22, that taxes are not levied in a lump sum fashion, but proportionally to all incomes (profits or wages). In that case it is easy to compute that (31) becomes:

$$F'(\ell)(\ell_0 - \ell) = \frac{\varepsilon}{\varepsilon - 1} \cdot \frac{\eta}{\eta - 1} \beta F(\ell) \tag{43}$$

all other equations remaining the same. In such a case employment and output are totally unaffected by the level of taxes and government spending, and there is 100 per cent crowding-out. The reason is intuitively simple: while the income effect of taxes still continues to induce a higher amount of work, inversely the proportional taxation of labour income discourages work. In this particular instance, the two effects cancel exactly.

Normative rules for government policy

We have just seen that in general fiscal policy can be effective in changing employment, output and private consumption, in a way somewhat similar to what would occur in a Walrasian setting. So a question one is naturally led to ask is: what should be the normative rules for government fiscal policy? Should they mimic the rules which would be derived in a comparable Walrasian model, or should they be 'biased' in a Keynesian manner, say by increasing government spending or reducing taxes? We shall now study this problem, beginning with the derivation, as a benchmark, of the 'classical' prescriptions.

Classical normative policy

The 'classical' policy prescription is most easily obtained by computing the 'stationary first best' state of our economy. This will be obtained through maximization of the representative consumer's utility subject to the global feasibility constraint, i.e.:

$$U(c, c', \ell_0 - \ell, g) \quad \text{s.t.}$$
$$c + c' + g = F(\ell)$$

which yields the conditions:

$$\frac{\partial U}{\partial c}=\frac{\partial U}{\partial c'}=\frac{\partial U}{\partial g}=\frac{1}{F'(\ell)}\frac{\partial U}{\partial(\ell_0-\ell)}. \tag{44}$$

It is easy to verify that this first best solution can be obtained as a stationary Walrasian equilibrium, corresponding to (22)–(28) taking both $1/\varepsilon$ and $1/\eta$ equal to zero, provided the government adopts the following rules:

$$g=\tau \tag{45}$$

$$\frac{\partial U}{\partial g}=\frac{\partial U}{\partial c}. \tag{46}$$

(45) simply tells us that the government's budget should be balanced. (46) tells us that the government should push public spending to the point where its marginal utility is equal to that of private consumption. In other words the government should act as a 'veil' and pick exactly the level of g the household would have chosen if it was not taxed and could purchase directly government goods.

Normative policy under imperfect competition

We shall now derive the optimal rule for the government under imperfect competition. In order to simplify analysis, we shall study only the balanced budget case $g=\tau$.[3] In that case, prices are constant in time and equations (22)–(28) simplify to:

$$\frac{w}{p}=\left(1-\frac{1}{\eta}\right)F'(\ell) \tag{47}$$

$$\frac{\partial U}{\partial c}=\lambda p \qquad \frac{\partial U}{\partial c'}=\lambda p \tag{48}$$

$$\frac{\partial U}{\partial(\ell_0-\ell)}=\lambda w\left(1-\frac{1}{\varepsilon}\right) \tag{49}$$

$$c+c'+g=y=F(\ell). \tag{50}$$

All equilibrium values depend on the level of g chosen by the government. To find its optimal value, let us differentiate $U(c,c',\ell_0-\ell,g)$ with respect to g:

$$\frac{\partial U}{\partial c}\cdot\frac{\partial c}{\partial g}+\frac{\partial U}{\partial c'}\cdot\frac{\partial c'}{\partial g}+\frac{\partial U}{\partial \ell}\cdot\frac{\partial \ell}{\partial g}+\frac{\partial U}{\partial g}=0. \tag{51}$$

Differentiating also (50) with respect to g we obtain:

$$\frac{\partial c}{\partial g}+\frac{\partial c'}{\partial g}+1=F'(\ell)\frac{\partial \ell}{\partial g}. \tag{52}$$

Combining (47), (48), (49), (51) and (52) we finally obtain:

$$\frac{\partial U}{\partial g}=\frac{\partial U}{\partial c}\left[1-\left(\frac{\varepsilon+\eta-1}{\varepsilon\eta}\right)F'(\ell)\frac{\partial \ell}{\partial g}\right]. \tag{53}$$

We see that there will be a systematic bias with respect to the first best rule (46): if $\partial\ell/\partial g>0$, as soon as there is market power (that is, if either ε or η is not infinite), the government will be led to push its spending beyond the level that the consumer would freely choose. The converse result will hold if $\partial\ell/\partial g<0$.

Another way to view this result is to imagine that we start from the level of g that the consumer would have freely chosen. That level of g is characterized by adding the following equation to (47)–(50) describing the imperfectly competitive equilibrium:

$$\frac{\partial U}{\partial g}=\frac{\partial U}{\partial c}=\lambda p. \tag{54}$$

Let us consider now, starting from this level, a small increase in public spending dg, financed by supplementary taxes $d\tau=dg$, and let us compute the resulting utility increase:

$$dU=\left[\frac{\partial U}{\partial c}\cdot\frac{\partial c}{\partial g}+\frac{\partial U}{\partial c'}\cdot\frac{\partial c'}{\partial g}+\frac{\partial U}{\partial \ell}\cdot\frac{\partial \ell}{\partial g}+\frac{\partial U}{\partial g}\right]dg. \tag{55}$$

Using (47), (48), (49), (52) and (54), we obtain:

$$dU=\frac{\partial U}{\partial c}\left[\left(\frac{\varepsilon+\eta-1}{\varepsilon\eta}\right)F'(\ell)\frac{\partial \ell}{\partial g}\right]dg. \tag{56}$$

This shows that, as compared with the first best rule, the government should systematically bias its spending so as to increase the level of economic activity. The intuition is straightforward: because of imperfect competition the level of activity on the goods and labour markets is inefficiently low, as we saw before. When choosing its level of spending, the government should take into account not only the direct effect on the household's utility (which would yield the 'first best' rule $\partial U/\partial g=\partial U/\partial c$), but should also take into account the indirect utility gains which derive from the positive effect of its macroeconomic policy on activity. This 'second best' policy prescription is thus different from the 'first best' classical one.

Should one, however, believe that the normative policy is biased in a 'Keynesian' manner? This is not the case, at least for two reasons. First, even when $\partial \ell / \partial g$ is positive, what leads to the activity increase is not government spending *per se* via a 'Keynesian' demand multiplier, but rather the taxes levied to finance them via a 'classical' labour supply effect. Normative analysis would then somehow call for higher taxes, hardly a Keynesian prescription. Secondly, the magnitude and even the sign of $\partial \ell / \partial g$ depends enormously on the method of taxation, making the direction of the bias extremely difficult to assess. Using again the example on pp.22 and 26, under proportional taxation the government should use exactly the 'classical' prescription. So whatever bias exists in the normative prescriptions, it is definitely not of a Keynesian type.

Conclusions

We constructed in this chapter a simple prototype model of imperfect competition with rational expectations and objective demand curves, studied its various properties, and compared them with those of the basic 'classical' and 'Keynesian' models.

We may first note that this model of imperfect competition clearly generalizes the corresponding Walrasian one, which can be obtained as a limit case by making the parameters η and ε go to infinity.

As for the 'positive' properties of the model, we saw that they stand somehow halfway between the Keynesian and classical ones: the inefficiency properties very much resemble those of a Keynesian fix-price–fix-wage model. On the other hand, the response to government policy, fiscal or monetary, is very much of a 'classical' nature.

The normative implications of such models for government action are also very important, and we saw that they were neither Keynesian nor classical. Moreover simple variations on the above model show that they will depend crucially on the nature of rigidities in the price system. It is thus quite urgent to develop models with more sophisticated rigidities than those arising from simple market power and to explore their positive and normative properties. This should be the object of further research.

Appendix

We shall in this appendix derive, under a more general form, the objective demand curves used in the text (cf. notably (16) and (17)), and show how all the results extend without modification to general aggregator functions Λ and V.

The objective demand curves

When computing the objective demand curve for the product he sells, each price-maker has to forecast the demand forthcoming to him for any value of (1) the price or wage he determines and (2) prices and wages set by other agents. Following the methodology developed in Bénassy (1988, 1990), we see that the natural definition of objective demand at a price–wage vector (P, W) is simply the demand forthcoming at a fix-price equilibrium corresponding to (P, W), which we shall now compute.

We may note before actually starting computations that, according to a traditional result in imperfect competition, each agent will set the price of the good he controls at a level high enough for him to be willing to serve all demand forthcoming, and actually even more. We are thus, in 'fix-price' terminology, in a situation of generalized excess supply where each agent is constrained in his supply (but unconstrained in his demand) and thus takes the level of his sales as a constraint.

Consider first firm j. For given prices and wages its optimization program is:

$$\max p_j y_j - \sum_{i=1}^{n} w_i \ell_{ij} \quad \text{s.t.}$$

$$F[\Lambda(\ell_{1j}, \ldots, \ell_{nj})] = y_j$$

where y_j is determined by the demand of other agents and thus exogenous to firm j. The solution in ℓ_{ij} to this program is:

$$\ell_{ij} = \phi_i(W) F^{-1}(y_j) \tag{A1}$$

where $\phi_i(W)$, a function associated to Λ by duality, is homogeneous of degree zero in its arguments. As an example, if Λ is the CES function (3), then

$$\phi_i(W) = \frac{1}{n}\left(\frac{w_i}{w}\right)^{-\varepsilon} \tag{A2}$$

where w is the aggregate wage index given by (4) in the text.

Consider now old household i. It owns a quantity of money $\bar{m}_i$ and seeks to maximize its second-period consumption index c_i' (8) under the budget constraint:

$$\sum_{j=1}^{n} p_j c_{ij}' = \bar{m}_i.$$

The result of this maximization is:

$$c'_{ij}=\phi_j(P)\frac{\bar{m}_i}{p} \tag{A3}$$

where $\phi_j(P)$, associated by duality to V, is homogeneous of degree zero in all prices, and p is the aggregate price index associated to V, given by:

$$p=\sum_{j=1}^{n}p_j\phi_j(P). \tag{A4}$$

As an example again, if V is the CES function (10), then:

$$\phi_j(P)=\frac{1}{n}\left(\frac{p_j}{p}\right)^{-\eta}. \tag{A5}$$

Consider now the government and assume it has chosen a level g_i for the level of public consumption index attributed to household i. The government will choose the specific g_{ij}s to minimize the cost of doing so, and will thus solve the program:

$$\min \sum_{j=1}^{n}p_jg_{ij} \quad \text{s.t.}$$

$$V(g_{i1},\ldots,g_{in})=g_i$$

which yields the solution in g_{ij}:

$$g_{ij}=\phi_j(P)g_i \tag{A6}$$

where $\phi_j(P)$ is the same as in (A3). The cost to the government is pg_i.

Let us finally consider young household i. Merging its two budget constraints (12) and (13) into a single one, we find that it will determine its current consumptions c_{ij} through the following maximization program:

$$\max U(c_i,c'_i,\ell_0-\ell_i,g_i) \quad \text{s.t.}$$

$$\sum_{j=1}^{n}p_jc_{ij}+\sum_{j=1}^{n}p'_jc'_{ij}=w_i\ell_i+\pi_i-p\tau_i$$

where the right-hand side (and notably the quantity ℓ_i of labour sold) is exogenous to household i. Given the assumptions on U (separability, homotheticity), the solution will be such that the value of current consumptions is given by:

$$\sum_{j=1}^{n}p_jc_{ij}=\gamma(p'/p)(w_i\ell_i+\pi_i-p\tau_i) \tag{A7}$$

where $\gamma(p'/p)$ is the propensity to consume. Maximizing c_i under budget constraint (A7) yields the current consumptions c_{ij}:

$$c_{ij} = \phi_j(P)\gamma(p'/p)(w_i\ell_i + \pi_i - p\tau_i). \tag{A8}$$

We have now determined all components of the demand for goods. Output y_j will be equal to the sum of demands for good j, i.e.:

$$y_j = \sum_{i=1}^{n} c_{ij} + \sum_{i=1}^{n} c'_{ij} + \sum_{i=1}^{n} g_{ij} \tag{A9}$$

which, using (A3), (A6) and (A7) yields:

$$y_j = \phi_j(P)\left[\frac{\bar{M}}{p} + G + \gamma(p'/p)\sum_{i=1}^{n}(w_i\ell_i + \pi_i)/p - \gamma(p'/p)\theta\right]$$

$$G = \sum_{i=1}^{n} g_i \qquad \bar{M} = \sum_{i=1}^{n} \bar{m}_i \qquad \theta = \sum_{i=1}^{n} \tau_i. \tag{A10}$$

We shall use the global incomes identity:

$$\sum_{i=1}^{n}(w_i\ell_i + \pi_i) = \sum_{j=1}^{n} p_j y_j. \tag{A11}$$

Combining (A4), (A10) and (A11) we obtain the final expression for the objective demand addressed to firm j:

$$Y_j = \phi_j(P)\frac{1}{1-\gamma}\left[\frac{\bar{M}}{P} + G - \gamma\theta\right]. \tag{A12}$$

If the number n of producers is large, p, p' and thus γ are taken as exogenous to firm j and the elasticity of Y_j with respect to p_j is that of the function ϕ_j.

We can now compute the objective demand for type i labour by adding the ℓ_{ij}s, $j = 1, \ldots, n$ given by (A1) and replacing y_j by the objective demand Y_j just derived, which yields:

$$L_i = \phi_i(W)\sum_{j=1}^{n} F^{-1}(Y_j) \tag{A13}$$

where the Y_j are given by (A12). Again with large n, the elasticity of L_i with respect to w_i is equal to that of $\phi_i(W)$.

Now formulas (16) and (17) in the text are simply obtained by replacing $\phi_i(W)$ and $\phi_j(P)$ by the specific forms (A2) and (A5), and using the fact that the values of $\bar{m}_i$, g_i and τ_i are the same for all n households.

General aggregator functions

We shall now show that all results derived in the text with the specific CES aggregator functions (3) and (10) are valid as well with general forms for Λ and V, and notably that the crucial (22) and (25) hold unchanged.

Indeed the first order conditions for programs (A_j) and (A_i) are in the general case:

$$\frac{w_i}{p_j}=\left(1-\frac{1}{\eta_j}\right)\frac{\partial F_j}{\partial \ell_{ij}} \tag{A14}$$

$$\frac{\partial U_i}{\partial(\ell_0-\ell_i)}=\lambda_i w_i\left(1-\frac{1}{\varepsilon_i}\right) \tag{A15}$$

where η_j and ε_i are the absolute values of the elasticities of the functions Y_i and L_i. Looking at formulas (A12) and (A13), we see that for large n these elasticities are actually those of the functions ϕ_j and ϕ_i, so that:

$$\eta_j=-\partial \log \phi_j(P)/\partial \log p_j=\eta_j(P) \tag{A16}$$

$$\varepsilon_i=-\partial \log \phi_i(W)/\partial \log w_i=\varepsilon_i(W) \tag{A17}$$

Because of the homogeneity and symmetry properties of the original functions Λ and V these elasticities are the same at all symmetric points, and we denote them as η and ε:

$$\eta_j(p,\ldots,p)=\eta \qquad \forall p,\forall j \tag{A18}$$

$$\varepsilon_i(w,\ldots,w)=\varepsilon \qquad \forall w,\forall i. \tag{A19}$$

Combining (A14), (A15), (A18) and (A19) at a symmetric equilibrium, we obtain (22) and (25).

Notes

I wish to thank Huw Dixon and Neil Rankin for their comments on a first version of this chapter. The usual disclaimer applies fully here.

1. Of course all the concepts that follow would be valid with a different number of households and firms, but using the same number n will simplify notation at a later stage.
2. These were initially introduced in the macrosetting by Weitzman (1985).
3. The case of an unbalanced budget $g \neq \tau$ is studied in Bénassy (1991b).

2 Imperfect competition and macroeconomics: a survey

Huw David Dixon and Neil Rankin

Introduction

The importance of imperfect competition has long been recognized in many areas of economics, perhaps most obviously in industrial economics and in the labour economics of trade unions. Despite the clear divergence of output and labour markets from the competitive paradigm in most countries, macroeconomics where it has used microfoundations has tended to stick to the Walrasian market-clearing approach. However, over the last decade a shift has begun away from a concentration on the Walrasian price-taker towards a world where firms, unions and governments may be strategic agents. This chapter takes stock of this burgeoning literature, focusing on the macroeconomic policy and welfare implications of imperfect competition, and contrasting them with those of Walrasian models.

We seek to answer three fundamental questions. (1) What is the nature of macroeconomic equilibrium with imperfect competition in output and labour markets? With monopoly power in the output market causing price to exceed marginal cost, and union power leading to the real wage exceeding the disutility of labour, we would expect imperfectly competitive macroeconomies to have lower levels of output and employment than Walrasian economies, a Pareto inefficient allocation of resources, and the possibility of involuntary unemployment. Few would disagree that deviations from perfect competition will probably have undesirable consequences. (2) To what extent can macroeconomic policy be used to raise output and employment in an imperfectly competitive macroeconomy? (3) If policy can raise output and employment, what will be the effect on the welfare of agents?

Whilst there may be fairly general agreement over the answer to (1), we believe that there are no truly general answers to (2) and (3). In Walrasian models there is only one basic equilibrium concept employed: prices adjust to equate demand and supply in each market. There are, however, many different types of imperfect competition, which can differ in fundamental respects, as has been seen in Industrial Organization and the 'New' International Trade Theory. We would thus expect the theory of imperfectly competitive macroeconomies to embrace 'classical' models, with monetary neutrality and a vertical aggregate supply curve, as well as 'Keynesian' models. Imperfect competition, however, not only opens new channels of influence for monetary and fiscal policy, but also creates the possibility that an increase in output may be welfare improving. The First Fundamental Theorem of Welfare Economics tells us that the Walrasian equilibrium is Pareto optimal. But with imperfect competition, the market prices of goods and labour generally exceed their shadow prices, so policies that succeed in expanding output are very likely to increase welfare. The survey considers several cases of such policy effects, which are in stark contrast to those in Walrasian economies.

In the second section we present a general framework which nests much of the theoretical literature on imperfectly competitive macroeconomies, and enables us to explore the effects of imperfect competition on output and labour markets. In the third and fourth sections we explore monetary and fiscal policy respectively, concentrating on the mechanisms through which policy effects occur in an imperfectly competitive economy. We inevitably have been forced to omit several closely related areas of potential interest, amongst which are the 'mesoeconomic' approach developed by Ng(1980, 1982a, 1986); open economy applications (these are surveyed in Dixon, 1994a); macroeconomic models of bargaining (McDonald and Solow, 1981; Jacobsen and Schultz, 1990); and the 'insider–outsider' literature (Lindbeck and Snower, 1989). We have also omitted the 'coordination failure' literature (see Silvestre, 1993 and an earlier version of this chapter, Dixon and Rankin, 1991).

A general framework

The models constructed in much of the recent literature on imperfect competition have shared some common features. In this section we outline a generic model of an imperfectly competitive economy that provides a baseline, and we will use variants of it in subsequent sections to derive particular results. The three main points we make in this section are: (1) that imperfect competition in either output or labour markets will tend to reduce equilibrium output and employment; (2) that the introduction of union wage-setting will tend to generate 'involuntary' unemployment; and (3) that the model will possess a (unique in this case) Natural Rate, with money being neutral. This section thus highlights the 'classical' properties of imperfect competition, as a prelude to subsequent sections which will extend the framework to models with less classical effects for monetary and fiscal policy.

There are n produced outputs X_i, $i=1,\ldots,n$. Households' utility function takes the form

$$[u(X_1,\ldots,X_n)]^c[M/P]^{1-c}-\theta N^e \qquad 0<c<1, e\geq 1 \tag{1}$$

where u is a degree-one homogeneous subutility function, P is the cost-of-living index for u, M is nominal money holdings, and θN^e is the disutility of supplying N units of labour, $N\leq H$. Since preferences are homothetic over consumption and real balances we can aggregate and deal with one 'representative' household. Most studies further simplify (1). First, a specific functional form is assumed for $u(\cdot)$ – notably Cobb–Douglas or CES. Second, the labour supply decision may be made a [0,1] decision – to work or not to work for each individual household. We can represent this for our aggregated household by setting $e=1$, so that θ is the disutility of work. Most models of imperfect competition incorporate money using the standard temporary equilibrium framework (see Grandmont, 1983, for an exposition) by including end-of-period balances in the household's utility function. Whether it should be deflated by the current price level as in (1) depends on the elasticity of price expectations, as we discuss on p.47 below. As regards firms, we assume there are F firms in sector i, each with a loglinear technology

$$X_{if}=B^{-1}N_{if}^{a} \qquad a\leq 1. \tag{2}$$

The special case of constant returns where $a=1$ is widely used.[1]

We have now to add the macroeconomic framework. Turning first to aggregate income–expenditure identities, income in each sector must equal expenditure Y_i on that sector, and national income must equal total expenditure Y. We will introduce fiscal policy in the fourth section. In this section, the government merely chooses the total money supply M_0. In aggregate, the household's total budget consists of the flow component Y and the stock of money M_0. Given (1), households will choose to spend a proportion c of this on producer output, and to save a proportion $1-c$ to accumulate money balances M.[2] Hence the income–expenditure identities coupled with (1) imply that in aggregate:

$$Y=c[M_0+Y] \quad \text{or} \quad Y=\frac{c}{1-c}M_0. \tag{3}$$

Given total expenditure, households allocate expenditure across the produced goods. Since preferences are homothetic, the budget share of output

i, α_i, depends only on relative prices. Hence total expenditure on sector i, Y_i, is given by

$$P_iX_i \equiv Y_i = \alpha_i(P_1, \dots, P_n)Y \qquad (4)$$

where α_i is homogeneous of degree zero in $P_1, \dots, P_n$. We will assume symmetric preferences, so that if all prices of outputs are the same then $\alpha_i = 1/n$.

How are wages and prices determined? As a benchmark let us consider the Walrasian economy with price-taking households and firms. Furthermore, let us assume perfect mobility of labour across sectors, so that there is a single economy-wide market and wage W. The labour supply from (1) is then

$$N^s(W/P) = [\theta e]^{1/[1-e]}[W/P]^{1/[e-1]}. \qquad (5)$$

The additive separability combined with degree-one homogeneity of (1) rules out any wealth effect on labour supply, so that only real wages matter. Assume a single, representative, price-taking firm per sector.[3] Then sector i's labour demand takes the form

$$N^d(W/P_i) = [a/B]^{1/[1-a]}[W/P_i]^{-1/[1-a]}. \qquad (6)$$

In a symmetric equilibrium where $P_i = P$, (5)–(6) determine equilibrium real wages, employment and output in the representative sector. Equilibrium sector output X^* is given by $B^{-1}[N^*]^a$. Under symmetry $\alpha_i = 1/n$ in (4) so that the nominal price level is

$$P^* = \frac{c}{1-c}\frac{M_0}{nX^*}. \qquad (7)$$

Nominal wages and prices adjust to equate aggregate demand with equilibrium output. This is an entirely 'classical' model with full employment and neutral money.

What difference does the introduction of imperfect competition make? Let us assume that each output is monopolized by a sole producer ($F = 1$) and that there are many sectors. The large 'n' means that the monopolist treats the general price index P as exogenous when it makes its decisions. Before we proceed, it should be noted that the elasticity of demand $\varepsilon_i(P_1, \dots, P_n) \equiv -[\partial \ln X_i/\partial \ln P_i]_{P\,\mathrm{const}}$ from (4) is homogeneous of degree zero in prices, due to homotheticity. In a symmetric equilibrium $\varepsilon_i(P_1, \dots, P_n)$ will thus take the same value irrespective of the price level: $\varepsilon^* \equiv (1, \dots, 1)$. We will assume gross substitutability in general, so that $\varepsilon^* > 1$. If the individual firm maximizes its profits treating the general price level as given, then its labour demand is easily obtained as

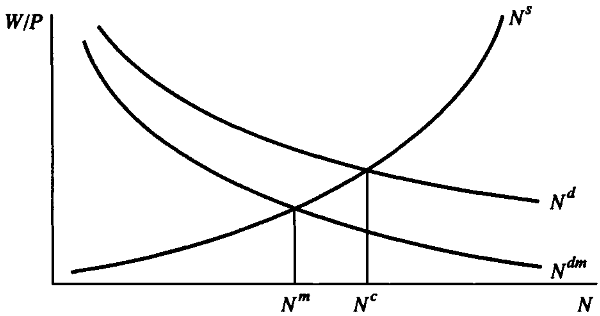


Figure 2.1 Equilibrium with and without monopolistic firms

$$N^{dm}(W/P_i) = [1 - 1/\varepsilon^*]^{1/[1-a]} N^d(W/P_i). \tag{8}$$

Since $\varepsilon^* > 1$ and $a < 1$, labour demand is smaller for any given real wage, as we would expect. Equilibrium under symmetry is depicted in figure 2.1. For a given supply curve, in a symmetric equilibrium the effect of monopolistic competition is simply to reduce sectoral employment (and hence output) from N^c to N^m. Note that money will still be neutral, since (5) and (8) are both homogeneous of degree zero in (W, P). The degree of monopoly μ is $1/\varepsilon^*$. So the less elastic is demand when prices are all equal, the higher the marking up of price over marginal cost and the lower equilibrium output. Imperfect competition in the output market has thus reduced total output and employment, although (since the labour market is competitive) households are on their labour supply curve N^s.

How will the introduction of unions alter matters? To take the simplest case, consider an economy-wide monopoly union that has the unilateral power to set the nominal wage. The union predicts, given the wage it has set, what prices will be set by firms and the resultant level of employment.[4] At the aggregate level, the trade-off between real wages and employment faced by unions in symmetric equilibrium is given by (8) multiplied by the number of sectors n. Several assumptions may be made about union preferences (see Oswald, 1985). Here we will simply assume that the union's objective function is to maximize the total surplus, or wage revenue less disutility, earned by employed workers.[5] If we let $e = 1$, then there is a constant marginal disutility of labour θ. Each employed worker earns $W/P - \theta$ as surplus. The union's problem is thus

$$\max_{W/P} [W/P - \theta] N^{dm}(W/P). \tag{9}$$

Since the elasticity of labour demand with respect to W/P is constant in (8),

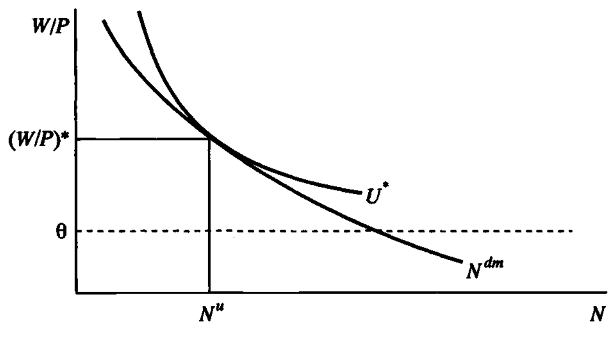


Figure 2.2 Equilibrium with an economy-wide union

the solution to (9) has the property that the union chooses the real wage as a constant mark-up over θ

$$W/P=\theta/a. \qquad (10)$$

This is depicted in figure 2.2, where we show the union's maximum utility indifference curve U^*. The less elastic is the demand for labour (the lower a), the higher is the wage set by the union. Since the monopolistic and competitive firms have the same real wage elasticity of labour demand, the real wage chosen by the union is the same, though employment is lower with monopolistic firms. Again, since (10) is homogeneous of degree zero in (W, P), money is neutral. This model illustrates the point that the introduction of wage-setting unions leads to involuntary unemployment. Since the union marks up the wage over the disutility of labour (from (10)), the unemployed households are worse off than the employed, and furthermore the employed would be willing to work more for less.

Turning to the case of sectoral unions, wages in each sector may now differ. The union is assumed to control entry into employment in that industry so that employed 'insiders' are isolated from the potential competition of 'outsiders' (see Lindbeck and Snower, 1989, for a discussion). In a large economy with many sectors, each individual union will take the general price level of goods consumed by its members as given (in contrast to the centralized union) and in a non-cooperative Nash framework it will also treat other sectoral unions' wages as given. However, each union will take into account the effects of its wages on its own sector's price P_i and hence on output and demand for labour in its sector. The sectoral labour demand curve is thus essentially a relation between nominal wages W_i and employment, because the firm bases its employment on the own-product wage W_i/P_i, in contrast to the union's objective function which depends on the real consumption wage W_i/P.[6]

Consider the elasticity of the sectoral unions' demand for labour with respect to W_i. The labour-demand equation stems from the price-cost equation which equates the own-product wage to the marginal revenue product of labour

$$\frac{W_i}{P_i}=\frac{\varepsilon-1}{\varepsilon}aB^{-1}N_i^{-[1-a]}. \tag{11a}$$

If we take logs and differentiate N_i with respect to W_i, taking into account that P_i depends upon X_i which depends upon N_i, we obtain the money-wage elasticity of labour demand as

$$-\frac{\mathrm{d}\ln N_i}{\mathrm{d}\ln W_i}=\frac{1}{1-a+a[1-\eta]/\varepsilon} \tag{11b}$$

where η is the elasticity of $[\varepsilon-1]/\varepsilon$ with respect to P_i, which captures the effect of a rise in W_i (and hence P_i) on the mark-up of price over marginal cost, $\varepsilon/[\varepsilon-1]$. This can take either sign, although it is perhaps more reasonable to assume that $\partial\varepsilon/\partial P_i>0$ (demand becomes more elastic as you raise price), so that $\eta>0$ (the mark-up falls as W_i rises). If the sectoral union maximizes its surplus with respect to W_i, subject to the labour demand implicitly defined by (11), the equilibrium real wage given symmetry across sectors is

$$\frac{W}{P}=\frac{\theta}{a}\frac{\varepsilon^*}{\varepsilon^*-1+\eta}. \tag{12}$$

This is a higher real wage than with the centralized union (cf. (10)) so long as $\eta<1$. If utility is CES then $\varepsilon(P_1,\ldots,P_n)$ is equal to the constant elasticity of substitution, so $\eta=0$ and the comparison of (12) with (10) is unambiguous.

The sectoral union thus tends to set a higher nominal wage, with a consequent lower level of employment, despite the fact that the money-wage elasticity of its labour demand is likely to be higher than for a centralized union.[7] The reason is that it sees no effect of its own behaviour on the general price level P at which its members consume. A centralized union takes the general rise in P which it causes into account, and so restrains its wage pressure. This can also be seen as an example of externalities: the price rise caused by a sectoral union is mostly borne by members of other unions. For a detailed discussion of the effect of different union structures on wage determination, see Calmfors and Driffill (1988).

An alternative to sectoral unions – which are organized by industry – are craft unions, which are organized by labour skills. Blanchard and Kiyotaki (1987) use these. Each union sells a different type of labour, some of which is required by every firm. The number of labour types is assumed large, so each W_h has a negligible effect on the general index of wages W and no

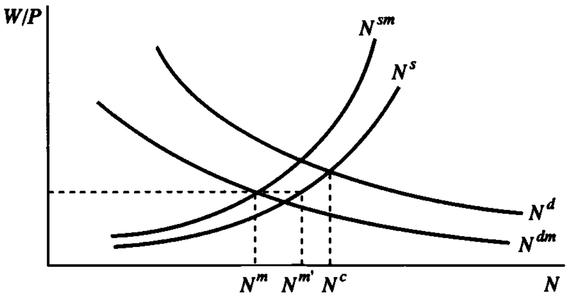


Figure 2.3 Equilibrium with craft unions and monopolistic firms

union sees itself as affecting any firm's output. Firm i's demand for type h labour, which is obtained by minimizing the cost of producing a given output X_i, is

$$N_{ih} = k_n[W_h/W]^{-\sigma} X_i^{1/a} \qquad k_n = \text{const} \tag{13}$$

where $\sigma(>1)$ is the constant elasticity of technical substitution between labour types. Blanchard and Kiyotaki further assume increasing marginal disutility of work, i.e. $e>1$ in (1), so that the union's surplus is[8]

$$[W_h/P]N_h - \theta N_h^e. \tag{14}$$

Maximizing (14) subject to (13) (aggregated over all i) taking W and $X_i, \ldots, X_n$ as given, union h chooses the level of labour sales

$$N^{sm} = [1 - 1/\sigma]^{1/[e-1]} N^s(W_h/P). \tag{15}$$

Since $\sigma > 1$, labour sales are smaller for any given real wage than in the competitive case (5), as we would expect. Combining (15) with (8) determines equilibrium under symmetry amongst firms and unions, as in figure 2.3. Whether the employment level is higher or lower than with sectoral unions depends on technological substitutability between labour types. With $\eta = 0$ and $e = 1$, employment is lower if $\sigma < 1/[1 - a[1 - 1/\varepsilon]]$. Money is clearly still neutral.

In this section we have presented a simple general framework that captures some common features of much of the recent work on imperfect competition and macroeconomics. We will now proceed to see how extensions of this general framework yield less orthodox results.

Monetary policy

Imperfect competition by itself does not create monetary non-neutrality, as we have seen.[9] It is the combination of imperfect competition with some other distortion[10] which generates the potential for real effects. The nature of this other distortion provides us with a natural method for classifying models of monetary policy effectiveness. First, the largest part of the literature combines imperfect competition with small lump sum costs of adjusting prices ('menu' costs), which are intended to represent the administrative costs of printing new price lists, etc. Examples of this approach are Mankiw (1985), Akerlof and Yellen (1985a), Blanchard and Kiyotaki (1987), Bénassy (1987), Caplin and Spulber (1987), Ball and Romer (1989a, 1990, 1991). A second group of studies may be interpreted as taking the same general framework of imperfect competition in a monetary temporary equilibrium, but as relaxing an implicit assumption often unconsciously made there: namely, that of unit-elastic expectations of future with respect to current prices. These include the seminal study of Hart (1982) and applications and extensions by Dehez (1985), d'Aspremont *et al.* (1989a, 1990), Silvestre (1990) and Rankin (1992, 1993). Thirdly, studies by Dixon (1990b, 1992), Fender and Yip (1993), and Moutos (1991) look at the imperfect competition combined with a small nominal rigidity in some sector of the economy. Common to all three approaches is that the same distortions in the presence of perfect competition would not cause monetary policy to affect output significantly. It is the interaction between minor, and perhaps intrinsically uninteresting, distortions and imperfect competition which generates major departures from neutrality. This can be viewed as an instance of the theory of the second best at work: monetary policy is not capable of causing Pareto improvements given either imperfect competition or the other distortion on its own, but given both together, it is.

Menu costs

We take Blanchard and Kiyotaki's (1987) model for our illustration, though the central ideas appear first in Mankiw (1985) and Akerlof and Yellen (1985a). Very similar points were simultaneously made by Bénassy (1987), and Parkin's (1986) study also has strong parallels. That this approach has also been influential outside the realm of pure theory is shown by Layard and Nickell's adoption of the Blanchard–Kiyotaki framework for their widely known empirical studies of UK unemployment (Layard and Nickell, 1985, 1986; Layard, Nickell and Jackman, 1991). The model in the absence of menu costs has already been described. Suppose now that the price- and wage-setting agents face administrative costs of changing prices and wages (e.g. for a restaurant, the cost of reprinting its menus). Such costs are lump sum in nature: they do not depend on the size of the price or wage change. If they are large enough to outweigh the forgone profits or utility of not adjusting a price or a wage when an increase in the money supply

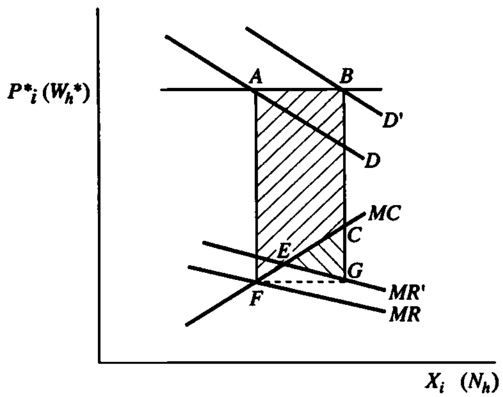


Figure 2.4 Effects of a demand shift on profits or utility

occurs, the firm or union still has to decide whether to meet the increase in demand, or whether to ration it. This is where monopoly is important: since price exceeds marginal cost and wage exceeds marginal disutility in the initial equilibrium, firms and unions will prefer to satisfy the extra demand (up to a point), since a profit or surplus is made on every extra unit sold. This is illustrated in figure 2.4, where the trapezium *ABCF* indicates the increase in the firm's profits or the union's utility. By contrast, under perfect competition, the price (wage) equals marginal cost (disutility) initially, and the firm (union) would lose profits (utility) if it satisfied an increase in demand, and so would choose to ration its customers.[11]

Once general equilibrium spillover effects have been taken into account, the size of the horizontal shift in the goods demand curves, and hence the size of the increase in output, will be in percentage terms equal to the increase in the money supply. This may be seen from the macroeconomic aggregate demand function (3) whence, together with the goods demand function (4) and the labour demand function (13), we may derive the multipliers

$$\frac{\mathrm{d}\ln X}{\mathrm{d}\ln M_0}=1 \qquad \frac{\mathrm{d}\ln N}{\mathrm{d}\ln M_0}=1/a. \tag{16}$$

Although figure 2.4 is a partial equilibrium diagram in which the position of the marginal cost (disutility) curve depends on the general wage (price) index, these two indices may validly be assumed unchanged provided menu costs are binding for all agents. Hence no shift in the curve is needed to depict the new general equilibrium.

The limit of the possible increase in employment and output (always given large enough menu costs) is reached when demand equals the competitive supply in either market. Beyond this, even monopolistic agents will choose to ration any further increase. Whether the limit is first hit in the goods or labour market depends whether the real wage is above or below its Walrasian level (respectively). In figure 2.3, for example, the maximum employment level as M_0 is increased occurs at $N^{m\prime}$, where the supply constraint in the labour market becomes binding. This brings out the formal similarity between the menu cost models and the 'disequilibrium' models of the 1970s (Barro and Grossman, 1971; Bénassy, 1975; Malinvaud, 1977 – see Bénassy, 1990 for a more recent survey). Within the class of equilibria for which menu costs bind, the economy behaves exactly as if it were in a quantity-constrained equilibrium.[12] In particular, within the neighbourhood of the initial frictionless equilibrium, it behaves as if in a regime of generalized excess supply, or 'Keynesian unemployment'. Big increases in the money supply will shift it into a regime of 'repressed inflation' (as happens in figure 2.3) or 'classical unemployment', depending on whether labour or goods supply constraints are reached first. This similarity between excess supply and monopoly was first exploited by Bénassy (1976, 1978) and Negishi (1979) as a means to 'endogenize' prices in disequilibrium models. Their models, however, use the concept of 'subjective' demands, introduced by Negishi (1961), rather than the 'objective' demands used here.

Note that the increase in output also constitutes a Pareto improvement. This is for three reasons: first, the shift in demand for labour brings a utility gain to the household equal to the area *ABCF* in figure 2.4; second, households receive an increase in profits from firms; and third, real money balances increase, which increases households' utility directly. An interesting interpretation of the Pareto suboptimality of the initial monopolistic equilibrium is to view it as resulting from a lack of cooperation amongst price-setting agents. Bénassy (1987) and Blanchard and Kiyotaki (1987) both point out that an agreement by all firms and households simultaneously to lower their prices and wages by x per cent would produce exactly the same real reallocation as a money supply increase of x per cent in the presence of binding menu costs. In either case real balances, and thus real demand and output, rise by x per cent, with no relative price changes. Monetary policy can thus be seen as a substitute for a cooperative agreement to lower prices. The failure to lower prices when acting independently is explained by Blanchard and Kiyotaki as due to an 'aggregate demand externality':[13] a lowering of one agent's price benefits all others to the extent that it slightly reduces the general price index and so raises real money balances and aggregate demand.[14] However, the private gain to the price-cutter is outweighed by the private loss due to the too-low relative price which would result. The monopolistically competitive equilibrium is therefore a form of economy-wide 'Prisoner's Dilemma'. In the absence of menu costs, when to expand the money supply would have no beneficial effect, a welfare-enhancing measure would be to impose an all-round wage and price cut by a prices and incomes policy.

Before the menu cost model can be taken seriously, it must tackle the obvious objection that in practice administrative costs of price and wage adjustment are very small. Because such costs are lump sum, once they are dominated by the forgone profits or utility of not re-optimizing in the face of a money supply increase, they will have no effect at all: an agent who has decided to adjust her price will adjust it to the same level as in the absence of menu costs, since the cost depends on the fact of the adjustment and not on its size. A large part of the research into menu costs has been concerned with overcoming this objection. The key observation is provided by Akerlof and Yellen (1985a) and Mankiw (1985), who point out that the opportunity cost of non-adjustment is second order in the size of the money supply shock. That is, if we take a Taylor approximation to firm i's forgone profits of not increasing P_i, or to union h's forgone utility of not increasing W_h, it will contain no term in ΔM_0, only in $(\Delta M_0)^2$ and higher powers of ΔM_0. We explain why below. By comparison, the increase in output is first order, i.e. proportional to ΔM_0, as is clear from (16). Thus, as ΔM_0 tends to zero, the ratio of the size of menu cost necessary for non-adjustment to the change in output which it sustains also tends to zero. For non-infinitesimal but small changes in the money supply, it follows that only a very small menu cost will be required in order to sustain non-adjustment. For instance, Blanchard and Kiyotaki's calculations show that with a 5 per cent money supply increase and $\theta=\sigma=5$, $e=1.6$ and $a=0.8$, the minimum menu cost for households to prefer non-adjustment is equal to only 0.112 per cent of GDP, and for firms to only 0.018 per cent of GDP.

To see why the opportunity cost is second order, consider again figure 2.4. (We have deliberately magnified the diagram in the neighbourhood of equilibrium, which permits us to approximate curves as straight lines.) If the price were to be adjusted following the shift in demand, the new output would be determined by the intersection E. Since output under no adjustment is greater than this, the forgone profits (utility) of not adjusting are measured by the cumulated surplus of MC over MR', given by triangle CGE. As the size of ΔM_0, and thus of the demand shift AB, is squeezed towards zero, the area of this triangle clearly falls with the square of AB, and thus with the square of ΔM_0 (note that distance BG always equals AF since demand elasticity is a constant). Intuitively, the reason why the cost is only second order is that 'objective functions are flat on top': in the neighbourhood of a maximum the slope is close to zero, so that the loss from being only slightly away from the optimum is also very close to zero.

Several extensions of this basic analysis are made by Ball and Romer. In their (1989a) study, they show that if the money supply is stochastic then the welfare cost of price rigidity, as measured by the fall in expected utility, becomes second rather than first order, i.e. proportional to the variance of the money supply. Despite this, parameter values exist which will drive the ratio of second order menu costs to second order welfare losses to zero.[15] In their 1991 study, Ball and Romer show that for a given money supply increase, there exists an intermediate range of values for the menu cost such that two equilibria coexist: one with no and one with full price adjustment. For menu costs in this range it can thus be argued that if the outcome is no adjustment, price rigidity is due to a coordination failure: if each agent expected the others to adjust, she would want to adjust too. In yet a third study, Ball and Romer (1990) address the problem for the basic menu cost model that, although theoretically acceptable parameter values exist which keep rigidity-sustaining menu costs small, these values are still unrealistic empirically. The lowest value of competitive labour supply elasticity ($1/[e-1]$) used in Blanchard and Kiyotaki's numerical illustrations is $1\frac{2}{3}$ (as in the cited example), which is much higher than most econometric estimates. Ball and Romer suggest a solution by showing that rigidities in nominal prices are made more likely if there are also rigidities in real – or relative – prices. Their general argument is as follows. Suppose agent i has the indirect utility function

$$U_i = W(M_0/P, P_i/P). \tag{17}$$

Agents in the model are 'farmers'.[16] i's optimum relative price is clearly determined from the first order condition $W_2(M_0/P, P_i^*/P) = 0$ (the subscript denoting a partial derivative), whence

$$\partial(P_i^*/P)/\partial(M_0/P) = -\underset{(+)}{W_{12}}/\underset{(-)}{W_{22}} \qquad (\equiv \pi, \text{ say}). \tag{18}$$

If a change in aggregate demand causes only a small change in i's desired relative price, real rigidity is said to be 'high', so π is an inverse measure of it. The second order approximation to i's private utility cost of not adjusting P_i after a change ΔM_0, given that others do not adjust, is

$$PC \approx [-(W_{12})^2/2W_{22}]\,[\Delta M_0]^2 = \tfrac{1}{2}\pi W_{12}[\Delta M_0]^2. \tag{19}$$

Thus the smaller is π, i.e. the greater the real rigidity, the smaller is the menu cost needed to ensure that i does not adjust his price. Ball and Romer flesh out this simple framework with two explicit models of real rigidities: a 'customer market' model in which firms face kinked demand curves due to ignorance by their customers of prices elsewhere; and a model with an 'efficiency wage' in the labour market.[17] Hence by combining three 'distortions' – imperfect competition, menu costs and real rigidities – an empirically plausible model of non-neutrality can be obtained.[18]

Are the results modified in a dynamic setting? Caplin and Spulber (1987) and Caplin and Leahy (1991) examine this question. Rules for optimal price- and wage-setting over time when subject to menu costs take an '*Ss*' form: when the deviation of P_i from its no-menu cost optimum, P_i^*, which follows a stochastic process determined by the process for the money supply, hits a ceiling S or a floor s, an adjustment is made to P_i, bringing $P_i^* - P_i$ back to some 'return point'. Caplin and Spulber show that when the money supply process involves only non-negative shocks to the money stock, money – rather surprisingly – turns out to be neutral in the aggregate. However when the process is symmetric, allowing negative as well as positive shocks, shocks do affect aggregate output, as shown in Caplin and Leahy.[19] Although these models are dynamic, the firm's optimization problem is treated as static. Dixit (1991), in a partial equilibrium analysis, shows that with dynamic optimization menu costs as small as fourth order can sustain price rigidities. The considerable difficulties of aggregating across agents with different initial situations have, however, so far discouraged the rapid development of this branch of the literature.

Non-unit-elastic price expectations

Hart (1982) was the first to show that imperfect competition could generate policy effectiveness. The 'policy' he considers, however, is not strictly monetary policy at all: it is an increase in the stock of a 'non-produced' good. Although Hart is reluctant to interpret this as 'money', we can validly do so if we view Hart's framework as one of temporary monetary equilibrium in the manner of Patinkin (1965) and Grandmont (1983). In this case the key assumption necessary for monetary policy effectiveness is that – implicitly or explicitly – agents' expectations of future prices are 'non-unit-elastic' in current prices. We present here a simple version of Hart's model under this monetary reinterpretation. We also examine another question debated in the literature for which Hart's model has been the framework: whether imperfect competition in the goods market can cause unemployment even at a zero wage.

Relative to the second section we slightly modify the household's problem, to

$$\max u(X, M/P^e) \quad \text{subj to} \quad M_0 + Y = PX + M. \tag{20}$$

P^e is the subjective expectation of next period's price level, so M/P^e is expected future consumption (taking the simplest possible case, in which the household has no future income).[20] Our earlier utility function (1) is just the special case of this in which price expectations are unit-elastic in current prices: $P^e = k_1 P$. Hart's (1982) utility function is the special case of it in which expectations are zero-elastic: $P^e = k_2$. Here we posit an arbitrary expectations function, $P^e = \phi(P)$. If preferences are homothetic, the resulting consumption demand function then takes the form

$$X = \alpha(P/\phi(P))\,\frac{M_0 + Y}{P} \tag{21}$$

where $\alpha'(\cdot)$ could take either sign. From this we have the price elasticity

$$-\frac{\partial \ln X}{\partial \ln P} \equiv \varepsilon(P) = 1 - [1 - \varepsilon_\phi(P)]\varepsilon_\alpha(P/\phi(P)) \tag{22}$$

where ε_α, ε_ϕ are the elasticities of the functions $\alpha(\cdot)$, $\phi(\cdot)$.

Now assume a competitive goods market, and constant returns to labour such that $X = N$. Then $P = W$, and the demand for labour is just (21) with N replacing X and W replacing P. This is the function faced by the r unions in a typical local labour market.[21] With no utility of leisure and equal rationing of its members, the appropriate maximand for the typical union is just its money wage revenue. Cournot competition amongst unions then requires that at an interior (unemployment) solution, where each union supplies $(1/r)$th of the market, we have

$$\varepsilon(W) = 1/r. \tag{23}$$

This equation (if it has a solution) defines W, and thus P, independently of M_0. We thus have complete price rigidity, and consequently a standard Keynesian-type multiplier of money on output, as is easily shown by setting $Y = PX$ in (21) and solving for X.

What is the role of non-unit-elastic expectations? This may be seen from (22): with a unit elasticity ($\varepsilon_\phi = 1$), $\varepsilon(P)$ equals 1, a constant. Clearly, it is then the case that no value of W can satisfy (23). It would be easy to show that here the equilibrium must be a full employment one, if we were explicitly to take into account the upper bound on a union's labour sales imposed by its exogenous labour endowment. However at full employment, monetary policy cannot affect output. Full employment would also be the outcome, whatever the elasticity of expectations, under perfect competition in the labour market, which highlights how imperfect competition is essential for the policy effectiveness result.[22] Rankin (1995) shows that if we take a more general production technology, then given a sufficient degree of concavity, unemployment will result even with unit-elastic expectations. In this case any divergence of the expectations elasticity from unity, above or below, turns out to cause a positive effect of money on output. Thus, since the expectations elasticity is an arbitrary subjective parameter, money will almost always affect output positively. Such robustness substantially strengthens Hart's original findings.

However, are non-unit-elastic expectations 'irrational'? An overlapping-generations extension of the model in which this can be investigated is provided in Rankin (1995). The very concept of an expectations elasticity, of course, implies that an element of learning is involved in expectations formation: under the extremely strong assumption of fully 'forward-looking' expectations, money is – unsurprisingly – neutral. But if our criterion of rationality is the more moderate one that expectations should converge on the truth, then we can show that any expectations elasticity is consistent with this. Moreover, a unit elasticity does not guarantee a correct short-run forecast: this will only be the case for 'step' increases in the money supply.[23]

Dehez (1985), d'Aspremont *et al.* (1989a, 1990), Silvestre (1990) and Schultz (1992) all consider models similar to Hart's but in which imperfect competition is in the goods market, and a competitive labour market is assumed. This is in order to consider whether goods market imperfect competition can cause a situation in which labour demand is bounded above, such that even at a zero wage, demand falls short of the economy's labour endowment. Assuming no utility of leisure, the market-clearing wage is then zero. Although such unemployment is formally 'voluntary', the situation is clearly one in which an extreme degree of wage flexibility is required, and only the smallest degree of inflexibility would in practice cause true involuntary unemployment.

The necessary condition for this type of unemployment is that a firm's marginal revenue should turn negative at some finite output level. This may be illustrated as follows. Assuming Cournot competition amongst F identical firms, in equilibrium we must have '$MR = MC$', i.e.

$$P[1 - 1/F\varepsilon(P)] = W. \tag{24}$$

By relating P back to X/F (the output of a typical firm) through the demand function, we may plot MR in the usual way as in figure 2.5. Here we have drawn MR as cutting the horizontal axis at some output level X_0/F, which is assumed to be less than the full employment one H/F. Silvestre (1990) shows that this can happen if, for example, the expectations elasticity is zero and utility is CES with an elasticity of substitution less than $1/F$. In this case as W, and thus MC, is lowered towards zero, output and employment can go no higher than X_0/F: unemployment exists even at zero wage. A non-unit expectations elasticity is clearly necessary for this to occur: if $\varepsilon_\phi = 1$ then from (22) $\varepsilon(P) = 1$, so by (24) MR is positive for all P (and thus for all X).[24]

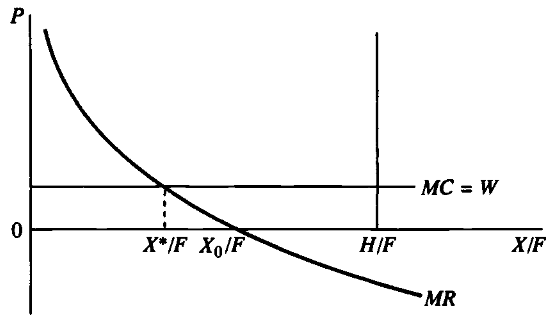


Figure 2.5 Negative marginal revenue at high output

Small nominal rigidities

There are in an economy many possible sources of nominal rigidity, which may occur in only a small part of the economy, but which may in the presence of imperfect competition cause significant non-neutrality of money. As is implied in the work of Ball and Romer (1990) and Haltiwanger and Waldman (1989), the relationship of strategic complementarity between the nominal choices of agents in general means that small nominal rigidities anywhere can induce some aggregate price rigidity (see also Dixon, 1994b). The origin of the nominal rigidity may be outside the domestic private sector – for example, for a small country with a fixed exchange rate, in the nominal price of tradeables: Dixon (1990a) and Rivera-Campos (1991) study this case. Prices, subsidies, welfare payments and taxes set by the government are also often 'rigid' in the sense of being set in nominal terms for a given period. One of the most significant of such nominal rigidities is unemployment benefit. Dixon (1990b), Fender and Yip (1993), and Moutos (1991) focus on this. The presence of such nominal rigidities can have very different implications in a unionized economy from those in a Walrasian economy. We will briefly look at the example of unemployment benefits.

Unemployment benefits are set in nominal terms by governments, and revised at regular intervals (in the UK, at the annual budget). In between revisions they are fixed. The level of unemployment benefits is important in a unionized economy because it alters the marginal trade-off between employment and unemployment for union members. If we take the baseline model and assume Cobb–Douglas preferences, no utility of leisure ($e=0$), constant returns to scale ($a=B=1$) and a perfectly competitive output market, then the market-clearing wage is (as depicted in figure 2.6)

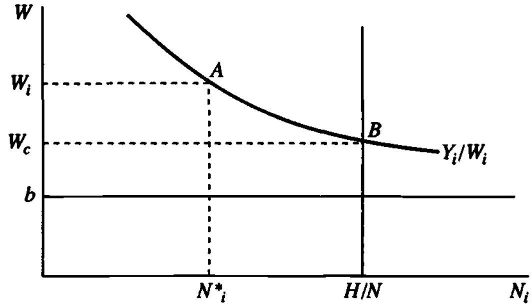


Figure 2.6 Equilibrium with the wage as a mark-up on unemployment benefits

$$W_c = \sum_{i=1}^{n} Y_i/nH = \frac{c}{1-c}\frac{M_0}{nH}. \tag{25}$$

The presence of unemployment benefits whose nominal level is fixed at b will not influence the level of wages so long as $b < W_c$, which means that it is worthwhile working (the replacement ratio is below unity). Except for the fact that benefits provide a floor for wages, money is neutral in the Walrasian case.

With unions, however, things may be different. Suppose that households are grouped into r unions in each sector who behave as Cournot quantity-setters. If union k seeks to maximize the 'surplus' $[W_i - b]N_{ik}$ earned by its members, treating the general price level P as fixed then, as is shown in Dixon (1990b), the equilibrium nominal wage becomes a mark-up over the benefit level

$$[W_i - b]/W_i = 1/r \tag{26}$$

so long as $W_i > W_c$. This is depicted in figure 2.6. The important point to note here is that in a unionized economy the nominal wage becomes tied to the benefit level. Furthermore even levels of benefit below the competitive wage can lead to involuntary unemployment, depending on the level of the money supply. So $W_i > W_c \geq b$, and employed households earn more than the unemployed, as at point A. Otherwise employment is always at B. This contrasts with the Walrasian economy, in which benefits can cause unemployment only if they are above the market-clearing wage. As a result of the nominal rigidity introduced by unemployment benefits, with a unionized labour market there will be standard Keynesian multiplier effects. Again, this contrasts with the Walrasian economy in which there will be full employment and a zero multiplier so long as $b \leq W_c$.

Fiscal policy

The same set of factors which make monetary policy effective on output will generally also make fiscal policy effective. This should not surprise us, since as just seen these factors work by endogenously producing some form of price stickiness, and we have long been familiar with the idea that price stickiness makes any policy that influences aggregate demand effective. However unlike in the case of monetary policy, imperfect competition by itself is in general enough to cause significant effects of fiscal policy on output. This is for several reasons. First, it is of the essence of price and wage determination in imperfectly competitive markets that elasticity of demand matters. Government policies which influence the elasticity of demand therefore have the potential to alter relative prices in a way that is absent in a price-taking economy. Second, imperfect competition influences the distribution of income between wages and profits. Where income distribution affects equilibrium, such as where there are income effects on labour supply, the degree of competition can alter the impact of government spending. A third reason is that in practice fiscal policy is not generally symmetric: governments tend to concentrate spending on particular areas. The exact microeconomic mix of expenditures turns out to have a significant macroeconomic influence which is much greater than in a Walrasian environment. Finally fiscal policy affects activity by inducing entry and exit of firms to and from the economy. Imperfect competition is here generally combined with increasing returns in production. In order to see the operation of these mechanisms clearly, in this section we abstract from the factors of the previous section which gave rise to monetary non-neutrality.[25]

Elasticity effects of the spending mix

When the demand for output has two components, private and public, its price elasticity is simply the weighted average of the individual elasticities. An increase in government spending, by increasing the share of public expenditure in the total, shifts the elasticity of demand towards that of public spending. If the latter is higher (lower) than the elasticity of private spending, overall demand elasticity rises (falls), and consequently the degree of monopoly tends to decrease (increase). Given the general finding that raising monopoly power lowers output, output could be expected to rise (fall).

This mechanism has been formalized and emphasized by numerous authors.[26] In practice it seems reasonable to argue that public spending is less price-elastic than private spending for most economies. This is obvious if the government fixes its spending, and its allocation between sectors, in real terms (zero elasticity), but it is also true if it fixes spending in nominal terms (unit elasticity). Such an argument implies a negative impact of an increase in spending on output. In general terms, governments often conceive of policies as affecting the trade-off faced by market participants. For example, in 1957 the British Chancellor of the Exchequer Thorneycroft argued that 'if ... money national income was pegged ... wages could push up prices only at the expense of employment: the onus of choice was, as it were, placed on the unions' (Dow, 1964, p.101). It is also possible to view one reason for the shift from volume planning to cash planning of UK public spending in the 1970s, and for the general attempt to reduce the scale of public spending in the 1980s, as being the desire to weaken monopoly (particularly labour monopoly) power, with the aim of countering the trend rise in unemployment.

Income effects on labour supply

Even in a Walrasian economy, one way in which fiscal policy may affect output is through the labour supply. A balanced budget increase in government spending will have a positive effect on output if leisure is a 'normal' good in households' preferences, by virtue of the higher tax burden which causes a lower demand for leisure and thus stimulates labour supply. Up to now we have deliberately excluded income effects on labour supply by the use of the utility function (1). Now we relax this assumption and show how imperfect competition strengthens such an effect, since it leads to a higher proportion of income entering the household's budget constraint in the form of profits.

The following simple example is taken from Mankiw (1988); other models exhibiting the same transmission mechanism are constructed by Dixon (1987) and Startz (1989). The representative household in a barter economy has Cobb–Douglas utility over goods and leisure

$$U=\left(\prod_{i=1}^{n} C_i^{1/n}\right)^c [H-N]^{1-c}. \qquad (27)$$

This implies that the price elasticity of private sector demand for each good is unity. To abstract from the 'elasticity' effects discussed above, government spending in each sector is taken to be fixed at G_i in nominal terms, so that real government spending, $g_i = G_i/P_i$, is also unit-elastic. On the production side we assume there are constant returns to scale ($X_i = N_i$), and thus marginal cost equals the wage, W. Given F firms per sector, the unit elasticity of demand implies that under Cournot–Nash equilibrium there will be a fixed mark-up of the price over the wage, with

$$\frac{P_i - W}{P_i} \equiv \mu = 1/F \Rightarrow \frac{W}{P_i} = 1 - \mu. \tag{28}$$

Firms' nominal and real profits in sector i are

$$\Pi_i = [P_i - W_i]N_i \qquad \Pi_i/P_i = \mu N_i. \tag{29}$$

Profits are immediately distributed and government spending is financed by lump sum taxation, so the household has the budget constraint

$$\sum_{i=1}^{n} P_i C_i + W[H-N] = WH + \sum_{i=1}^{n} \Pi_i - \sum_{i=1}^{n} G_i. \tag{30}$$

Since Cobb–Douglas utility implies constant expenditure shares, we can immediately write down the households' spending on leisure as

$$W[H-N] = [1-c]\left[WH + \sum_{i=1}^{n} \Pi_i - \sum_{i=1}^{n} G_i\right]. \tag{31}$$

The macroeconomic system is completed by assuming a symmetric goods market equilibrium with a competitive, clearing labour market. Using (28) and (29) in (31) yields an equation for N

$$N = cH - \frac{1-c}{1-\mu}[\mu N - g]. \tag{32}$$

We now have, differentiating (32), the following balanced budget spending multiplier

$$\frac{\mathrm{d}N}{\mathrm{d}g} = \frac{1-c}{1-c\mu}. \tag{33}$$

As the degree of monopoly, μ, increases from zero to one, we see that the multiplier rises from $1-c$ to unity. Thus it approaches the macro textbook multiplier for a high degree of monopoly. This may be understood in either of two ways. First, a higher mark-up increases the profit feedback from firms to households per unit increase in output. This boosts consumption spending and so the multiplier. Alternatively viewed, a higher μ lowers the real wage, $1-\mu$. The income effect on labour supply of the increased taxation is thereby strengthened, as may be seen from the term $1-\mu$ in (32). This is because the 'propensity to spend on leisure' is a constant, $1-c$: a lower real wage means more leisure is consumed per unit increase in exogenous income. The mechanism demonstrated here is not specific to a barter economy: Dixon (1987) obtains essentially the same outcome in a monetary economy with money-financed expenditure (see also Molana and Moutos, 1992, for a discussion of taxation in this model).

Effects of sectoral spending asymmetries

One of the most important ways that fiscal policy differs from monetary policy is in its inherently microeconomic content. This is obvious in the case of taxation: most taxes levied by governments alter supply-side incentives. It is, however, also true in the case of government expenditure: the government decides not just how much to spend, but also on what to spend it. The issue of how to allocate government expenditure is given much consideration by politicians, and quite rightly is seen by many as having important economic consequences. Some of these stem from the intrinsic value of government expenditure – on health, education and so on. However, in this section we will rather consider the case where government expenditure is 'waste'. We will also assume that apart from possibly different levels of government expenditure the 'fundamentals' of each market are the same – technology, the number of firms, union and consumer preferences. This rules out fairly obvious reasons for expenditure decisions based on differential employment effects due to capital intensity, import content and so on. By what mechanism can the allocation of government expenditure influence aggregate employment?

In an economy with perfect labour mobility and a competitive labour market, there can only be a single market-clearing wage in the economy. Whilst fiscal (or monetary) policy might influence this, it cannot influence relative wages. However, if there are sectoral unions, then these can in principle determine wages in their own sectors, and relative wages can then vary. In effect the union can be seen as an institution which limits labour mobility: the employed union 'insiders' are protected from the competition of 'outsiders', who may either be unemployed or employed in other industries. Since relative wages can then differ across sectors, the allocation of government expenditure amongst sectors has a foothold from which to influence aggregate output and employment.

In order to illustrate this, we outline the approach in Dixon (1988, 1991). In each of n sectors, there is a monopoly union which sets the nominal wage W_i in that sector, according to a Stone–Geary utility function (as is commonly used in empirical work – see Pencavel, 1984; Dertouzos and Pencavel, 1984). Each union sets its wage treating the general price level as given. Households have Cobb–Douglas preferences, there is no disutility of labour ($\theta = 0$), and constant returns to labour. Dixon derives a 'reaction

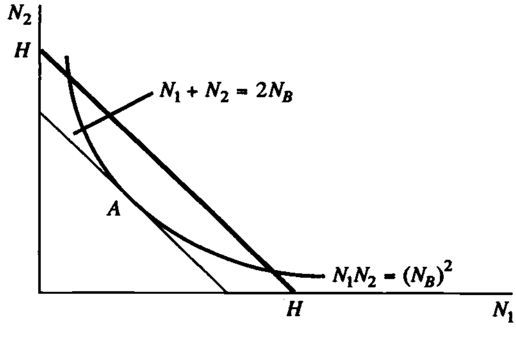


Figure 2.7 Sectoral and aggregate employment levels

function' for the sectoral union which states the nominal wage it wishes to set given the level of demand in its sector and the cost of living. The demand in that sector is determined by the sector-specific level of government expenditure (fixed in nominal terms) and the level of nominal national income. Given that prices are a mark-up on costs (determined by Cournot oligopoly), we can solve for the equilibrium nominal wage and employment in each sector for a given government policy.

The equilibrium employment equation is given by

$$N_i = N_B y_i^{1/n} \tag{34}$$

where N_B is a constant (determined by union preferences and the degree of monopoly in the product market), and y_i is the ratio of nominal expenditure in sector i (Y^i) to the geometric average of sectoral expenditures ($[\Pi_{i=1}^{n} Y_i]^{1/n}$). This yields the fundamental Natural Range property

$$\prod_{i=1}^{n} N_i = (N_B)^n \prod_{i=1}^{n} y_i^{1/n} = (N_B)^n. \tag{35}$$

That is, the product of sectoral employment levels is constant, defining a rectangular hyperbola in employment space. We can thus graph the combinations of possible equilibrium employment levels when $n=2$ as in figure 2.7. Total isoemployment isoquants are represented by negatively-sloped 45° lines, $N_1 + N_2 = N$. The total employment constraint is set by the aggregate labour supply, H. There is then a range of feasible aggregate employment levels: with a symmetric fiscal policy, aggregate employment is minimized at A with $N = 2N_B$; as we move away from the positively-sloped 45° line total employment increases up to full employment at H. For any given government policy, there is a unique equilibrium on the rectangular hyperbola $N_1 N_2 = (N_B)^2$. By altering the mix of government expenditure across sectors, the economy is made to move along the hyperbola, with the resultant change in aggregate employment.

Thus, in the unionized multisector economy the government's allocation of expenditure across sectors determines aggregate employment. This stands in total contrast to the Walrasian economy. In this case, perfect mobility of labour ensures that there is a single wage W for all workers, and furthermore that so long as $W > 0$ there will be full employment at H (since $\theta = 0$). Switching expenditure from one sector to another merely serves to cause exactly offsetting changes in employment to maintain full employment. The reason for the difference with imperfect competition is that the presence of unions means that wages may differ across sectors, and that as demand shifts across sectors relative wages alter, and thus changes in sectoral employment need not cancel out. The particular functional forms give rise to the specific 'natural range' result found in these studies, but the existence of a natural range in general does not depend upon them (see Dixon, 1988, Theorems 1 and 2).

Given that the government can increase total employment within the natural range, will it want to? Recall that we are treating government expenditure as waste. It can be shown that the real government expenditure multiplier in this model is less than unity (higher prices crowd out private expenditure – see Dixon, 1991, Proposition 6). However, despite this, government policy that increases total employment will increase the total utility of households (Dixon, 1991, Theorem 2). This is an interesting and possibly counterintuitive result. In unionized (as opposed to Walrasian) labour markets the real wage will usually exceed the marginal disutility of labour. Each employed worker thus earns a 'surplus': as total employment goes up, there is an increase in the total surplus as unemployed people become employed.

Fiscal effects on entry and exit

All the imperfectly competitive economies considered so far have treated the number of firms as fixed. In this sense they are 'short-run' analyses. One strand of the literature, beginning with Weitzman (1982) and developed further by Snower (1983), Solow (1986), Pagano (1990), and Green and Weale (1990), focuses on entry and exit of firms as the explanation of unemployment and the effects of fiscal policy. In simple terms, policy which induces entry will tend to increase competition in the market.[27] Thus fluctuations in the number of firms can influence output, with more firms leading to higher output.

Increasing returns in production are an essential feature of these models, and the ultimate source of the imperfect competition. Weitzman (1982) claimed to have explained involuntary unemployment on the basis of increasing returns and goods market imperfect competition alone, but his model, and the very similar one by Solow (1986), lack any treatment of the supply side of the labour market. Pagano (1990) builds an overlapping-generations version of the Weitzman model, completing it with a Walrasian labour market. This eliminates involuntary unemployment, but permits employment to fluctuate by allowing a variable labour supply. Fiscal policy in the form of a tax cut financed by bond issues is shown to reduce output and employment in his framework: the tax cut raises the interest rate and causes capital decumulation, reducing long-run output. The basic mechanism is the same as in Diamond's (1965) growth model, where the continuous birth of new households implies that 'Ricardian equivalence' fails to hold. However imperfect competition here reinforces the negative impact, because as firms are driven out of business the degree of monopoly increases, tightening the monopolistic restriction of output.

A further role for fiscal policy arises owing to the possible existence of multiple equilibria, which are a common feature of models with increasing returns. Pagano shows that there may be situations in which by changing taxation the government can eliminate a low output, Pareto dominated equilibrium, forcing an economy which has settled there to move to a superior one. (For other examples of multiple equilibria, see Cooper and John, 1988; Kiyotaki, 1988; Frank, 1900, and the rest of the 'coordination failure' literature described in Dixon and Rankin, 1991.)

Conclusions

What has imperfect competition added to the macroeconomic interest of the Walrasian model? First, it generates a suboptimally low level of output and employment, which is an apparently pervasive feature of real economies. This is suggested by any partial equilibrium model of imperfect competition, but the macromodels in addition enable us to see how inefficiently low output results from coordination failure amongst imperfectly competitive agents. Second, closely associated with low output, imperfectly competitive economies typically generate unemployment. When there is imperfect competition in the labour market, such unemployment is involuntary in the sense that there are individuals who would prefer to work more at the prevailing wage. Even where it is voluntary, as when the labour market is competitive, it is above the Pareto efficient level of unemployment.

Our focus has been on policy effectiveness. As regards fiscal policy, imperfect competition adds several important new mechanisms whereby policy can affect output, and modifies others. It is notable that, so long as money remains neutral, there is no general presumption in favour of a positive rather than a negative effect of a fiscal expansion on output. The transmission mechanisms are different from those of the Keynesian multiplier, and the sign of the effect depends on features of little importance in a Walrasian economy, such as relative price elasticities of private and public sector demands, or the sectoral allocation of spending. We may be tempted to think of these as 'supply-side' mechanisms, but this would be incorrect, since they work mainly via demand. Imperfect competition tends to undermine the textbook demand-side/supply-side dichotomy. However, the most critical difference between fiscal policy in Walrasian and imperfectly competitive economies is on the welfare side. Since output and employment are inefficiently low, it is much more likely that a fiscal policy change which increases output will bring about an increase in welfare (even if not necessarily a Pareto improvement). This is never true in Walrasian models, where if government expenditure is pure 'waste', an increase will always reduce welfare, irrespective of the change in employment.

As regards monetary policy, we emphasized from the start that we need some second distortion in addition to imperfect competition to generate real effects. The importance of imperfect competition is that without it the distortion would cause no, or only negligible, non-neutralities. Monetary policy, unlike fiscal policy, almost never has a negative effect on output, and where money is non-neutral the general behaviour of the economy is much closer to that of traditional macroeconomic theory. The reason is that there is then some form of endogenous nominal rigidity, i.e. a tendency of prices and wages to respond only weakly to aggregate demand. The study of imperfectly competitive macroeconomies thus tends to reinforce the view – which is still not especially widespread – that to generate some type of nominal rigidity is an essential part of any explanation of traditional macroeconomic policy effects.

What are the promising directions for future research? Two relatively unexplored areas are extensions to the open economy and to dynamic models. Work on the former exists primarily in the shape of studies of exchange rate pass-through, by Dornbusch (1987), Giovannini (1988), Froot and Klemperer (1989) and others (see the survey by Dixon, 1994a). This could profitably be merged with studies of policy effectiveness in the open economy such as Dixon (1990a): an example of this is Rivera-Campos (1991). Work on dynamic models exists in the studies by, amongst others, Caplin and Spulber (1987), Caplin and Leahy (1991), Pagano (1990), and Rankin (1992) and Jacobsen and Schultz (1994). This is still a disparate set of contributions: in particular, the complex strategic issues which potentially arise in the intertemporal setting have yet to be incorporated into macroeconomics. Other macroeconomic areas in which imperfect competition has been and will continue to be widely applied, but which we have not attempted to cover here, are the recent theory of endogenous growth (see, for example, Grossman and Helpman, 1991), and the theory of economic development (Murphy *et al.*, 1989). In view of the importance of nominal rigidities to traditional short-run macroeconomic questions, much future work is likely to focus on models which generate these. Serious questions remain for the dominant menu cost approach, such as whether it is reasonable that for a sufficiently large monetary shock neutrality will prevail. A difficult but potentially rewarding sequel would be to model not the direct administrative costs of price adjustment, but the indirect costs resulting from uncertainty and asymmetric information: some macroeconomic implications of these have begun to be explored by, for example, Andersen and Hviid (1990).

Notes

This chapter was originally published in *Oxford Economic Papers* (1994), pp.171–95; we would like to thank Oxford University Press for their kind permission to reproduce it here. We also acknowledge the very helpful comments of three referees, and of many seminar and conference participants.

1. A quite common simplification of the above separate treatment of households and firms is to assume a single type of agent (the 'farmer') who produces output using only his own labour as an input. This is used for example in Blanchard and Fischer (1989, ch. 8) and Ball and Romer (1989a, 1990, 1991), and has the advantage that the model reduces to one in a single type of market – namely for goods, with the labour market being suppressed.
2. Replacing the Cobb–Douglas form for subutility over aggregate consumption and money by a more general homothetic form makes no difference to the constancy of c, unless a different deflator for M is used. This becomes important in models with non-unit-elastic expectations, see p.47.
3. So formally $F=1$ but perfect competition is imposed by the assumption of price-taking: this enables a comparison with the monopoly case below which is not distorted by different numbers of production units.
4. When there is a centralized union, we assume that firms' profits are received by a separate rentier class of household which entirely consumes them. If the union received them, it would effectively control the whole economy and so would obviously choose the first best, competitive outcome.
5. This is consistent with the maximization of household's utility (1) if there is equal rationing of workers. With, instead, all-or-nothing rationing and random selection of workers, it is consistent with expected utility maximization if $e=1$, since (1) then exhibits risk-neutrality.
6. However, one of the best known models with sectoral unions does not fit this pattern. In Hart (1982) consumers are able to buy output only in one sector, so

one sector's output is neither a gross complement nor a gross substitute for another's and the money-wage elasticity of labour demand is not affected by having sectoral rather than centralized unions.

7. If goods are gross substitutes a rise in W_i and thus P_i will cause consumers to switch to other goods *j*, something which would not happen if all goods' prices rose together.
8. This objective function still derives directly from the household's utility function (1): we should think of each household as now constituting a separate union.
9. This point was not recognized in some early literature, which tended to regard any situation in which agents face downward-sloping demand curves as generating *ipso facto* demand management effectiveness. A simple fallacy is to argue that a money supply increase shifts outwards agents' demand curves causing them to raise output, forgetting that in a general equilibrium context cost curves will shift up by an exactly offsetting amount. In several studies Ng (1980, 1982a, 1982b, 1986) claims that imperfect competition breaks the classical dichotomy despite this, but his argument also rests on proving the existence of a local continuum of equilibria: see the interchange with Hillier *et al.* (1982). The clearest statement of the need for distortions in addition to imperfect competition is in Blanchard and Kiyotaki (1987).
10. 'Distortion' is not an ideal term because not all the extra factors we consider are necessarily sources of failure to achieve Pareto optimality in a competitive economy, although they could be.
11. Jones and Stock (1987) claim that imperfect competition is not necessary for the result. They assume 'near rationality', as introduced by Akerlof and Yellen (1985a, 1985b). Behaviour is 'near-rational' if the forgone utility or profits is less than some small fixed amount. If the failure of rationality is a failure to adjust prices optimally, then this is formally equivalent to menu costs. However, Jones and Stock assume it takes the form of a 'rule of thumb' in which competitive firms increase output whenever demand increases, which is clearly different from the notion of menu costs.
12. The formal similarities are explored in depth in Madden and Silvestre (1991, 1992).
13. Clearly a 'pecuniary' rather than a 'technological' externality. The term 'externality' is misleading in so far as pecuniary externalities are not usually held to cause market failure: the underlying source of the market failure here is of course just the imperfect competition itself.
14. The total effect of a fall in P_i on firm *j*'s profits is negative to the extent that it is undercut by a rival, but these relative price effects cancel out when all prices and wages are reduced.
15. It would seem to be a limitation of Ball and Romer's analysis that risk-aversion is present in their model only incidentally. They use the utility function from Blanchard and Kiyotaki's deterministic model without any modification: there is no separate risk-aversion parameter; risk-aversion simply happens to be present in the utility function due to the assumption of increasing marginal disutility of work ($e > 1$).

16. See n. 1 above.
17. This is almost identical to Akerlof and Yellen (1985a), who however do not comment on the help which their efficiency wage assumption provides in keeping opportunity costs small.
18. Again, 'distortion' may be a misleading label for certain kinds of 'real rigidities', since on Ball and Romer's definition they could consist of no more than, for example, highly elastic labour supply; yet it is clearly appropriate for those which derive from, for example, imperfect information.
19. Blanchard and Fischer (1989, ch. 8) provide intuitive explanations for these results.
20. As does Hart, we now assume that consumption is a scalar and that there is no disutility of work.
21. To allow us to assume that unions take their firms' customers' incomes (Y) and their members' consumption prices as given, Hart postulates many (identical) locations, each with its own labour and goods market, such that workers at one location are dispersed, *qua* consumers, amongst other locations.
22. It is true, as Patinkin (1965) showed, that non-unit-elastic expectations by themselves make money non-neutral. But here we are concerned with something stronger than mere non-neutrality: we are looking for a positive effect of money on output. Under perfect competition money is here non-neutral in that a change in M_0 affects real balances M_0/P – but this is an uninteresting non-neutrality.
23. In Rankin (1992) it is shown that with imperfect competition the assumption that expectations are validated does not tie down a unique long-run equilibrium. This still depends on the expectations elasticity, unlike in Walrasian models, and consequently so does the response to monetary growth.
24. Schultz (1992) however challenges the robustness of the zero-wage unemployment result. Extending the model to include overlapping generations of consumers, he shows that MR is always positive in such a world.
25. For an analysis where monetary non-neutrality and fiscal policy are combined, see, for example, Rankin (1993).
26. Amongst others, Solow (1986), Svensson (1986), Dixon (1990a), Thomas (1990), Rankin (1993), Jacobsen and Schultz (1994).
27. It is worth noting that this does not hold with the Dixit–Stiglitz (1977) CES version of monopolistic competition universally used in the menu cost literature (see Hart, 1985a, for a discussion of this fundamental microeconomic issue). Consequently most of the models with entry use Salop's (1979) 'competition on a circle' model.

3 Notes on imperfect competition and New Keynesian economics

Richard Startz

Introduction

New Keynesian economics is a counter-revolution against the ascendancy of the 'rational expectations' and 'new classical' schools which dominated macroeconomic research through much of the late 1970s and the 1980s. The New Keynesian approach uses the standard tools of microtheory; that is New Keynesians model consumers, workers, and firms as rational, maximizing agents. Further, markets clear. Thus, the 'modelling techniques' of the New Keynesians have much in common with those of the new classical school. The output of New Keynesian models, in contrast, follows along traditional Keynesian lines. Three broad, interrelated results emerge. First, the aggregate economy has multipliers, so an initial shock in the supply or demand for goods will be magnified in general equilibrium. Second, economic fluctuations are frequently not Pareto optimal. Recessions involve real welfare losses that, in principle, are avoidable. Third, government policy can be effective in manipulating output and can be welfare improving in so doing.

New Keynesian models invariably involve market failure, almost always some form of imperfect competition, and usually monopolistic competition. Why the link between New Keynesian models and imperfect competition? To illustrate the problem with the assumption of perfect competition, consider the following question for a hypothetical graduate school qualifying exam.

In this economy, consumers maximize utility and firms maximize profits. Markets are complete and perfectly competitive for all products and across all time periods. Assume any regularity conditions necessary to ensure an Arrow–Debreu equilibrium.

Specifically, the endowments, tastes, and technologies of the economy are as described in the following 666 equations.

...

Question 1: Describe the dynamic path of the optimal government intervention in the face of a temporary, one-unit shock to endowments as described in equation 111. (Ignore any role of income distribution in the social welfare function.)

The answer, of course, is that the government ought not intervene. All 666 equations above are just a red herring. If the conditions of the economy ensure Pareto optimality, then there is no (non-redistributive) role for government. I take Robert Lucas' oft-quoted advice that a good model doesn't leave $20 bills lying on the sidewalk as a shorthand reminder of this fact.

One way to summarize the 'Keynesian' research programme in the 40 years following the *General Theory*, is as the work of macroeconomists collectively flunking the just-stated exam question. The paradigm of the early postwar period rested on simple general equilibrium models, Hicks' ISLM diagram and Samuelson's multiplier/accelerator model. These models had only loose underpinnings in maximizing behaviour. While each sector of the economy was modelled assuming rational, and usually competitive, behaviour on the part of agents, general equilibrium predictions seemed to suggest unexploited and unexplained profit opportunities. Nonetheless, both theoretical models and the very successful programme of building econometric forecasting models went along merrily producing 'policy recommendations'. It was this contradiction between the well known results of mathematical general equilibrium theory and the practice of applied macroeconomic model building that led macroeconomics into scientific disrepute in the 1970s and 1980s.

The new classical economists accuse the Keynesians of being unscientific for making the claim that there are specific macroeconomic policies for improving social welfare – the $20 bill on the sidewalk being opportunities left unexploited by the private sector. The Keynesians, in turn, point to the recessions of the 1930s and early 1980s as periods of great social inefficiency. They claim these episodes disprove the classical theories. Good science would not lead society to fall flat on its face – having tripped over a $20 bill in plain sight – because theory 'proved' the impediment could not exist. In other words, the new classical economists maintain a Bayesian prior that agents 'behave rationally' and the Keynesians' prior is that the empirical behaviour of the economy is very different from what would be generated by atomistically competitive rational agents. Neither side is inclined to allow evidence to move their posterior.

New Keynesian economists, like the new classicals, build models in which agents behave rationally.[1] Like the Keynesians, they believe that the real economy departs in important ways from an Arrow–Debreu equilibrium. In a sense, New Keynesians believe we do observe $20 bills, and seek to explain their presence.

New Keynesian economists study that very narrow subset of non-Arrow–Debreu equilibria in which aggregate externalities and positive feedback play an important role.[2] The models are necessarily built on incomplete markets or imperfect competition. The most common failure of perfect markets has been monopolistic competition. In part, this choice results from the belief that monopolistic competition is pervasive in a modern economy. In part, the choice reflects technological improvements which have made these models easy to manipulate, specifically the invention of the Dixit–Stiglitz model of product variety. Even studies which employ essentially identical models of monopolistic competition do so for disparate purposes. Rather than a simple linkage, there are many different ties between imperfect competition and the New Keynesian economics.

New Keynesian work has a mostly macroeconomic-modelling appearance. It is nonetheless useful to consider the microeconomic mechanisms which cause a departure from the usual perfect equilibrium. There are several. One important mechanism is to put a wedge between private and social costs so that the economy operates always inside the omniscient-central-planner-production-possibility-frontier (OCPPPF). Demand fluctuations move equilibrium around inside the frontier. Blanchard and Kiyotaki (1987), Bryant (1983), Diamond (1982), Dixon (1987), Hart (1982), Mankiw (1988) and Startz (1989) fall in this genre. A second mechanism has scale fluctuations causing endogenous reallocation of input factors across production technologies of differing productivity. Shleifer and Vishny (1988) is one such. Finally, one ought to consider the role of imperfect competition in explaining endogenous movement of the OCPPPF. This is the realm of 'New Growth Theory,' Romer (1986) and many others.

The strands of New Keynesian economics

For purposes of exposition, I divide the work of the New Keynesians into four strands: real models of monopolistic competition and aggregate demand, real models of aggregate demand and search equilibrium, nominal models of monopolistic competition and rigid prices,[3] and aggregate demand and multiple equilibria. Finally, I will comment on some work that does not quite fit in any of these categories. While I reference a number of studies in what follows, nothing here should be construed to be a literature review. There are now four long – and excellent – survey papers. See Bénassy (1993), Dixon and Rankin (chapter 2 in this volume), Gordon (1990), and Silvestre (1993). The Winter 1993 *Journal of Economic Perspectives* symposium on 'Keynesian Economics Today' provides competing views on the value of the New Keynesian agenda. Matsuyama (1992) nicely exposits much of the intuition connecting monopolistic competition and complementarity.

Monopolistic competition and aggregate demand

The models of Hart (1982), Dixon (1987), Mankiw (1988) and Startz (1984, 1989, 1990) present models of imperfect competition which can be used as foundations for the Keynesian cross. There are several ways of thinking about what monopolistic competition buys you. The first is that the social return to labour is above the private return. Since part of production goes to profits rather than to factor inputs, less labour is sold than would be optimal. Something which expands the economy has the potential to move the economy closer to a first best. The second way to think about the nature of equilibrium is that monopoly profits create an externality that generates a positive feedback through aggregate demand. An increase in demand generates economic profits which are returned to the firms' owners. The owners, whose wealth has increased, increase their demand, further increasing output and profit. Thus, a 'Keynesian' multiplier results.

To illustrate, consider a slightly changed version of Mankiw (1988) with some elements of Startz (1989) thrown in. Suppose that the representative consumer has a utility function defined over two goods, C and G, and leisure, L. Utility is given by

$$U = \alpha \log C + \beta \log G + (1 - \alpha - \beta) \log L$$

where $0 < \alpha, \beta, \alpha + \beta < 1$.

The agent sells part of her labour endowment, ω, at a real wage rate w. C and G are produced with identical technologies so the price of each is 1. Real (*per capita*) profits, if any, are Π and lump sum taxes are T. Suppose that the consumer's allocation of G is supplied by the government. Thus the consumer's budget constraint is

$$C = w(\omega - L) + \Pi - T$$

and the government's budget constraint is $G = T$.

Taking G, T, and Π as exogenous, the representative consumer will demand

$$C = \frac{\alpha}{1-\beta} [w\omega + \Pi - T].$$

Imposing the government budget constraint, GNP and welfare are

$$Y = C + G = \frac{\alpha}{1-\beta} [w\omega + \Pi] + \frac{1-\alpha-\beta}{1-\beta} [G] \quad \text{(aggregate demand)}$$

$$U = (1-\beta)\log C + \beta \log G + k_1$$

where k_1 is an unimportant constant. If the government sets G to maximize welfare, the economy ends up at

$$C = \alpha[w\omega + \Pi]$$
$$U = \log[w\omega + \Pi] + k_2.$$

Suppose that the economy is competitive, or at least that Π is exogenous to the economy as well as to the individual agent. Then this equilibrium is the same as agents would have chosen for themselves if G were provided privately rather than by the government. In other words, the government is replicating a *laissez-faire* equilibrium. What happens if the government decides to undertake 'expansionary' policy and increase G one unit? By inspection of the GNP equation, GNP rises,

$$dY/dG = \frac{1-\alpha-\beta}{1-\beta},$$

although by less than one, as consumption of the good named C is partially crowded out. Since G had been set to maximize welfare, the increase in G necessarily reduces welfare. This is very much the new classical result. Government policy that moves the economy away from *laissez-faire* may be 'effective', in that output rises, but it is clearly undesirable.

Now introduce imperfect competition. Skipping a careful specification of the industrial structure, assume producers mark up prices over marginal cost by the factor μ. Marginal real profit from sale of an additional unit of output is $\partial\Pi/\partial Y = (\mu-1)/\mu$. Combining this relation with the aggregate demand equation gives

$$dY = \frac{\alpha}{1-\beta}[w\omega + d\Pi] + \frac{1-\alpha-\beta}{1-\beta}[dG] \quad \text{(aggregate demand)}$$

$$d\Pi = \frac{\mu-1}{\mu}dY \quad \text{(aggregate profit)}$$

where $d\Pi$ enters the aggregate demand equation via income in the consumption function. From inspection of the two equations, the balanced budget government spending multiplier is

$$\frac{dY}{dG} = \frac{1}{1-\dfrac{\alpha}{1-\beta}\dfrac{\mu-1}{\mu}} \cdot \frac{1-\alpha-\beta}{1-\beta}, \quad \frac{1-\alpha-\beta}{1-\beta} \leq \frac{dY}{dG} < 1.$$

This New Keynesian multiplier is larger than the effect of government spending in the competitive version. At the macroeconomic level, the

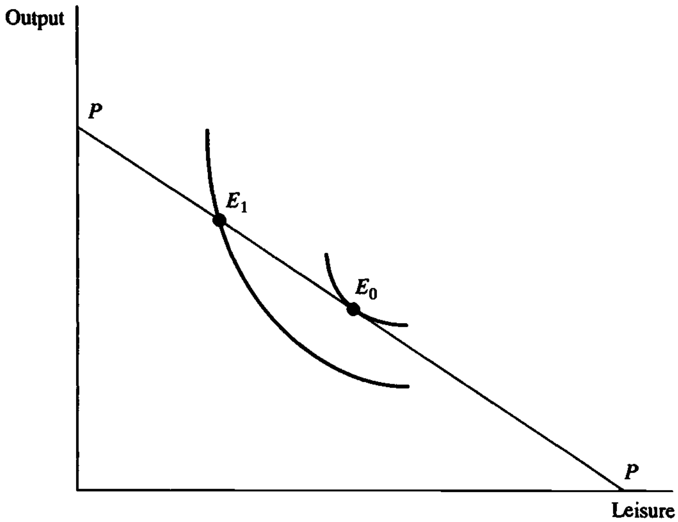


Figure 3.1 Fiscal policy under perfect competition

difference is due to the additional path of positive feedback through aggregate profits. At the microeconomic level we have a 'strategic complementarity', a term coined by Bulow, Geanakoplos and Klemperer (1985) and brought into the macroeconomics literature by Cooper and John (1988). Increased demand due to government spending increases aggregate profit, which leads individual consumers to follow more aggressive strategies, that is to increase their own demands.

What happens to welfare? Suppose government spending is set initially at the *laissez-faire* level for this economy. An increase in G has two effects. First, at the *laissez-faire* point, the marginal utilities from C and G are equal, since that is what private agents would have chosen. An increase in G increases utility by less than the crowding-out of C reduces it. Exactly as in the competitive model, increased G distorts private choice and reduces welfare by a (literally) second order amount. Second, the expansion of demand increases aggregate profits and therefore income. The increased income increases welfare by a first order amount. Thus, starting from the *laissez-faire* point, some finite amount of expansionary fiscal policy is necessarily welfare improving.

At the macroeconomic level, imperfect competition buys us a route for positive feedback, aggregate demand multipliers, and welfare improving fiscal policy. The question is, what is going on in the microeconomic structure? The answer is that private agents regard aggregate profits as exogenous, when in fact profits are endogenous. Aggregate profits act as an

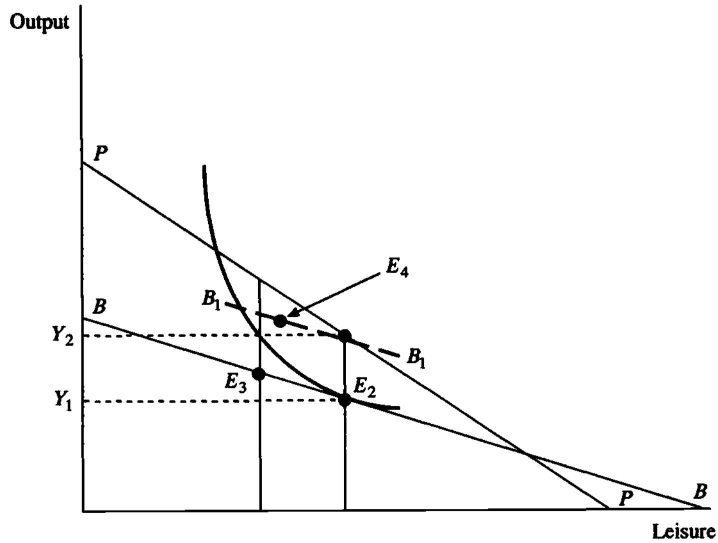


Figure 3.2 Fiscal policy under imperfect competition

externality. Atomistic decisions do not lead to a first best solution. This is an example of what we call a 'coordination failure'. It is not surprising that social policy which moves the equilibrium closer to the first best is welfare improving.

All this has a very 'macroeconomic' look. What might an intermediate-micro diagram look like? In figure 3.1, we show indifference curves between output (C and G combined) and leisure. Production exhibits constant marginal product and takes place under a regime of perfect competition. The PP line is the production possibility set and coincides with the consumer's budget line. Under *laissez-faire*, equilibrium is at E_0. Suppose the government introduces a lump sum tax and uses the revenue to transfer G to agents. Equilibrium moves to a point like E_1. (The budget line is not tangent to the indifference curve because the consumer is being forced to take more G and less leisure than she would choose privately.) While fiscal policy is expansionary, it is also, obviously, a bad idea.

In figure 3.2, the production function PP has constant marginal cost plus a fixed cost component. Because of the monopolistic mark-up, the consumer's budget line, BB, is more shallow than the production function. Assume that equilibrium is initially at zero profits and that this is reflected in the height of the budget line. The consumer chooses point E_2. The vertical distance between Y_1 and Y_2 is the producer's marginal profit, which is just enough to cover the fixed cost, leaving zero total profit. Suppose we repeat the earlier exercise. The consumer, taking profits to be exogenous, 'initially' moves to E_3.[4] But at this point profits have risen because the BB line is flatter than the PP line. As a result each consumer's budget line rises, further increasing demand and thus further increasing profits. In general equilibrium, the budget line rises, say to something like B_1B_1, a segment of which is shown, and the economy moves to E_4. (The income expansion path given the constrained goods/leisure choice runs through E_3 and E_4.) Not only is output up but, as drawn, so is welfare. This last result depends on the balance between the leftward movement of E_3 reducing welfare and the upward movement to E_4 increasing welfare. (With Cobb–Douglas preferences, it turns out that a small intervention is always good and a large one is always bad.)

All this relied on the social production function, PP, not moving. So the analysis given assumes away any changes in the strategic games played by producers. As the most simple illustration of this limitation, the movement shown in figure 3.2 holds constant the number of firms even though in the presence of fixed costs, average costs may be lower at E_4 than at E_2. Since a point like E_4 has positive aggregate profits, the final long-run equilibrium ought to involve some further movement of the PP curve. In Startz (1989), free entry increases the number of firms until the aggregate fixed costs are high enough to eliminate profits. In that model the movement of the PP curve is just enough to return output to its prefiscal policy level.

The model as stated is really incomplete, because the production technology and market structure are unspecified. A number of models use monopolistic competition to support an increasing average returns to scale technology. With the Chamberlinian large numbers assumption, each producer takes her competitor's actions as fixed. The Dixit–Stiglitz model makes it particularly easy to compute the monopoly mark-up, as each firm faces a constant elasticity demand curve. What is necessary at the micro level is for there to be support for an aggregate positive feedback loop which 'disappears' at the atomistic level. Monopolistic competition provides such a structure, but it is not the only way to get there. Hart (1982) and others use Cournot–Nash equilibria.

Aggregate demand and search equilibrium

Diamond (1982) and Diamond and Fudenberg (1989) produce a coordination failure in a model of probabilistic search for trading partners. Agents can enjoy the fruits of their labour only by exchanging their produce with a trading partner. The more likely an agent is to find a trading partner, the greater the *ex ante* value of working. If more agents choose to work, then any given agent is more likely to find a trading partner. Thus the decision to

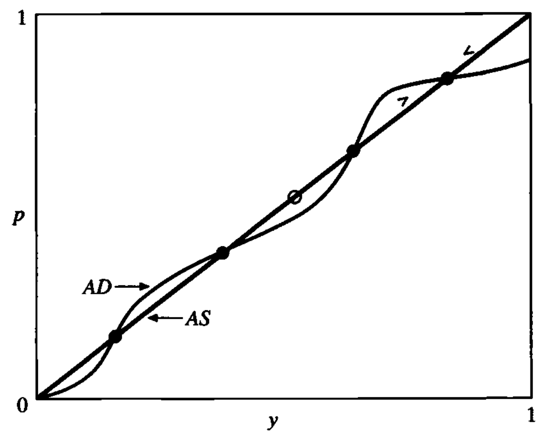


Figure 3.3 Strategic complementarity in Diamond's model

produce increases the probability of success for all other traders, acting in this manner as an aggregate demand externality.

To illustrate, consider a much simplified variant of Diamond (1982). Each agent owns a coconut palm. Climbing the tree costs disutility $c < 1$. If the tree is climbed, the agent harvests one coconut which is either sold, returning one util, or rots. The equilibrium probability of selling the coconut is p. Maximizing expected utility, the agent climbs her tree if $p \cdot 1 + (1 - p) \cdot 0 > c$. Assume that c is distributed among agents uniformly on the unit interval. The number of coconuts picked, y, is then p. (The first best solution is $y = 1$.) Think of this as giving the aggregate supply of coconuts.

Aggregate demand is written as the probability of finding a buyer as a function of the number of agents searching, $p(y)$. The function increases monotonically from $p(0) = 0$, with the exact form depending on the search technology. This is the source of strategic complementarity. Figure 3.3 shows one possibility.

As drawn, there are several rational expectations equilibria, none of which is a first best. The equilibria alternate between stable and unstable. To see this, suppose agents mistakenly believe that the probability of a sale is given by the point '∘', where aggregate supply is above aggregate demand. Producing at point '∘' they will go to market and find too few buyers and therefore will reduce output. The choice of equilibrium depends on an exogenous self-fulfilling prophecy as to the level of p. In other words, the economy can be moved from one static equilibrium to another by Keynesian 'animal spirits'.

Now add a government to the model which engages in taxes and

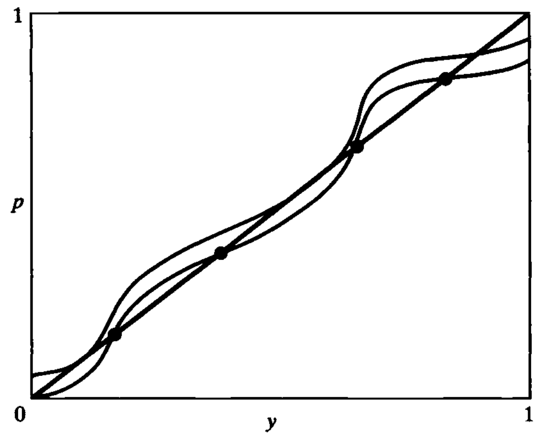


Figure 3.4 Increase in aggregate demand

purchases, thus increasing the level of market activity at any level of output. Aggregate demand is increased, as shown in figure 3.4.

The stable equilibria move to higher levels of output. Since more output is always better in this model, fiscal policy improves welfare as well as increasing output.

The level of market activity acts as a public good and increased government provides more of it. One ought to ask why the government can provide something private agents can not. In the previous section, fiscal policy distorted private choice so no individual agent would voluntarily undertake the kinds of trades imposed by the government. In this section, we might assume some transaction cost of trading with the government. Every agent wants there to be expansionary fiscal policy in order to increase p, but no agent will voluntarily accept the transactions cost.

Imperfect markets enter in that no price clears the goods market – there is no auctioneer. (Monopolistic competition is not used in this model.) The positive feedback loop operates through increases in demand increasing the probability of making a sale, which increases output and thus demand. Atomistic agents treat p as exogenous, even though p is endogenously determined at the aggregate level.

Monopolistic competition and rigid prices

The greatest intellectual problem in macroeconomics is to understand the role of nominal prices. Economic theory is clear that only relative prices matter and use of a fiat numéraire is a mere bookkeeping convenience.[5]

Most Keynesian economists, at least in the USA, believe that as an empirical matter changes in the real money supply move real GNP and that in the short run nominal prices are rigid. The nominal money supply is therefore an effective tool to manage output.

The development of the menu cost literature as an underpinning for rigid nominal prices is the first serious theoretical development in this area. The seminal works are Akerlof and Yellen (1985b), Blanchard and Kiyotaki (1987), Kiyotaki (1985) and Mankiw (1985). There is also a growing modern body of work examining rigid prices *per se* (see Ball, Mankiw, and Romer, 1988, Rotemberg, 1987). Chapter 8 in Blanchard and Fischer's *Lectures on Macroeconomics* is devoted to this topic. Gordon's excellent survey in the *Journal of Economic Literature* (1990) 'sifts and criticizes' a decade's work in this area.

The best exposition of the menu cost literature is Mankiw (1985). What follows is a variant of that model and of Blanchard and Kiyotaki (1987). Suppose all firms are price-setters, that aggregate demand depends on the real money supply, and that a firm's market share depends on its price relative to the nominal price level. The profit function[6] of the ith firm is $\pi_i = \pi(p_i/p, M/p)$. Initially, there is some set of prices and corresponding quantity allocations which are privately optimal given the nominal money supply.

Suppose the nominal money supply is now increased Δ per cent. In the absence of transaction costs, all nominal prices will rise by Δ per cent. Relative prices are the ratio of two nominal prices and therefore remain unchanged. The real money supply, profits, and quantity allocations will be as before.

Suppose instead there is a very small cost c to changing the nominal price, a 'menu cost'. If the gain from changing price is less than c, prices will remain at their initial levels. The gain from changing prices is $\partial\pi/\partial p_i[(1+\Delta)p_i - p_i]/p$. The initial optimum was found by setting $\partial\pi/\partial p_i$ equal to zero, so for small Δ, $\partial\pi/\partial p_i$ is approximately zero. In other words, in the neighbourhood of the optimum, the rate of change in the objective function is second order small. Even though menu costs are first order small, the potential gain from changing prices is smaller and each firm will leave prices fixed. As a result, relative prices remain unchanged, but the real money supply rises by Δ and aggregate demand increases. Further, if the level of output was below first best, as is likely under imperfect competition, welfare rises.

The key insight is not that transaction prices might lead to sticky prices, but that *very small* transaction costs might lead to sticky prices. Imperfect competition is essential here. The more elastic the demand faced by firms, the greater the profit loss from having prices a given distance away from the optimum. Under perfect competition, demand is infinitely elastic and transaction costs will never cause sticky prices.

Aggregate demand and multiple equilibria

One form of early (pre-New) Keynesian theory created a set of equations with multiple equilibria as solutions and suggested that economic fluctuations might be the economy's jumping from one equilibrium to another. Obviously one wants a theory that says more than that a set of non-linear equations has more than one solution. In particular, multiple solutions are much more interesting when they can be ranked (by a weak Pareto criterion, for example). One such theory is the 'sunspot' literature, in which agents agree on some inherently irrelevant signal as a coordination device. Elements of the New Keynesian literature give aggregate demand a central role in coordination. It is natural to combine the question of coordination with imperfect competition. Perfect competition may result in multiple equilibria, but aggregate signals are irrelevant to atomistic perfect competitors once they have conditioned on the price vector. In contrast, imperfect competitors do care about quantities. The Diamond model discussed above is one example where aggregate demand matters to individual agents so that changes in expectations of aggregate demand can move the economy from one equilibrium to another.

Kiyotaki (1988) presents a model of multiple expectational equilibria centred around the investment sector. The production function exhibits increasing returns to scale. The role of monopolistic competition is to support an increasing returns to scale equilibrium. If firms are optimistic about the future, they invest heavily now. The capital stock is large and therefore output and demand are high, justifying the optimistic expectations. If instead, expectations are pessimistic, a low level equilibrium is reached.

Shleifer (1986) and Shleifer and Vishny (1988) present models which specifically emphasize aggregate demand spillovers. In Shleifer (1986), firms receive cost-reducing inventions at various times. However, they may end up innovating, putting the inventions to work, in a coordinated fashion. Innovation reveals the invention, which can be copied by other firms after a lag. To extract the maximum rent from the invention, firms wish to innovate during periods of temporarily high demand. Innovation itself raises demand. Shleifer demonstrates how endogenous cycles can result.[7]

In Shleifer and Vishny (1988), each industrial sector has a high fixed cost/ low marginal cost monopolist and a competitive fringe. If demand is low, it does not pay for the monopolist to operate. If demand is high, the monopolist produces. Since the monopolist has a high marginal product, aggregate output rises, calling forth monopolists in other sectors. Effectively, positive feedback operates through a market structure in which society shifts to a higher production possibility frontier during good times.

In a slightly simplified version of the model, each consumer spends $y=\pi+L$ on each commodity, where π is aggregate profits and L is inelastically supplied labour. Fringe firms transmute one unit of labour into one unit of output. The monopolist, if it operates, buys a cost-reducing technology at cost F which allows it to produce $\alpha>1$ units of output per unit of labour. The monopoly price is a scintilla under unity, capturing the entire market. The monopolist operates if its profit, $\pi=y-y/\alpha-F$, is positive. Suppose F is distributed according to $H(\cdot)$, which is uniform on $[0, H^{max}]$. Let F^* be the breakeven cost. Equilibrium is defined by two equations in income; the first defines the breakeven point; the second gives aggregate demand.

$$F^*=\frac{\alpha-1}{\alpha}y \qquad \text{(breakeven)}$$

$$y=\int_0^{F^*}\left(\frac{\alpha-1}{\alpha}y-F\right)\mathrm{d}H(\cdot)+L. \qquad \text{(aggregate demand)}$$

Together, the equations give the multiplier relation

$$\frac{\mathrm{d}y}{\mathrm{d}L}=\frac{1}{1-\frac{1}{H^{max}}\left(\frac{\alpha-1}{\alpha}\right)^2 y}.$$

The multiplier is greater than 1, but decreases monotonically as the size of the economy grows relative to costs of better technology (y/H^{max}). Note that higher aggregate income is associated with more extensive imperfect competition and moving closer to the first best. The high fixed cost/low marginal cost technology assumed in this model does not permit competition to support an efficient equilibrium.

The illustrative model here demonstrates positive feedback and has a unique equilibrium. The actual model in Shleifer and Vishny adds informational considerations to produce multiple equilibria.

Related work

The primary work on abstract New Keynesian theory is Cooper and John (1988). The authors introduce the term 'strategic complementarity' to macroeconomics, emphasizing the importance of positive feedback, and give examples using both game theory and imperfect competition. We know that for a single individual facing a competitive market there is a sense in which 'most' goods are substitutes rather than complements.[8] The authors argue that understanding complementarity is central to understanding the macroeconomy. Akerlof and Yellen (1985a, 1985b) serve as underpinnings for the menu cost literature. In addition, the authors explain why small deviations from perfectly competitive behaviour on the part of individual agents can yield large and persisting deviations from the first best at the aggregate level. While not yet greatly exploited in the literature, this may provide a basis for a better understanding of inefficient business cycle fluctuations without requiring enormous departures from the competitive paradigm. Finally, while not a New Keynesian model *per se*, the work by De Long, *et al.* (1990), on 'noise traders', presents a related model of positive feedback in an efficient market.

Conclusions

The New Keynesian economics is too incomplete, and this is too incomplete a review, to provide a final word on the literature. However, several elements stand out. First, the New Keynesian economics is about the strategic coordination of decision-making through aggregate demand. For this, perfect competition is a non-starter. Second, New Keynesians search for paths of positive feedback and strategic complementarity. Finally, imperfect competition in general and monopolistic competition specifically are endemic in New Keynesian economics. However, the literature is too new for there to be any agreement on how they should be used.

Notes

This chapter originated as background material for Robin Marris' *Reconstructing Keynesian Economics with Imperfect Competition: A Desk-Top Simulation* (1991). This version was revised for the University of Warwick Macroeconomics Workshop (July 1993). I have benefited from discussion with Jean-Pascal Bénassy, Russ Cooper, Huw Dixon, and others, but the usual disclaimer of assignation of blame applies most strongly. The chapter is intended to give a personal view of the ongoing development of the 'New Keynesian economics'. The view is my own and I have not attempted to paint a picture shared by the profession as a whole. Because some important work is not discussed here, this is just not a valid literature review, but assuming the reader understands the context, I will proceed without further apology.

1. Not all New Keynesians *believe* economic agents are rational, but agents ought to be modelled *as if* they were. As a matter of history of thought, optimizing

models drive out non-optimizing models. To quote Gordon (1990), 'Any attempt to build a model based on irrational behaviour or submaximizing behaviour is viewed as cheating'.

To be more careful, both New Keynesian and new classical economists sometimes consider carefully controlled departures from the economist's definition of 'rationality'. Both groups are interested in topics such as 'bounded rationality' and 'adaptive learning' under uncertainty.

2. As a rule of thumb, the number of definitions of 'New Keynesian economics' is approximately equal to the number of New Keynesian economists. (This rule of thumb may sound circular, but is in fact resolved in general equilibrium.) The term 'New Keynesian' seems to have become a marketing rubric for ideas in opposition to the new classical school, rather than a single school centred around what Keynes said. The benefit of this is broad inclusion. The cost is that we muddy the history of thought.
3. Gordon (1990) tells us: 'The task of new-Keynesian economics is to explain why changes in the aggregate price level are sticky . . .' I emphasize here real, as well as nominal, models. If polled, I suspect the majority of macroeconomists would pick Gordon's definition.
4. Remember that there were no profits in figure 3.1, so the analysis ended here.
5. The inflation rate is a relative price. But the serious disputes between classicals and Keynesians lie in the failure of neutrality, not super-neutrality.
6. Assume the function to obey any necessary regularity conditions, in particular, to be differentiable in both its arguments. This is a substantive assumption, not just one of convenience. Perfect competitors face infinitely elastic demand and therefore their profit functions are not differentiable in their own-price.
7. Interestingly, stabilization policy can be harmful in such a model. If there are fixed costs of innovation, smoothing the business cycle may eliminate the temporary rents needed to draw forth innovation.
8. Note also that proofs about the uniqueness of Arrow–Debreu equilibria pass through much more easily if all goods are gross substitutes.

Part II

Goods market imperfections

4 Optimal labour contracts and imperfect competition: a framework for analysis

Russell Cooper

Introduction

This chapter investigates the importance of labour market relations to the macroeconomic inefficiencies associated with imperfectly competitive product markets. Hart (1982) argues that models of imperfect competition are capable of generating macroeconomic results such as an underemployment equilibrium, multiplier effects and so forth.[1] Related results on coordination failures, i.e. the existence of Pareto ordered Nash equilibria, for this class of models are reported by Heller (1986), Kiyotaki (1988), Roberts (1987, 1988) and Cooper (1987).[2]

Some, but not all, of the studies on the macroeconomic implications of imperfectly competitive product markets appear to rest on employment relations between firms and workers which are *privately suboptimal.* For example, Hart (1982) stresses the importance of market power by the *suppliers* of labour in his study of imperfectly competitive economies.[3] The 'contracts' between workers and firms in Hart's model are *not* privately optimal in that the allocation does not lie on a contract curve between the contractants.

Another example of this point arises in Weitzman's (1983, 1985) studies advocating share contracts. Weitzman supposes that a 'wage system', in which the wage is predetermined and employment is demand determined, exists for the trading of labour services. He then contrasts the welfare properties of this system with alternative compensation schemes assuming that product markets are imperfectly competitive. Cooper (1988a) shows that an imperfectly competitive economy with this structure of labour contracts will generate underemployment equilibria and multiplier effects using a model similar to that of Hart (1982). That contracts such as these produce socially suboptimal behaviour is not surprising since these contracts are generally not even privately optimal. Hence, an important aspect of Weitzman's argument for the introduction of a share system appears to rest on the assumption that agents trade labour services in a privately suboptimal fashion.

An important issue in this line of research is thus the role of labour markets in coordinating the activities of firms in the product market. If there is an underemployment equilibrium, why are there no forces operating in the labour markets to remedy the situation? Are the inefficiencies and other Keynesian features reported in these studies merely a consequence of some form of suboptimal labour arrangement which limits the coordination of activities? To the extent that the implications of imperfect competition reported are a consequence of the assumed contractual structure, imperfectly competitive product markets seem less important. As a consequence, this approach is subject to the criticism that the 'Keynesian features' are simply a product of an arbitrary contracting structure.

There are two approaches to addressing this concern. First, one might argue that these contracts are empirically relevant. Second, one can study a model with imperfectly competitive product markets *and* privately optimal labour contracts. This chapter takes the second approach by introducing privately optimal labour contracts into a simple general equilibrium model of imperfect competition. The results indicate that economies with imperfectly competitive product markets can experience underemployment equilibria even in the presence of optimal labour contracts. In contrast to Hart (1982), this underemployment reflects the product market power of the firm–union coalition and not the bargaining power of the union in its relationship with the firm. The framework can then be used to study the comparative static properties of these models, as well as their policy implications.

Overview of the model

Before delving into the details of the model, it is useful to begin with an informal overview of the basic approach to this problem. The model has two key components. First is the design of a labour contract between a group of workers and a firm. Second is the interaction of this coalition of workers and a firm with other such coalitions in the determination of a product market equilibrium.

Figure 4.1 displays the interactions between a single firm, the group of workers it contracts with and the product market in which it sells. Labour contracts are negotiated prior to the determination of the state of the economy so that the maximal gains to risk sharing are possible. The state is fully described by the level of endowment of a non-produced good, preferences and technology.

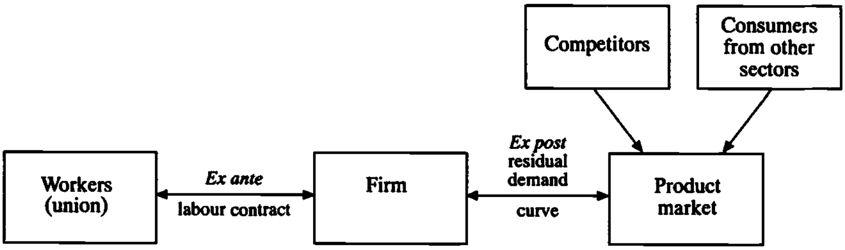


Figure 4.1 Interactions between the firm, its workers and its product market

Assume that both the firm and the workers costlessly observe the state so that contingent compensation and employment rules are feasible. Note that the contractants negotiate both compensation and employment rules rather than granting the firm the latitude to select employment *ex post* given wages. This is in keeping with the literature on optimal labour contracts.[4] Any inefficiencies in the equilibrium allocations will not be the consequence of direct restrictions on the contracts. Following the specification of the labour contracts, the random endowments, preferences and technology of the agents in the economy are determined and the product markets open.

The economy is composed of a number of sectors producing different products. Demands in the economy are structured to highlight the specificity of production relative to consumption. In particular, workers and firms are assumed to consume all products in the economy except for the output they produce. This structure of demands is common to models of coordination failures and imperfect competition.[5]

In each sector, there are a small number of firms who act as oligopolists. The output of these firms is determined, on a state-contingent basis, in their labour contract. In designing this contract, the workers and the firm recognize the market power of their coalition in the product markets. In equilibrium, there is no incentive for a given worker–firm pair to change output *given* the actions of other worker–firm coalitions.

The model of imperfect competition explored here has as a key feature the underutilization of labour resources. As a consequence, increases in demand can be met by expanding output without the need to induce workers to supply more labour by increasing real wages.[6] In addition, wages do not decentralize employment decisions due to the presence of optimal labour contracts and prices do not decentralize product market behaviour because of imperfect competition. Evidence that wages and prices fluctuate little relative to quantities is consistent with this class of models.

Structure of the economy and a characterization of equilibrium

Consider an economy composed of S sectors indexed by $s = 1,2,\ldots,S$. In each sector there are F identical firms producing a homogeneous product. Firms in each of the S sectors produce a distinct commodity so that there are S produced goods in the economy. There are N workers per firm in each sector who provide labour services to their respective firms. This allocation of workers to firms is not really a restriction given the symmetry in the model. For simplicity we will often term a group of N workers, a union, and view contracts as negotiated between a union and a firm.

There is a non-produced good in the economy which is the endowment of the *outsiders*.[7] These outsiders trade their endowments for the goods produced in each of the S sectors.

Preferences

In models of this genre, identical homothetic preferences are often used to avoid aggregation problems and distribution effects. These preferences imply that demand curves are linear in income which eases the characterization of the Nash equilibrium in the product markets. We take a step further and assume that all workers and firms have Cobb–Douglas preferences over the commodities they consume. This yields a closed form characterization of the equilibrium which is useful for expositional purposes and tractability.[8] As noted above, agents do not consume the commodity they are active in producing. Hence, workers and firms in sector s consume the goods produced in the other sectors and the non-produced good. Relaxing this assumption would lead firms to recognize that output expansions would influence the position of their demand curves so that the underemployment effects would be weakened.

Workers in sector s have preferences given by

$$U(c_{-s}, m, n) = U(\Pi_{k \neq s}(c_k)^{\alpha} m^{\beta} - rn) \qquad (1)$$

where n equals one if the worker is employed and zero if the worker is unemployed and c_k denotes consumption of the good produced in sector k. The consumption of the non-produced good is denoted by m. The disutility of work equals r where $0 \leq r < 1$. Assume that $(S-1)\alpha + \beta = 1$, so that preferences are homothetic and that $U(\cdot)$ is strictly increasing and strictly concave so that workers are risk-averse. The demands generated by these preferences are

$$c_i = (\alpha/p_i)Y \text{ for } i \neq s \text{ and } m = \beta Y \qquad (2)$$

where Y is the worker's income.

Firms (i.e. a technology which is wholly owned by an agent in this economy) in sector s spend the proceeds from their production activities on commodities other than their output and the non-produced good. Firms' (shareholders') preferences are given by

$$V(c_{-s}, m) = V(\Pi_{k \neq s}(c_k)^{\alpha} m^{\beta}) \tag{3}$$

where $V(\cdot)$ is increasing and concave. Since firms and workers in the same sector have identical ordinal preferences, firms' demands are given by (2) with income equal to the profit level per firm, π. Firms' attitudes towards risk are characterized by $V(\cdot)$.

The outsiders split their demands evenly across the S sectors and consume the non-produced good as well. Their preferences are Cobb–Douglas with demands given by

$$c_k = (\tau / p_k)M \text{ for } k = 1,2,\ldots,S \text{ and } m = (1 - S\tau)M \tag{4}$$

where M is the aggregate endowment of the outsiders.

Technology

The presence of imperfect competition in product markets derives from some element in the industry which prevents the free entry and exit of firms. These barriers to entry are certainly important in understanding the long-run behaviour of a particular industry. Here, the model is very short-run in nature so that these barriers will not be explicitly introduced into the analysis. That is, we will simply assume that there are F firms per sector and not model the source of this market power. An extension of this framework to study dynamics of entry and exit is provided in Chatterjee and Cooper (1989).

The technology will simply be a proportional relationship between output and labour input with one unit of labour producing Θ units of output. Variations in Θ, i.e. productivity shocks at the sectoral or aggregate level, can then be studied.

Assume that work sharing is not feasible so that workers are either employed or unemployed. This may reflect features of technology which make hours and people imperfect substitutes. This structure is imposed so that the model will generate employment fluctuations rather than variations in hours, and allows us to investigate the role of unemployment insurance. See the related discussion in Hansen (1985).

Optimal labour contracts

We focus on the contracting problem between an arbitrary firm f of sector s and a group of N workers who form a union. The structure of unions and the size of their membership is taken as exogenous for this exercise. Since the economy is symmetric, we focus on symmetric Nash equilibria. The contract devised for this firm and group of workers will therefore be optimal for all other union–firm pairs in a symmetric equilibrium.

The exogenous variables, which are unknown to workers and firms when contracts are negotiated, include: the endowments of the outsiders, (M), the preferences of the outsiders, (τ), the preferences of the workers and firms, (α, β), and technology, (θ). Denote by Γ the vector of these parameters describing the state of the economy. In addition to the variables included in Γ, payoffs to the worker–firm coalition depend on the *ex post* quantity decisions of other coalitions in the economy. These variables are conjectured, on a state-by-state basis, by the contractants when negotiating a contract *ex ante*. In equilibrium, these conjectures are correct. So a labour contract will be written contingent on the realization of the random variables in the economy, given conjectures about the *ex post* quantity decisions of other coalitions in the economy. All expectations used in describing the contracting problem will be relative to the distribution of Γ, which is common knowledge.

The optimal contract between firm f of sector s and that firm's N workers is characterized by three schedules: $w^e(\Gamma)$, $w^u(\Gamma)$, and $L(\Gamma)$. The first schedule is the state-contingent wage paid to employed workers, the second is the state-contingent level of severance pay and the third is the employment rule contingent on Γ as well. To simplify notation, denote a contract by δ and eliminate the superscripts and subscripts. The union–firm choose δ to maximize

$$E\{V(z\pi) + \sigma[(L/N)U(zw^e - r) + (N - L/N)U(zw^u)]\} \tag{5}$$

subj to: $L(\Gamma) \leq N$ and

w^e, w^u, L non-negative for all Γ.

In the objective function, σ represents the bargaining weight attached to workers' expected utility. The ratio L/N is the probability that a given worker is employed if L workers are employed under the optimal contract in some state. The firm's utility function depends on the product of the firm's profit π and a variable z (which is defined below and also appears in the workers' payoffs). The profit level is

$$\pi = R(L, \Gamma) - w^e L - w^u(N - L). \tag{6}$$

The function $R(L, \Gamma)$ is the firm's revenue if it employs L workers in state Γ.

The variable z appears in the payoffs of both the firm and the workers and is a price deflator for the market basket of goods generated by the Cobb–Douglas preferences. For later reference,

$$z=\beta^{\beta}\Pi_{k\neq s}(\alpha/p_k)^{\alpha}. \tag{7}$$

Assuming for the moment that the constraint on the employment level not exceeding N does not bind, the first order conditions for the problem can be summarized by

$$U'(w^e z-r)\sigma/N=U'(w^u z)\sigma/N=V'(z\pi) \text{ and} \tag{8}$$

$$R_L(L,\Gamma)=r/z \text{ for all } \Gamma \tag{9}$$

where R_L is the derivative of $R(L,\Gamma)$ with respect to L.[9]

(8) implies that risks are shared efficiently between the firm and its workers. Because of the strict concavity of $U(\cdot)$, (8) implies that workers are indifferent between the states of employment and unemployment, i.e. the level of severance pay compensates the unemployed worker for the lost wage income less the utility of leisure (r). As a consequence, workers' welfare is independent of employment status. Note though that as long as the firm is not risk-neutral, the utility level of workers will depend on the level of economic activity in the economy.

(9) represents the employment rule in the optimal contract. Since workers are indifferent between employment and unemployment, the only cost to the firm of employing an additional worker is the value of that worker's leisure in terms of the consumer basket, r/z. The gain to employing another worker is the marginal revenue gained from selling an extra unit of output. If the constraint that $L\leq N$ binds, then employment is set at N and (9) does not hold with equality.

The important element about (9) is that once Γ is determined, this expression determines the firm's action in the *ex post* product market game. The union and the firm act as a coalition in this product market game with their strategy determined in the *ex ante* labour contract. In contrast to Hart (1982), the employment decision reflects the *joint* market power of the union–firm coalition in product markets and not the labour market power of the union *per se*. That is, the employment rule is independent of σ. There are no pressures to force the economy to full employment: the optimal contract allows some workers to remain unemployed as a means of obtaining the largest surplus possible given the market power of the firm.

Of course, the level of wages does depend on σ as this parameter determines how the surplus from the market game will be shared. This separation of employment and compensation arises from the structure of preferences (i.e. no income effects) for the contractants.[10] The decisions of the union–firm in the product market game are thus equivalent to that of a firm facing a marginal cost of labour (in terms of the numéraire) of r/z.

(8) and (9) completely specify the employment, wage and severance pay schedules in an optimal contract between firm f in sector s and its N

workers. Recall that we have omitted any notation regarding the identity of this firm or its sector of operation. Given the symmetry of the problem, the conditions describing this optimal contract also characterize that for other union–firm pairs in other sectors. The only change that is necessary is to define z to correspond to the sector of interest. Given this optimal contract, we are able to compute the Nash equilibrium in the product market.

Goods market equilibrium

To complete our characterization of an equilibrium, we need to model the interaction of firms in product markets, *ex post*. We will concentrate on symmetric Nash equilibria for this economy. The labour contracts negotiated *ex ante* with their unions, specify an employment rule for a firm contingent on the realization of random variables in the economy. The interactions of these employment rules will then determine the equilibrium. The revenues earned by firm f of sector s are:

$$R(L_s^f, \Gamma) = p_s q_s^f = (E_s/Q_s) q_s^f \tag{10}$$

where $q_s^f = L_s^f \theta$ for all f and Q_s is the total output in sector s.

The Cobb–Douglas preferences imply that each commodity has a constant budget share for each of the consuming agents in this economy. Hence, the price in sector s is simply the total expenditure in that sector, E_s, divided by the total output in sector s, Q_s.

Denote by q_s the level of output for each firm in sector s if they are all producing at the same level. Since the expenditure on sector s is independent of q_s, we can solve for the symmetric partial equilibrium in sector s by finding that level of output per firm in that sector which is a best response to the output levels chosen by the other firms in that sector *given* the level of expenditure on that sector. To do so, use (10) to solve (9) yielding

$$R_L(L, \Gamma) = \theta E_s \eta / F q_s = r/z_s \tag{11}$$

where η is defined as $(1 - 1/F)$. Note that (11) holds only when the constraint that $L \leq N$ is not binding.[11] The variable η will serve as a useful measure of the market power of firms in the economy. As $F \to \infty$, $\eta \to 1$ while $\eta \to 0$ as the industry approaches monopoly. Recall that the sectors are symmetric so that η is not indexed by the sector. Using (11), the price in sector s is simply

$$p_s = r/(\eta \theta z_s). \tag{12}$$

It is important to note that z depends on prices in other sectors as well as parameters of preferences so that, in (11) and (12), we index it by s.

To characterize the symmetric Nash equilibrium for this economy, note that (11) and (12) hold for all s. Since z_s depends on the prices in other sectors, we can solve for the output price, common to all sectors, using (7). This will be

$$p = (r/\eta\theta)^{1/\beta}/\alpha^{(1-\beta)/\beta}\beta \tag{13}$$

where θ is a productivity shock (assumed for now to be) common across firms and sectors. Note that prices can be determined independent of output levels as long as $L \leq N$. This is a feature of homothetic preferences and constant marginal costs. This implies that variations in the endowments of outsiders will cause output but not prices to change.

To determine the output level of an arbitrary firm in sector s (given that all firms are producing the same level of output in the partial Nash equilibrium), we need to specify the level of expenditure on sector s. Using the Cobb–Douglas preferences, this is given by

$$E_s = \alpha F \Sigma_{k \neq s} p_k q_k + \tau M. \tag{14}$$

Again note that the summation here is over sectors other than s. Using (12), in a symmetric Nash equilibrium, (14) can be rewritten as

$$E_s = F_p[\alpha \Sigma_{k \neq s} q_k + \tau \Phi M] \tag{15}$$

where $\Phi = 1/Fp = \theta\eta z/Fr$.

This expression implies that the total expenditure on sector s depends on the level of economic activity in all the other sectors and is proportional to the endowment of the outsiders. This is similar to a simple model of income determination in which there are autonomous expenditures and an expenditure term which depends on the level of economic activity.

Inserting (15) into (11) implies

$$q_s = \alpha \Sigma_{k \neq s} q_k + \tau \Phi M = \alpha \Sigma_{k \neq s} q_k + \tau M/Fp. \tag{16}$$

So the per firm level of output in sector s is a linear function of the aggregate level of activity in other sectors and the per firm endowments of the outsiders (in terms of the produced commodity). For S large enough (so that each sector is small), (16) implies that the level of output per firm in sector s is a linear function of the aggregate output level in the economy.

Figure 4.2 displays a 'reaction curve' for sector s for a given level of activity in other sectors. Note that this is not literally a 'reaction curve' for a particular firm but rather an expression for the sector-specific Nash equilibrium output level for given output levels of other firms. The intercept of this 'reaction curve' is the sector-specific level of autonomous expenditures.

As the output levels in other sectors increase, all of the firms in sector s will expand as well. So even though the firms in a given sector are producing

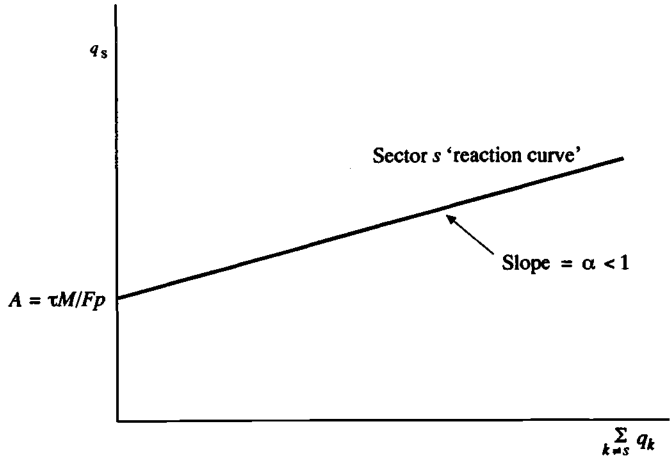


Figure 4.2 A reaction curve for sector s

perfect substitutes, positive interactions (strategic complementarities) are present between firms across sectors in a symmetric Nash equilibrium.

Since (16) holds for all s, we can use it to compute the output level per firm in each sector in a symmetric Nash equilibrium. This output level is given by

$$q = \frac{\tau M/Fp}{1-(S-1)\alpha} = A/\beta \tag{17}$$

where A is the level of autonomous expenditure on the produced good. Recall that β is the budget share of the numéraire good and represents the share of income that leaks out of the income stream connecting the firms in the economy.

To guarantee that this level of output is feasible, we need to check that $q \leq N\theta$ since each firm had available only N workers. This inequality is satisfied as long as M is not too large and β not too small. We will use (17) as an equation for the equilibrium level of output even though it contains the price in it. From (13), we know that the equilibrium price level is independent of output.

When the level of output given in (17) exceeds θN, then the economy is at full employment and the expressions derived in this section will not apply. Each firm will then simply employ all N workers and prices in each sector will adjust so that markets clear at full employment.

Finally, and perhaps most importantly, note that the equilibrium levels of output, employment and prices are actually independent of a number of labour market variables in this economy: the number of workers N and the bargaining weight σ. This independence of the equilibrium of the number of workers is also a property of many efficiency wage models (see, for example Weiss, 1980 and Yellen, 1984). If the level of per firm output satisfying (17) is less than N, we have an underemployment equilibrium. Furthermore, increasing the number of workers in this economy (assuming that they all join a union) will only increase the unemployment rate and will not alter the form of the optimal contract between the union and the firm. Again, the optimal contract specifies employment to obtain the maximal amount of payoffs for the coalition of workers and the firm, and this decision rule is independent of the number of workers N and the bargaining weight. This economy therefore has the property that there are no forces at work to reduce the rate of unemployment since the agents in the economy have no incentive to employ the excess labour force.

Overall, the equilibrium for this economy is quite similar to that described by Hart and others, although here the labour relations are the consequence of a privately optimal contract. Hart (1982) provides a discussion of multiplier effects and the presence of a balanced budget multiplier for his economy. Those types of comparative static results can be obtained for this economy as well.[12] (17) can be used directly to evaluate the effects of variations in M as long as the economy operates in an underemployment region. (13) and (17) can be used to evaluate the effects of changes in aggregate productivity.[13] Moreover, though this economy had a unique equilibrium, it is possible to construct economies with multiple equilibria, along the lines of Heller (1986) and/or Kiyotaki (1988), in which labour services are traded with labour contracts. We thus see that the main conclusions of the previous work on the implications of imperfect competition are robust to the introduction of privately optimal labour contracts.

Conclusions

The goal of this chapter has been to provide a framework of analysis for coordination failures in imperfectly competitive economies. In contrast to other studies this area, this chapter includes a representation of labour markets through a contracting framework. This approach provides some perspective on the manner in which adjustments in the labour market influence the operation of product markets. To the extent that labour contracts bind workers and firms into a coalition that seeks to gain its share of surplus, there are no forces at work in labour markets to move the economy towards an efficient outcome. The model can then be used in a straightforward manner to derive some comparative static properties of imperfectly competitive economies and to conduct some policy experiments.

This framework is intended as a basis for other research on the implications of macroeconomic imperfect competition.[14] One can amend the model to study a number of dynamic issues such as the role of entry and exit. Further, the contracting problem can be made more interesting by introducing asymmetric information, and the aggregate implications of this distortion can be investigated.[15] Finally, the framework may be useful in studying the interaction between the organization of labour relations (for example, the nature and extent of unionization) and aggregate behaviour.

Notes

This is a revised version of Cooper (1986). Much of this research was conducted while I was a Visiting Scholar at the Board of Governors of the Federal Reserve System. I am very grateful to that institution for its support. Seminar participants at the 1986 NBER Summer Institute, the University of Delaware, the University of Iowa, the University of Pennsylvania, the University of Western Ontario and the 1987 NBER Conference on Labor Markets and the Macroeconomy provided numerous helpful comments and suggestions. Dale Mortensen provided an excellent presentation of the original paper at the 1987 NBER Conference on Labor Markets and the Macroeconomy which helped considerably in this revision. Comments from Robin Boadway, Randy Wright, Jon Strand and Peter Howitt and two anonymous referees are also appreciated. Financial support from the National Science Foundation, NSF No. SES86-05302, is gratefully acknowledged. This chapter previously appeared in the *Canadian Journal of Economics* (1990), pp. 509–22, and we would like to thank Toronto University Press for their kind permission to reproduce it here.

1. Related results on multipliers in similar models are reported by Startz (1989) and Mankiw (1988). The term 'underemployment equilibrium' refers to the possibility that labour may not be fully employed in equilibrium. This is not necessarily an inefficient outcome as it could arise in an Arrow–Debreu economy as well. However, underemployment equilibria are more likely to arise in models with imperfect competition.
2. Cooper and John (1988) discusses the relation of these results to other coordination models associated with trading externalities (Diamond, 1982 and Howitt, 1985) and production externalities (Bryant, 1983).
3. Hart (1982) notes that these results are also dependent on the inelastic labour supply of workers. Dixon (1987, 1988) adopts a similar model of union behaviour.
4. For a review of that literature, see Cooper (1987) and Rosen (1986).
5. See, for example, Cooper and John (1988) and Heller (1986).
6. This implication of imperfect competition is exploited by Hall (1987) and Cooper and Haltiwanger (1990) as well.
7. In dynamic models of imperfect competition, this outside good is often treated as fiat money. See, for example, Chatterjee and Cooper (1989).

8. An earlier version of this chapter described the implications of relaxing these strong assumptions.
9. These conditions are obtained by differentiating (5) on a state-by-state basis since there are no terms involving the rate of change in the state Γ.
10. If there were income effects and employment was not $\{0,1\}$, then market power would influence the level of employment through the disutility of work. In this case, the results would be quite sensitive to the specification of preferences. In addition, this specification implies that workers are indifferent between being employed and unemployed.
11. When $L = N$, the economy is at full employment and prices must vary to clear markets.
12. See Cooper (1986) for a characterization of these comparative statics.
13. Sector-specific productivity shocks can be shown to lead to economy wide fluctuations through changes in relative prices, see Cooper (1986).
14. Cooper (1986) spells out these applications in more detail.
15. See Kahn and Mookerjee (1988) for work in this direction.

5 Market power, coordination failures and endogenous fluctuations

Claude d'Aspremont, Rodolphe Dos Santos Ferreira and Louis-André Gérard-Varet

Introduction

Imperfect competition has become one of the main ingredients of the so-called 'New Keynesian economics'.[1] Much recent work has emphasized Keynesian features of imperfectly competitive macroeconomic models,[2] belonging either to the Cournotian tradition, as in Hart (1982), or to one of the monopolistic competition brands, corresponding to the Chamberlinian framework of Dixit and Stiglitz (1977) (Weitzman, 1985; Blanchard and Kiyotaki, 1987), or to Hotelling's spatial model (Weitzman, 1982). A somewhat paradoxical aspect of part of this literature is that it actually appears quite unkeynesian, not so much because Keynes himself paid little attention to imperfect competition, but essentially because some of its results, concerning in particular unemployment or the response to aggregate demand shocks, depend either on union power,[3] or on small adjustment costs ('menu costs')[4] and 'nearly rational' behaviour of price-setting agents,[5] and thus in a sense come back to 'classical' explanations, in terms of wage and price rigidities or frictional phenomena.

We do not want to deny the relevance of such explanations. We propose to explore in this chapter only some consequences of firms' monopoly power, in a context of full rationality, complete price flexibility, without adjustment costs and in the absence of exogenous shocks. Since we conform to the Keynesian priority given to the product market relatively to the labour market, and since it is the output market power that we want to emphasize, we shall (unrealistically) assume perfect competition in a homogeneous labour market. And, as an attempt to unify the theoretical framework, facilitating comparisons, we shall use a model of 'Cournotian monopolistic competition',[6] which generalizes and extends both the Cournotian and the Chamberlinian approaches to imperfect competition. The model involves several goods, each produced by several firms, and includes pure Cournot oligopoly and pure monopolistic competition as particular cases. When making a decision, the producer of a good is supposed to take as given the total quantity of that good produced by other firms and the prices of the other goods, while knowing the 'objective' demand, that is, acting on the basis of rational conjectures about consumers' behaviour.

The model further extends early imperfectly competitive models by assuming overlapping generations of consumers. This is of course a convenient step towards the dynamic analysis of intertemporal equilibria, involving in particular the study of the conditions in which endogenous fluctuations occur. Besides, coexistence of two generations of young and old consumers allows us to incorporate not only individual income and substitution effects on the demand elasticity, and hence on the firms' market power (and the case where the former dominates the latter, the complementarity case, will turn out to be interesting), but also redistributional effects, between the young and the old. Such effects may account for a significant variability of such market power, induced by changes in demand elasticity as the output price varies.

It is this variability, more than the existence of a strong market power in itself, that drives interesting effects.[7] Indeed, alone or in conjunction with decreasing marginal cost[8] (compatible with equilibrium because of imperfect competition), mark-up variability may be responsible for a reversal of the slope of the adequately defined 'oligopolistic labour demand' curve. This reversal, by having that curve cut the labour supply curve more than once, can naturally lead to a multiplicity of temporary equilibria in this model where, on the contrary, a unique equilibrium would have resulted from non-increasing returns together with weak mark-up variability (as under nearly competitive conditions or under substitutability). Multiple temporary equilibria, corresponding to different employment levels, may even be Pareto-ranked, giving rise to a situation of 'coordination failure'[9] which cannot occur under perfect competition (as we well know from the First Fundamental Theorem of Welfare Economics).

The same factors, increasing returns and variability of market power, that account for a multiplicity of temporary equilibria, also favour, through their influence on the form and movements of the oligopolistic labour demand curve, emergence of endogenous fluctuations. Given enough complementarity, such fluctuations can occur in a perfectly competitive world.[10] But we will show that if, in an extreme form of market power variability, marginal revenue becomes negative[11] when prices are not expected to change much in the next period (a condition which excludes the existence of a stationary equilibrium), and if returns are increasing (a condition that further excludes in our model the existence of a monotonic equilibrium), then *all* equilibria are characterized by persistent endogenous fluctuations in prices and output.

The chapter is organized as follows. In the second section we describe the model of Cournotian monopolistic competition and define the equilibrium concepts that we will use. The third section exhibits a multiplicity of Pareto ranked (temporary or stationary) equilibria, under increasing returns or under significantly variable market power. More complex intertemporal equilibria, generating endogenous fluctuations, are analysed in the fourth section. We then conclude.

The model

We use an overlapping-generations model in discrete time. In each period $t=0,1,2,\ldots$, there are $m+2$ commodities: m produced goods ($m \geq 2$), labour and money, which is both the unit of account and the only storable good. The economy is endowed with a constant stock of money M.

We consider two types of agents: consumers and firms. Each consumer lives for two periods, works and earns income when young, but consumes the produced goods both when old and when young. At each period and for every good $k=1,\ldots,m$, there are n firms producing good k. The labour market is assumed competitive. However, in the produced good markets we assume *Cournotian monopolistic competition*, in the sense that each producer of good k maximizes his profit, facing the 'true' demand for good k, net of the total quantity produced by the others and given the prices of the other produced goods.[12] In this section we shall first derive the demand functions from consumers' behaviour, then present the firm's problem and introduce both a temporary and an intertemporal concept of Cournotian monopolistic competition equilibrium. Finally we compare our model to others in the literature.

Consumers' behaviour and demand functions

In each period $t=-1,0,1,\ldots$, a continuum of consumers of mass L is born. Each individual lives for two periods and receives, while young, an income (profit or wage). Preferences of young consumers, identical for all individuals, are homothetic and separable, represented by the function

$$U[u(c), u(\hat{c})] - vl, \tag{1}$$

where $c \in \mathscr{R}_+^m$ and $\hat{c} \in \mathscr{R}_+^m$ denote respectively present and future consumption baskets, $l \in \{0,1\}$ is an (indivisible) labour supply and v is the (non-negative) disutility of work. The function U is twice-differentiable, strongly quasi-concave, increasing, homogeneous of degree one, and such that the indifference curves do not cut the axes. The function u is a symmetric CES function. We assume that the preferences of old consumers are represented by this same function u.

Consumers know and take as given the current prices $p \in \mathscr{R}^m_{++}$ and, when young and hence in the labour market, the current wage $w \in \mathscr{R}^m_+$. Also, young consumers are assumed to have identical rigid point expectations about future prices $\hat{p} \in \mathscr{R}^m_{++}$.

Separability of preferences entails the following convenient procedure for utility maximization of the young consumer:

1. Optimal consumption baskets are first chosen for each period, assuming given expenditure levels. Let

$$u(c) = m\left(\frac{1}{m}\sum_{k=1}^{m} c_k^{(s-1)/s}\right)^{s/(s-1)} \equiv C, \tag{2}$$

where $s \in \mathscr{R}_{++}\backslash\{1\}$ is the elasticity of intersectoral substitution.[13] Notice that the CES function u can be seen as defining an index for total consumption C, since the expression for C/m in (2) is just a mean of the quantities in the consumption basket. If the mean price is defined by the formula

$$P \equiv P(p) \equiv \left(\frac{1}{m}\sum_{k=1}^{m} p_k^{1-s}\right)^{1/(1-s)}, \tag{3}$$

then c_k^*, the kth element of the optimal consumption basket for an expenditure x in the first period of consumer's life, can be expressed as

$$c_k^* = \frac{1}{m}\left(\frac{p_k}{P}\right)^{-s}\frac{x}{P}, \tag{4}$$

so that $\sum_k p_k c_k^* = PC = x$. Clearly, similar expressions hold for the optimal consumption of an old consumer as well as for $\hat{c}_k^*$, the future consumption of a young consumer (as a function of the future price $\hat{p}_k$ and the future mean price $\hat{P}$).

2. An optimal intertemporal consumption decision is then taken, given a total income i. By homogeneity of the function U, we get

$$C^* = h(P/\hat{P})i/\hat{P} \text{ and } \hat{C}^* = [1 - (P/\hat{P})h(P/\hat{P})]i/\hat{P}, \tag{5}$$

where h is a twice-differentiable, decreasing function. We shall henceforward denote:

$\theta \equiv P/\hat{P}$ the real interest factor, and

$\alpha(\theta) \equiv \theta h(\theta)$ the propensity to consume.

Notice that $0 < \alpha(\theta) < 1$, since the indifference curves of U do not cut the axes. The optimal intertemporal consumption structure

$$\frac{C^*}{\hat{C}^*} = \frac{h(\theta)}{1-\theta h(\theta)} \equiv \frac{\alpha(\theta)}{\theta} \frac{1}{1-\alpha(\theta)} \equiv H(\theta) \tag{6}$$

is a twice-differentiable, decreasing function of the real interest factor. We shall denote:

$\sigma(\theta) \equiv -H'(\theta)\theta/H(\theta)$ the elasticity of intertemporal substitution.

A simple calculation shows that $\sigma(\theta) = 1 - (\alpha'(\theta)\theta)/(\alpha(\theta)[1-\alpha(\theta)])$, so that α is increasing (resp. decreasing) in the case of intertemporal complementarity, i.e. if $\sigma < 1$ (resp. substitutability, i.e. if $\sigma > 1$). The utility derived from optimal consumption

$$U(C^*, \hat{C}^*) = U(\alpha(\theta), \theta[1-\alpha(\theta)])i/P \equiv U^*(\theta)i/P \tag{7}$$

is linear in the real income i/P and increasing in the real interest factor θ. Indeed, using the first order condition $U_1' = U_2'\theta$, and Euler's identity $U = U_1'\alpha + U_2'\theta(1-\alpha)$, we have:

$$U^{*\prime}\theta = (U_1' - U_2'\theta)\alpha'\theta + U_2'\theta(1-\alpha) = U^*(1-\alpha) > 0. \tag{8}$$

3. Finally, the consumer chooses to work (resp. not to work) if $U^*(\theta)w/P - v > 0$ (resp. < 0). Indifference prevails in the case of an equality. Thus,

$$\frac{w}{P} = \frac{v}{U^*(\theta)} \tag{9}$$

is the real reservation wage, equal to the ratio of labour disutility to utility of optimal consumption.

Since demand is the same linear function of income for all young consumers, income distribution is immaterial (notice that we do not assume that shares of firms are uniformly distributed), so that aggregate demand by the young is a linear function of aggregate income I. In the same way, the distribution of cash balances among the old is also immaterial, and their aggregate demand is a linear function of the money stock M. More precisely, first notice that the aggregate expenditure X is given by the sum of the money stock, which is completely spent by the old, and of the aggregate income multiplied by the propensity to consume, corresponding to what the young spend:

$$X = \alpha(P/\hat{P})I + M. \tag{10}$$

Thus, as $I = X$ in equilibrium, we obtain the following expression for the aggregate demand:

$$\frac{X}{P} = \frac{1}{1-\alpha(P/\hat{P})} \frac{M}{P} \equiv D(P, \hat{P}), \tag{11}$$

which is the familiar product of the Keynesian multiplier and the autonomous demand. Since demand has the same structure for all consumers, whether young or old, given by (4), we then get the sectoral demand for good k:

$$d(p_k, P, \hat{P}) = \frac{1}{m}\left(\frac{p_k}{P}\right)^{-s} D(P, \hat{P}). \tag{12}$$

By symmetry, the function d is the same for all produced goods, so that an index k is not required.

If we now refer to the elasticity of aggregate demand $D(\cdot, \hat{P})$:

$$-\frac{\partial D}{\partial P}\frac{P}{D} = \alpha\left(\frac{P}{\hat{P}}\right)\sigma\left(\frac{P}{\hat{P}}\right) + 1 - \alpha\left(\frac{P}{\hat{P}}\right) \equiv \Delta\left(\frac{P}{\hat{P}}\right), \tag{13}$$

and using:

$$\frac{\partial P}{\partial p_k}\frac{p_k}{P} = \frac{1}{m}\left(\frac{p_k}{P}\right)^{1-s}, \tag{14}$$

we can easily compute the elasticity of sectoral demand for good k:

$$-\left(\frac{d}{dp_k} d(p_k, P, \hat{P})\right)\left(\frac{p_k}{d(p_k, P, \hat{P})}\right)$$
$$= \left[1 - \frac{1}{m}\left(\frac{p_k}{P}\right)^{1-s}\right] s + \frac{1}{m}\left(\frac{p_k}{P}\right)^{1-s} \Delta\left(\frac{P}{\hat{P}}\right) \equiv \delta\left(\frac{p_k}{P}, \frac{P}{\hat{P}}\right). \tag{15}$$

Hence, the sectoral demand elasticity δ appears as a weighted average of the elasticity of intersectoral substitution s and of the aggregate demand elasticity Δ, itself a weighted average of the elasticity of intertemporal substitution σ and of unity (corresponding to the spending behaviour of the young and the old, respectively). It is necessarily positive.

Producers' behaviour

Every good $k = 1, \ldots, m$ is produced by the same number n of firms ($n \geq 1$), with identical increasing, but not necessarily convex, isoelastic cost functions: y^{β} (with $\beta > 0$), giving the labour quantity required by output y.

Each producer j of good k takes as given the wage w in the labour market, while acting as a Cournot oligopolist in the market for the produced good k, on the basis of the 'true' residual demand for his output, contingent on the supply of the other firms in the same sector:

$$d(p_k, P(p_k, p_{-k}), \hat{P}) - Y^k_{-j}, \text{ with } Y^k_{-j} \equiv \Sigma_{i \neq j} y^k_i.$$

We assume that producer j knows the common mean price expectation $\hat{P}$, and conjectures (correctly at equilibrium) the vector p_{-k} of other commodity prices and the output Y_{-j}^k of good k produced by the other firms. These conjectures are essential for our notion of Cournotian monopolistic competition. The function $d(.,P(.,p_{-k}),\hat{P})$ being strictly decreasing in p_k, we may introduce the inverse demand function ψ for the output of firm j, which is:

$$p_k=\psi(y_j+Y_{-j}^k,p_{-k},\hat{P}) \Leftrightarrow y_j=d(p_k,P(p_k,p_{-k}),\hat{P})-Y_{-j}^k. \tag{16}$$

Then, for any given $\bar{Y},\bar{p},\hat{P}$ and w, the typical 'producer's problem' is the following profit maximization problem:

$$\max_{y\in\mathscr{R}_+} \pi(y)\equiv\max_{y\in\mathscr{R}_+} \psi(y+\bar{Y},\bar{p},\hat{P})y-wy^\beta \tag{17}$$

(for simplicity of notation, we omit the indices j and k). An optimal positive output y must satisfy the first order condition (with $\psi(y)$ standing for $\psi(y+\bar{Y},\bar{p},\hat{P})$):

$$\psi(y)\left[1+\frac{\psi'(y)y}{\psi(y)}\right]=w\beta y^{\beta-1}, \tag{18}$$

which is the familiar equality of marginal revenue and marginal cost. It must also entail a non-negative profit: $\psi(y)y\geq wy^\beta$ (otherwise it would be dominated by a zero output), so that (using (18)):

$$\varepsilon(y)\equiv-\frac{\psi'(y)y}{\psi(y)}\geq 1-\beta. \tag{19}$$

Lemma 1 and Lemma 2 in the appendix introduce restrictions (on the function ε and on the elasticity δ of sectoral demand, respectively), ensuring that the ratio of marginal revenue to marginal cost is decreasing *whenever* condition (19) is verified and the marginal revenue is non-negative. Then conditions (18) and (19) are sufficient for a *global* maximum of the profit function (unique in the positive domain), although that function may fail to be quasi-concave. Based on these lemmata, the following propositions give simple assumptions on s and σ implying that the first order condition together with the non-negative profit condition are necessary and sufficient for profit maximization at a positive level of output. We stick to two typical cases: overall (both intertemporal and intersectoral) complementarity and overall substitutability.

Proposition 1 (complementarity): Assume intersectoral complementarity: $s\leq 1$. Also, assume that the elasticity of intertemporal substitution σ satisfies, for any $\theta\in\mathscr{R}_{++}$,

$$\sigma(\theta) < 1 \text{ and } \frac{\sigma'(\theta)\theta}{\sigma(\theta)} \geq -[1-\sigma(\theta)] \tag{20}$$

or, equivalently, that the intertemporal expenditure structure $\theta H(\theta)$ is an increasing and concave function of θ. Then, the first order and non-negative value condition:

$$p\beta \geq p\left[1 - \frac{y/(y+\bar{Y})}{\delta\big(p/P(p,\bar{p}), P(p,\bar{p})/\hat{P}\big)}\right] = w\beta y^{\beta-1}, \tag{21}$$

where $p = \psi(y+\bar{Y}, \bar{p}, \hat{P})$ as defined in (16), is necessary and sufficient for y to be the unique positive solution to the producer's problem (17).

Proof: See appendix.

Proposition 2 (substitutability): Assume intersectoral substitutability: $s \geq 1$. Also, assume that the elasticity of intertemporal substitution σ satisfies, for any $\theta \in \mathcal{R}_{++}$,

$$1 \leq \sigma(\theta) \leq \min\{s, 1+\beta\} \text{ and } \sigma'(\theta) \geq 0. \tag{22}$$

Then, the first order and non-negative value condition (21) is necessary and sufficient for y to be the unique positive solution to the producer's problem (17) ($\bar{Y} > 0$ and hence $n > 1$, is required in condition (21) if $s = 1$).

Proof: See appendix.

Temporary and intertemporal Cournotian monopolistic competition equilibria

Two concepts of equilibrium are successively considered: temporary and intertemporal equilibrium. Temporary equilibrium imposes compatibility of optimal consumers' and producers' decisions in the particular period which is considered, but price expectations remain exogenous and are arbitrarily given:

Definition 1 (temporary equilibrium): A temporary equilibrium associated with the expected mean price $\hat{P}$ is a vector with components $y^*_{kj} \geq 0$ (the outputs chosen by each firm $j = 1, \ldots, n$ in each sector $k = 1, \ldots, m$), $p^*_k > 0$ (the product prices, with $k = 1, \ldots, m$), and $w^* \geq 0$ (the wage), such that:

1. for any firm j in any sector k,

$$y^*_{kj} \in \arg\max_{y \in \mathcal{R}_+} \psi\left(y + \sum_{i \neq j} y^*_{ki}, p^*_{-k}, \hat{P}\right) y - w^* y^{\beta},$$

with the sectoral inverse demand function ψ as defined in (16),

2. for any sector k, $\sum_j y^*_{kj} = d(p^*_k, P^*, \hat{P})$, with $P^* = P(p^*_k, p^*_{-k})$,
3. $\sum_k \sum_j y^{*\beta}_{kj} = L$ and $w^*/P \geq v/U^*(P^*/\hat{P})$ (full employment) or $\sum_k \sum_j y^{*\beta}_{kj} < L$ and $w^*/P = v/U^*(P^*/\hat{P})$ (underemployment).

Conditions 1 and 2 characterize a Cournotian monopolistic competition equilibrium in the product markets: every producer maximizes his profit, knowing the 'true' consumers' demand function and their expectation of the mean price in the next period, acting as a Cournot oligopolist in his own market, and taking as given the wage and the other products prices. Condition 3 defines the competitive equilibrium of the labour market by the equality of supply and demand, but taking into account the possibility that only part of the labour force is employed. In this case, since we assume perfect wage flexibility and no rationing, the wage must be equal to its reservation value.

It can be checked that all equilibria are necessarily symmetric in quantities, inside each sector, relatively to active firms (see Lemma 3 in the appendix). But the possibility that only a number $n^* < n$ of firms be active in equilibrium cannot be excluded when returns are increasing ($\beta < 1$). Also, symmetry is not always warranted in prices, and correspondingly in quantities across sectors, unless we strengthen the assumptions hitherto formulated.[14] For simplicity we shall however simply ignore asymmetric equilibria, and the kind of multiplicity they may introduce.

It will be appropriate to refer to a symmetric temporary equilibrium by a triplet (y^*, p^*, w^*) characterized, under the assumptions of either Proposition 1 or Proposition 2, by:

1. the first order and non-negative value condition:

$$p^*\beta \geq p^*\left[1 - \frac{1/n}{\delta(1, p^*/\hat{P})}\right] = w^*\beta y^{*\beta-1}, \tag{23}$$

with δ as given by (15) and (13),

2. the product market equilibrium condition:

$$mny^* = \frac{1}{1-\alpha(p^*/\hat{P})}\frac{M}{p^*}, \tag{24}$$

3. and the labour market equilibrium condition:

$$\begin{aligned} &mny^{*\beta} = L \text{ and } w^*/p^* \geq v/U^*(p^*/\hat{P}) \text{ (full employment)} \\ &\text{or } mny^{*\beta} < L \text{ and } w^*/p^* = v/U^*(p^*/\hat{P}) \text{ (underemployment)} \end{aligned} \tag{25}$$

The characterization of symmetric temporary equilibria can be further simplified by referring to just the equilibrium price p^*, which may be determined using workers' and firms' reservation wages. Let us first define the symmetric marginal revenue function

$$\rho(p;\hat{P}) \equiv p\left[1 - \frac{1/n}{\delta(1, p/\hat{P})}\right], \tag{26}$$

such that $1 - \rho(p;\hat{P})/p = 1/n\delta(1, p/\hat{P})$ is Lerner's index of monopoly power. Then the firms' reservation wage[15] is constructed from the first order condition (23) and is equal to the product of the marginal productivity by the marginal revenue. In a symmetric product market equilibrium, using condition (24), this is simply the following function of the price p (conditional on a given price expectation $\hat{P}$):

$$\omega_f(p;\hat{P}) = \frac{1}{\beta}\left(\frac{D(p,\hat{P})}{mn}\right)^{1-\beta} \rho(p;\hat{P}). \tag{27}$$

The function ω_f has elasticity:

$$\frac{\omega_f'}{\omega_f} p = (\beta - 1)\Delta(p/\hat{P}) + \frac{\rho'(p;\hat{P})}{\rho(p;\hat{P})} p. \tag{28}$$

Similarly the workers' (nominal) reservation wage is constructed from (25) as a function of the same variables:

$$\omega_l(p;\hat{P}) = \frac{v}{U^*(p/\hat{P})} p, \tag{29}$$

with elasticity (for $v > 0$, and using (8)):

$$\frac{\omega_l'}{\omega_l} p = \alpha(p/\hat{P}). \tag{30}$$

The following proposition summarizes all the simplifications coming from symmetry:

Proposition 3 (symmetric temporary equilibria): An underemployment symmetric equilibrium is determined by equalizing both reservation wages, and verifying that a solution p^* to the equation $\omega_f(p;\hat{P}) = \omega_l(p;\hat{P})$ satisfies:

$$L > (mn)^{1-\beta} D(p^*, \hat{P})^\beta \text{ and } \beta p^* \geq \rho(p^*;\hat{P}). \tag{31}$$

A full employment symmetric equilibrium is given by the solution p^{**} to the equation $L = (mn)^{1-\beta} D(p^{**}, \hat{P})^\beta$, whenever this solution satisfies:

$$\omega_f(p^{**};\hat{P}) \geq \omega_l(p^{**};\hat{P}) \text{ and } \beta p^{**} \geq \rho(p^{**};\hat{P}). \tag{32}$$

We finally give the definition of an intertemporal equilibrium which, in addition to the compatibility of optimal consumers' and producers' decisions in every period, imposes the fulfilment of price expectations.

Definition 2 (intertemporal equilibrium): An intertemporal equilibrium is a sequence of temporary equilibria, one for each period $t=0,1,2,\ldots$, associated with an expected mean price $\hat{P}_t$, characterized by a mean price P_t^* and such that, for any $t>0$, $P_t^*=\hat{P}_{t-1}$.

Clearly, as we restrict ourselves to symmetric equilibria, an intertemporal equilibrium is completely described by a sequence of triplets $(y_t^*, p_t^*, w_t^*)_{t\in\mathcal{N}}$. As suggested by the simplified characterization of symmetric temporary equilibrium given in Proposition 3, a symmetric intertemporal equilibrium can also be simply identified to a sequence of positive prices $(p_t^*)_{t\in\mathcal{N}}$.

Comparison with the literature

It is useful at this stage to compare our model, and its associate temporary equilibrium concept, with the two main classes of imperfectly competitive macroeconomic models: the one introduced by Hart (1982) and relying on a Cournotian approach, and the one following the Dixit and Stiglitz (1977) Chamberlinian framework, and popularized by Weitzman (1985) and Blanchard and Kiyotaki (1987).[16] Some important differences immediately appear. Contrary to ours, these models attach importance to labour market imperfections,[17] keep intertemporal decisions implicit,[18] and (with the exception of Weitzman, 1985) neglect the so-called Ford effects[19] that lead producers to refer to a demand function incorporating a multiplier, such as the one given by (11). Also they do not exhibit an overlapping-generations structure, and hence ignore in particular the related distributional effects on demand.[20] It is however easy to capture the meaningful features of these two classes of models, at least from the point of view of product markets, by associating them with two special cases of our own model, assuming a unitary elasticity of, respectively, intersectoral and intertemporal substitution.

Unitary elasticity of intersectoral substitution or pure Cournot oligopoly

Hart (1982) uses a single produced good and assumes that consumers only spend in one sector. The significant point is however that cross-price effects do not play any role in his model. Such is also the case when consumers spend a constant sum in each sector, as a result of a Cobb–Douglas intersectoral subutility function u, with unit elasticity. Indeed, as one can easily check from (15), with $s = 1$ the sectoral demand elasticity has the real interest factor $\theta = P/\hat{P}$ as its sole argument, and becomes independent of the relative price p/P:

$$\delta = 1 - \frac{1}{m} + \frac{1}{m} \Delta(\theta) = 1 + \frac{1}{m} \alpha(\theta)[\sigma(\theta) - 1]. \tag{33}$$

Notice that $\delta > 1 - 1/m \geq 1/n$ (for $n > 1$, and recalling that $m \geq 2$, by assumption), so that the marginal revenue is necessarily positive, as *assumed* by Hart. As it already appears from the analysis in Heller (1986), this is a crucial restriction, making in particular less likely the emergence of multiple Pareto ranked equilibria (as we shall see on pp.106ff.). Such restriction also excludes the cases of economies in which *all* intertemporal equilibria are inflationary and converge to the autarkical state (d'Aspremont *et al.*, 1991a) or are non-monotonic in prices and output (pp.123ff.).

Unitary elasticity of intertemporal substitution and pure monopolistic competition

Comparison with pure monopolistic competition models, in the line of Dixit and Stiglitz (1977), supposes a single firm in each sector: $n = 1$. Also, neglecting, as these authors and their followers systematically do, any indirect individual price effects (through the mean price), amounts to assuming an infinite number of sectors ($m = \infty$), leading to a sectoral demand elasticity equal to the elasticity of intersectoral substitution: $\delta = s$. Indeed, we know that the mean price P is a function of p_k, the individual price of good k, given by (3). The elasticity of this function, given by (14), becomes negligible if m is very large,[21] which is precisely the reason put forward by Dixit and Stiglitz for taking P as (approximately) constant when determining the demand elasticity for good k (see d'Aspremont *et al.*, 1994). If we look beyond the simplifying role of such an assumption, we can see that it leads to the important consequence that the sectoral demand elasticity is constant and larger than one (otherwise no equilibrium would exist), thus excluding again the possibility of a negative marginal revenue and, more fundamentally, ruling out any change in market power.

This limitation remains, even without resorting to an infinite number of sectors (and to a sole producer of each good), as long as we assume, with Weitzman (1985) and Blanchard and Kiyotaki (1987), but contrary to Dixit and Stiglitz (1977), that the intertemporal utility function U is restricted to being of the Cobb–Douglas type, *i.e.* that the corresponding elasticity of substitution is equal to one. Indeed, with $\sigma(\theta) \equiv 1$, the sectoral demand elasticity becomes a weighted average of s and 1:

$$\delta = \left[1 - \frac{1}{m} \left(\frac{p}{P} \right)^{1-s} \right] s + \frac{1}{m} \left(\frac{p}{P} \right)^{1-s} \tag{34}$$

With symmetry in prices, this elasticity is equal to $(1 - 1/m)s + 1/m$, a constant, necessarily larger than $1/n$ for an equilibrium to exist. As a consequence, the symmetric marginal revenue is always positive and, more significantly, market power (or mark-up over marginal cost) is constant, so that both multiple temporary equilibria (pp.106ff.) and endogenous fluctuations (pp.106ff.) are less likely to occur.

Multiplicity of Pareto ranked equilibria

We first examine the static properties of our model and study how imperfect competition may contribute to the emergence of multiple Pareto ranked equilibria. These static properties will be useful both to analyse temporary equilibria and, under stationarity, intertemporal equilibria.

Temporary equilibria

We have seen that underemployment equilibria are determined by equalizing the firms' and the workers' reservation wage functions, ω_f and ω_l respectively, and that a full employment equilibrium requires the value of ω_f at the full employment price p^{**} to be at least as large as the value of ω_l. This is illustrated in figure 5.1. Existence of at most one equilibrium is thus clearly ensured if the graph of ω_f can only cut the graph of ω_l from below.

That condition is satisfied if (using (28) and (30))

$$\frac{\omega_f'}{\omega_f} p = (\beta - 1)\Delta + \frac{\rho'}{\rho} p > \alpha = \frac{\omega_l'}{\omega_l} p \tag{35}$$

for any price higher than the full employment one. In particular this is afforded by the assumption of non-increasing returns ($\beta \geq 1$), together with the restriction: $\rho' p/\rho > \alpha$, for any admissible p. Multiple temporary equilibria can thus never arise in our model when competition is perfect (requiring both non-increasing returns and equality of price and marginal revenue, so that $\rho' p/\rho = 1 > \alpha$). Multiple equilibria are also impossible in the models of Weitzman (1985) and Blanchard and Kiyotaki (1987), once we exclude fixed costs and 'menu' costs, since the demand elasticity is constant, entailing as in the pure competition case a unitary elasticity of ρ, the marginal revenue function. More generally, non-increasing returns rule out

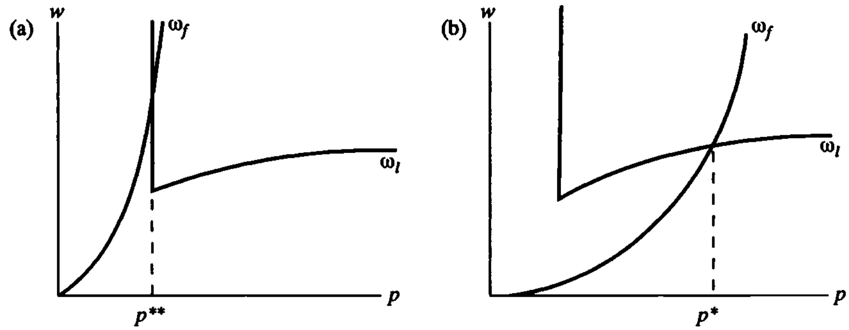


Figure 5.1 Unique equilibria
(a) Full employment (b) Underemployment

equilibrium multiplicity in the substitutability case,[22] where (as stated in Lemma 4 in the appendix) $\rho' p/\rho$ is always larger than α:

Proposition 4 (uniqueness): In the substitutability case, with non-increasing returns, and under the assumptions of Proposition 2, there is at most one symmetric temporary equilibrium.

There are however cases of multiple equilibria. The proposition suggests that we consider the case of increasing returns and the case of complementarity.

Increasing returns

That increasing returns are a possible source of multiplicity of symmetric temporary equilibria should be obvious from the preceding discussion. In order to focus on the specific role of increasing returns, let us neglect changes in market power, by resorting to the quite simple case of a Cobb–Douglas intertemporal utility function (leading to $\Delta = 1$ and $\rho' p/\rho = 1$). We assume that the (constant) propensity to consume α is larger than the elasticity of the cost function β, so that

$$\frac{\omega_f'}{\omega_f} p = \beta < \alpha = \frac{\omega_l'}{\omega_l} p. \tag{36}$$

The equality of the firms' and workers' reservation wages $\omega_f = Bp^{\beta} = Ap^{\alpha} = \omega_l$ (where A and B denote constant factors[23]) is thus satisfied at a unique price p^*, characterizing an underemployment equilibrium if conditions (31) are satisfied, that is:

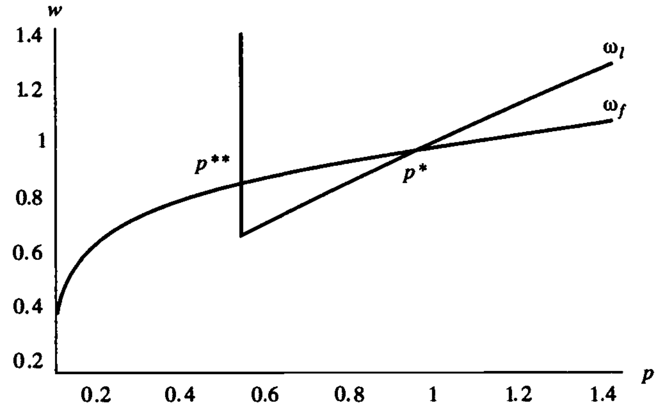


Figure 5.2 Coexisting full- and underemployment equilibria

$$p^* = \left(\frac{B}{A}\right)^{1/(\alpha-\beta)} > (mn)^{1/\beta - 1}\frac{M/L^{1/\beta}}{1-\alpha} = p^{**} \text{ and } \beta \geq 1 - \frac{m/n}{(m-1)s+1}. \quad (37)$$

In fact, as shown in figure 5.2, when these conditions hold, there is, besides the underemployment equilibrium, a full employment equilibrium corresponding to a price p^{**} (we have $Bp^{**\beta} > Ap^{**\alpha}$, so that conditions (32) are also verified).[24] Otherwise, except in the singular case $p^* = p^{**}$, or in the case $v = 0$ (implying $A = 0$), where the full employment equilibrium is unique, no equilibrium exists.

Multiplicity of equilibria does not in general imply that the equilibria are Pareto ranked. An old consumer obviously prefers the lowest price, here the full employment price p^{**}. A young consumer, as a worker, is unable to choose among underemployment equilibria: in any such equilibrium, the utility he derives from working is $U^*(p/\hat{P})\, w/p - v = 0$, which is the utility he gets when not working and not receiving any non-wage income. On the other hand, he definitely prefers, as a worker, a full employment equilibrium, if the wage is higher than its reservation value. The difficulty comes from his position as a shareholder, receiving a given fraction of total profits Π. So, in order to assert that p^{**} Pareto dominates p^*, one has to check whether

$$U^*\left(\frac{p^*}{\hat{P}}\right)\frac{\Pi^*}{p^*} \leq U^*\left(\frac{p^{**}}{\hat{P}}\right)\frac{\Pi^{**}}{p^{**}}. \quad (38)$$

Since $U^*(p/\hat{P})/p$ is decreasing (by (29) and (30), it has elasticity $-\alpha$), this inequality holds if Π is a non-increasing function of p, which is true in this example, since $pD(p,\hat{P})=M/(1-\alpha)$, so that $\Pi=pD(p,\hat{P})-(mn)^{1-\beta}Bp^{\beta}D(p,\hat{P})^{\beta}$ is in fact constant.

Multiple equilibria are thus Pareto ranked in this simple example with increasing returns and constant market power, giving rise to a *coordination failure*.[25]

Complementarity and significant variability of market power

We will now consider the complementarity case, putting aside increasing returns. For simplicity, we stick again to a straightforward class of examples,[26] assuming that the symmetric marginal revenue function ρ is positive over some price range but, as the price increases, eventually becomes negative since:[27]

$$\begin{aligned}\lim_{\theta\to\infty}\delta(1,\theta)&=(1-1/m)s+(1/m)\lim_{\theta\to\infty}\sigma(\theta)<1/n\\&<(1-1/m)s+1/m=\lim_{\theta\to 0}\delta(1,\theta).\end{aligned}\tag{39}$$

An example is given by a CES function U, with $\sigma<m/n-(m-1)s<1$.

In this class of examples, because of intertemporal complementarity (dominance of the income effect over the substitution effect), a higher price leads to a higher expenditure by the young consumers, thus increasing the weight of the generation with the lowest demand elasticity. So market power increases with price, and the firms' reservation wage correspondingly decreases (if marginal productivity does not increase too sharply with diminishing output). Since the firms' reservation wage function eventually becomes negative, it must indeed be decreasing for high prices. But it is increasing for low prices since, by (27), it converges to zero, as the price tends to zero. As a matter of fact, except in singular cases, it is this property, and not the fact that the function eventually takes negative values, that rules out equilibrium uniqueness. As shown in figure 5.3, if the marginal revenue is positive at the full employment price, there will be, for a labour disutility v sufficiently weak to ensure existence of an equilibrium, and by continuity of the reservation wage functions, at least two intersections of their graphs, and hence two equilibria. Notice that, contrary to the simple case of a unitary elasticity of intertemporal substitution with increasing returns, this case may entail, and indeed does entail in the example of figure 5.3, two underemployment equilibria.[28] Also, multiplicity of equilibria is compatible in the present case with the absence of labour disutility.

Are such equilibria Pareto ranked? From the discussion about the increasing returns case, in order to establish that lower price equilibria Pareto dominate higher price ones, it is sufficient to show that the utility derived from the profit income $U^*\Pi/p$ is non-increasing in price in the

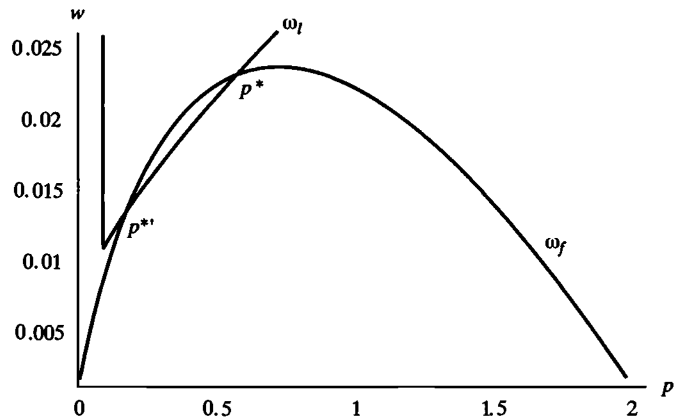


Figure 5.3 Multiple underemployment equilibria

relevant range. As the elasticity of this function is equal, by (8), to $-\alpha$ plus the elasticity of the function

$$\begin{aligned}\Pi(p;\hat{P}) &\equiv pD(p,\hat{P}) - \omega_f(p;\hat{P})(mn)^{1-\beta}D(p,\hat{P})^{\beta} \\ &= pD(p,\hat{P})(1/\beta)\,[\beta - 1 + 1/n\delta(1,p/\hat{P})],\end{aligned} \tag{40}$$

this amounts to showing that:

$$\frac{\partial \Pi}{\partial p}\frac{p}{\Pi} = 1 - \Delta - \frac{\delta_2'\theta/\delta}{(\beta-1)n\delta+1} \leq \alpha. \tag{41}$$

By referring to (13) and (15), and using $\Delta'\theta \geq -\Delta(1-\Delta)$ (implied by the assumption of Proposition 1),[29] together with the fact that $n\delta > 1$ in the relevant range, we indeed obtain:

$$\begin{aligned}\frac{\partial \Pi}{\partial p}\frac{p}{\Pi} &\leq (1-\Delta)\left[1 + \frac{1}{(\beta-1)n\delta+1}\,\frac{\Delta}{\Delta+(m-1)s}\right] \\ &< \alpha(1-\sigma)\left[1 + \frac{1}{\beta(1+(m-1)s)}\right] \leq \alpha.\end{aligned} \tag{42}$$

The last inequality is satisfied if we assume: $\sigma(\theta) \geq 1/[1+\beta(1+(m-1)s)]$ for the relevant values of θ (those which lead to $\delta(1,\theta) > 1/n$). A simple example of existence of multiple Pareto ranked temporary equilibria based on the CES utility function (as in the example of figure 5.3) is thus obtained for values of σ such that:

$$\frac{1}{1+\beta(1+(m-1)s)} \leq \sigma < m/n - (m-1)s < 1. \tag{43}$$

Stationary equilibria

Under temporary equilibria, expectations are exogenous and arbitrary, and one may always wonder whether the multiplicity of equilibria obtained should not be traced back to the degree of freedom thus allowed to the model-maker. Considering intertemporal equilibria, such degree of freedom is lost since expectations, even though rigid and devoid of cognitive foundations, are required to be fulfilled.

The study of intertemporal equilibria is often limited to stationary equilibria, allowing us to remove the arbitrariness of expectations without giving up the simplicity of the static analysis. Imposing stationarity may, however, be misleading. Indeed, in the symmetric temporary equilibrium context, it shares with the assumption of a unitary elasticity of intertemporal substitution the property that demand elasticity, and hence market power, is constant (since $\theta_t = 1$ for all t). As a result, we obtain, as in perfect competition, a unitary elasticity of the symmetric marginal revenue function. Thus, we should not be surprised to find that non-decreasing returns *per se* exclude (generically) the possibility of multiple stationary equilibria, independently of the characteristics of intertemporal substitution. By contrast, given existence, increasing returns necessarily entail such a multiplicity (except in a singular case).

The analysis of stationary equilibria is straightforward. Reservation wage functions take quite simple forms when the real interest factor is fixed at unity:

$$\omega_l(p;p) = Ap \text{ and } \omega_f(p;p) = Bp^{\beta}, \tag{44}$$

with $A = v/U^*(1)$ and $B = [1 - 1/n\delta(1,1)](1/\beta)[M/(mn(1-\alpha(1)))]^{1-\beta}$. If $A > 0$ and $B > 0$, and except in the case of constant returns ($\beta = 1$), there is a unique price, say p^*, at which the two reservation wages are equal. Under decreasing returns ($\beta > 1$), if this price is larger than the full employment price p^{**}, it sustains an underemployment stationary equilibrium. Otherwise the only stationary equilibrium has full employment. Under increasing returns ($\beta < 1$), *both* p^* and p^{**} are stationary equilibrium prices if $p^* \geq p^{**}$ and if the non-negative profit condition is satisfied. Otherwise no stationary equilibrium exists. Equality of p^* and p^{**} corresponds to a degree $\bar{v}$ of labour disutility satisfying:

$$\bar{v} \equiv U^*(1)Bp^{**\beta-1} = U^*(1)[1 - 1/n\delta(1,1)](1/\beta)(L/mn)^{1/\beta - 1}. \tag{45}$$

We can use $\bar{v}$ to summarize the preceding remarks in a proposition.

Proposition 5 (stationary equilibria): Under the assumptions of either Proposition 1 or Proposition 2, and referring to $\bar{v}$ as defined in (45),

- if returns are *decreasing* ($\beta > 1$), there exists a *unique* stationary equilibrium, with full employment if $0 \leq v \leq \bar{v}$ (one inequality strict), or with underemployment if $0 < \bar{v} < v$; in the *singular* case in which $v = \bar{v} = 0$, there exist multiple stationary equilibria, one with full employment and a continuum with underemployment, corresponding to any price higher than the full employment price; and, if $\bar{v} \leq 0 < v$, no stationary equilibrium exists;
- if returns are *constant* ($\beta = 1$), there exists a *unique* stationary equilibrium with full employment if $v < \bar{v}$; there exist multiple stationary equilibria, one with full employment and a continuum with underemployment, corresponding to any price higher than the full employment price, in the *singular* case in which $v = \bar{v}$; and no stationary equilibrium exists if $\bar{v} < v$;
- if returns are *increasing* ($\beta < 1$) and if $\beta \geq 1 - 1/n\delta(1,1)$, there exist *two* stationary equilibria, one of each type, if $0 < v < \bar{v}$, and only one stationary full employment equilibrium in the *singular* cases in which $0 < v = \bar{v}$ or $0 = v < \bar{v}$; there exist multiple stationary equilibria, one with full employment and a continuum with underemployment, corresponding to any price higher than the full employment price, in the *singular* case in which $0 = v = \bar{v}$; and no stationary equilibrium exists if $\bar{v} < v$ or if $\beta < 1 - 1/n\delta(1,1)$.

Are multiple stationary equilibria Pareto ranked? Because of stationarity and the associated constancy of both expenditure and market power, the answer is based on the same argument as in the context of temporary equilibria when the elasticity of intertemporal substitution is equal to one. Lower price equilibria are obviously preferred by all generations when old, and are at least no less desired by young consumers as wage earners (the utility derived from their worker's status is zero in any underemployment equilibrium and may be greater at full employment) and also as profit earners (with a constant real interest factor, the utility derived from the shareholder's status is proportional to real profits, and nominal profits are constant in this case).

Endogenous fluctuations

By its very simplicity, the study of stationary equilibria is both appealing and delusive. Indeed, stationary equilibria are but a special type of intertemporal equilibria, and should be viewed as such in a genuine dynamic analysis. Non-stationary equilibria may be quite complex, and it is in particular possible to find, along an equilibrium sequence, both underemployment and full employment phases. For the sake of simplicity, we shall however emphasize the two 'pure' dynamic regimes, corresponding to the full employment and the underemployment regimes. These two regimes will be successively examined in the next subsection. Then, in the following subsection, monotonicity of all equilibrium trajectories will be established for the substitutability case with non-increasing returns. In the two remaining subsections, we will then show how imperfect competition may contribute to the emergence of endogenous fluctuations, by allowing increasing returns on the one hand, and through variability of market power on the other.

Two regimes of dynamics

Full employment

The first significant feature of full employment intertemporal equilibria, whether stationary or not, is that stationarity prevails at least in output, given symmetry and the constancy of labour supply and productivity. Thus, the demand must satisfy, for any $t=0,1,2,\ldots$,

$$D(p_t,p_{t+1})=D(p_{t+1},p_{t+2}). \tag{46}$$

The second important feature of full employment intertemporal equilibria is that they are determined quite independently of market structure, according to a dynamic system which is the same, whether competition is perfect or imperfect. It is only through the two admissibility conditions, of non-negative profits and proper ranking of the reservation wages, that market power may have an influence, by ruling out inadmissible trajectories.

Because of homogeneity (of degree -1) of the demand function, the difference equation of order 2 given by (46) can be transformed into a difference equation of order 1 in the variable $\theta_t=p_t/p_{t+1}$:

$$D(\theta_t,1)=\theta_{t+1}D(\theta_{t+1},1), \tag{47}$$

or equivalently, since $[1-\alpha(\theta_t)]\theta_t=M/D\ (\theta_t,1)$ by (11):

$$[1-\alpha(\theta_t)]\theta_t=1-\alpha(\theta_{t+1}). \tag{48}$$

The left-hand side of (48) is an increasing function of θ_t (with elasticity $\Delta(\theta_t)$), so that (48) always has, for any θ_{t+1} in some interval of $\mathscr{R}_{++}$ (in general $\mathscr{R}_{++}$ itself), a unique solution in θ_t, and thus characterizes well-defined *backward* dynamics in the real interest factor.

In the simple cases that we are considering, the substitutability and the complementarity cases, α is a monotonic function. So it becomes possible to consider also well-defined *forward* dynamics: $\theta_{t+1}=\Phi_{FE}(\theta_t)$ on some interval of $\mathscr{R}_{++}$. As $\alpha(\theta)<1$ for any positive θ, the only fixed point of Φ_{FE} is

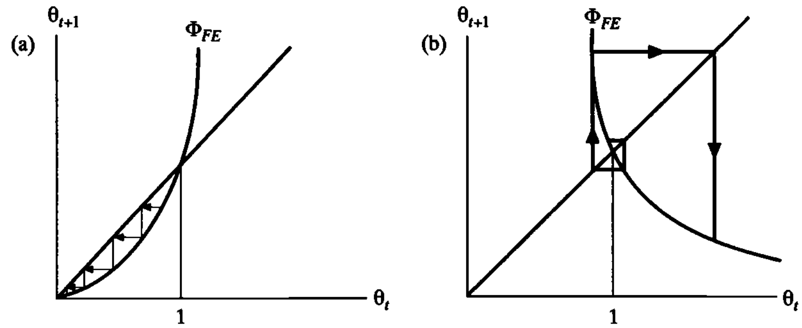


Figure 5.4 Dynamics in the full-employment regime
(a) Substitutability case (b) Complementarity case

clearly 1. Moreover, since Φ_{FE} is increasing in the case of intertemporal substitutability (α decreasing), decreasing in the case of intertemporal complementarity (α increasing), the dynamics are monotonic (both in θ_t and in p_t) in the first case, alternating (in both variables) in the second case[30] (see figure 5.4). Stability of the stationary equilibrium can be assessed by referring to the absolute value of Φ'_{FE} at $\theta_t = 1$, with

$$\Phi'_{FE}(1) = \left.\frac{d\theta_{t+1}}{d\theta_t}\right|_{\theta_{t+1}=\theta_t=1} = \frac{\Delta(1)}{\Delta(1)-1}. \tag{49}$$

When investigating existence of non-stationary intertemporal equilibria, one can of course use the dynamic system described by Φ_{FE}, keeping in mind that any sequence $(\theta_t)_{t\in\mathcal{N}}$ derived from an initial value θ_0 by successive iterations of Φ_{FE}, characterizes an equilibrium only if profits are non-negative and reservation wages are properly ranked in all periods, *i.e.* if, for any t,

$$\beta\left(\frac{L}{mn}\right)^{1-1/\beta} v/U^*(\theta_t) \leq 1 - 1/n\delta(1,\theta_t) \leq \beta. \tag{50}$$

Underemployment

Underemployment equilibria must satisfy, for any $t = 0,1,2,\ldots$, the equality of the two reservation wages:

$$\omega_f(p_t, p_{t+1}) = \omega_l(p_t, p_{t+1}), \tag{51}$$

i.e., using homogeneity (of degree β for ω_f and 1 for ω_l):

$$\omega_f(\theta_t, 1)p_{t+1}^{\beta-1} = \omega_l(\theta_t, 1). \tag{52}$$

As can be seen from (52), the dynamics in θ_t degenerate if the returns are constant ($\beta = 1$), or if labour has no disutility (i.e. if $v = 0$, leading to $\omega_l \equiv 0$). Any underemployment equilibrium is then constant in the real interest factor, hence quasi-stationary. Deflationary equilibria are excluded, since demand would be unbounded along the equilibrium path. Thus, in each one of those two situations (except in the singular case in which 1 is a solution to (52)), any equilibrium must be inflationary and declining: prices and the nominal wage increase at the same constant rate $1/\theta - 1$, with outputs decreasing at rate $1 - \theta$.

If $\beta \neq 1$ and $v \neq 0$, one can divide both sides of (52) by the corresponding sides of the same equation, applying to the next period, to get a difference equation of order 1 in the variable θ_t:

$$f(\theta_t) \equiv \frac{\omega_f(\theta_t, 1)}{\omega_l(\theta_t, 1)} = \frac{\omega_f(\theta_{t+1}, 1)}{\omega_l(\theta_{t+1}, 1)} \theta_{t+1}^{1-\beta} \equiv g(\theta_{t+1}). \tag{53}$$

Any solution to this equation characterizes an underemployment intertemporal equilibrium, provided the non-negativity condition on profits and the labour availability constraint are satisfied in each period t. Since the output derived from the equality of the two reservation wage functions is given by:

$$y_t = \left[\frac{1}{\beta v} U^*(\theta_t)(1 - 1/n\delta(1, \theta_t)) \right]^{\frac{1}{\beta - 1}}, \tag{54}$$

the two admissibility conditions are respectively:

$$\beta \geq 1 - 1/n\delta(1, \theta_t) \text{ and } \frac{L}{mn} > \left[\frac{1}{\beta v} U^*(\theta_t)(1 - 1/n\delta(1, \theta_t)) \right]^{\frac{\beta}{\beta - 1}}, \tag{55}$$

Using (28) and (30), elasticities of the left-hand and of the right-hand sides of (53) with respect to θ_t and θ_{t+1} are respectively:

$$\frac{f'}{f} \theta_t = \frac{\omega_f'}{\omega_f} \theta_t - \frac{\omega_l'}{\omega_l} \theta_t = (\beta - 1)\Delta + \frac{\rho'}{\rho} \theta_t - \alpha \tag{56}$$

and

$$\frac{g'}{g} \theta_{t+1} = \frac{\omega_f'}{\omega_f} \theta_{t+1} - \frac{\omega_l'}{\omega_l} \theta_{t+1} + 1 - \beta = (\beta - 1)(\Delta - 1) + \frac{\rho'}{\rho} \theta_{t+1} - \alpha. \tag{57}$$

Neither of these elasticities has a well determined sign, the same for all possible cases, or even, in some cases, the same for all values of θ_t or θ_{t+1}. Thus, contrary to the full employment difference equation (48), (53) does not necessarily lead to well defined dynamics, either backwards or forwards. However, putting together equality (56) and inequality (35), one sees that the condition entailing uniqueness of temporary equilibrium (that the elasticity of the firms' reservation wage be larger than the elasticity of the workers' reservation wage) naturally leads to well defined *backward* underemployment dynamics (since f is an increasing function in that case). As we have seen, that condition is in particular satisfied when competition is perfect, so that increasing returns are excluded (hence $\beta > 1$) and marginal revenue is equal to price (hence $\rho'\theta/\rho = 1$).

Also, in the substitutability case, we know from Lemma 4 in the appendix that the elasticity of the symmetric marginal revenue function ρ is always larger than the propensity to consume α. We can then conclude that, for $\beta > 1$, both sides of (53) are increasing. Thus, both the forward and the backward dynamics are well defined in this case. We can in particular consider a function Φ_{UE}, defined on some interval of $\mathscr{R}_{++}$, and describing the *forward* dynamics characterized by (53): $\theta_{t+1} = \Phi_{UE}(\theta_t)$. The function Φ_{UE} is increasing, with a derivative larger than 1 at 1, its unique fixed point:

$$\Phi'_{UE}(1) = \left.\frac{d\theta_{t+1}}{d\theta_t}\right|_{\theta_{t+1}=\theta_t=1} = \frac{(\beta-1)\Delta(1)+\rho'(1)/\rho(1)-\alpha(1)}{(\beta-1)(\Delta(1)-1)+\rho'(1)/\rho(1)-\alpha(1)}. \tag{58}$$

Before we show how imperfect competition can contribute to endogenous fluctuations, let us consider this particularly simple case, in which fluctuations are precisely excluded.

The substitutability case with non-increasing returns

Assume first decreasing returns ($\beta > 1$). The two dynamics are then essentially the same: they can both be defined forwards by explicit functions Φ_{FE} and Φ_{UE} (in full employment and underemployment, respectively), which are increasing, have 1 as their unique fixed point, with derivatives larger than 1 at this point (see (49) and (58)). So, the stationary equilibrium is always unstable. If $\theta_0 > 1$, θ_t would indefinitely increase, implying that the price would converge to zero. But this is impossible, since demand would then tend to infinity whereas production is bounded. Thus, besides the stationary equilibrium, only equilibria with explosive inflation (such that θ_t converges to 0) can exist. As a matter of fact, there generally exists a continuum of them, for any $\theta_0 < 1$, as we will see.

In order to assert existence of such equilibria, we must take the admissibility conditions (50) and (55) into account. Let us define, as a function of the degree of labour disutility, the value $\tilde{\theta}(v)$ of the real interest factor that leads to equality of firms' and workers' reservation wages at full employment:

$$(1/\beta)(L/mn)^{1/\beta-1}[1-1/n\delta(1,\tilde{\theta}(v))] = v/U^*(\tilde{\theta}(v)). \tag{59}$$

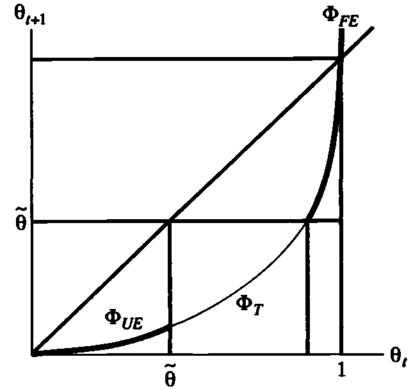


Figure 5.5 Dynamics with possible shift of regime

As $\rho'\theta/\rho > \alpha$ when substitutability prevails, the elasticity with respect to θ of the left-hand side of this equation is larger than the one of the right-hand side, so that a solution $\tilde{\theta}(v)$ is uniquely determined. A positive solution exists if $v > \underline{v}$, with

$$\underline{v} \equiv (1/\beta)(L/mn)^{1/\beta - 1} \lim_{\theta \to 0} \{[1 - 1/n\delta(1,\theta)]U^*(\theta)\}. \tag{60}$$

Also, notice that $\tilde{\theta}$ is increasing and that, taking $\bar{v}$ as defined in (45), we obtain $\tilde{\theta}(\bar{v}) = 1$. Finally, since by (54) output y_t increases with the real interest factor θ_t, any value of θ_t will be associated with underemployment if $\theta_t < \tilde{\theta}(v)$ and compatible with full employment if $\theta_t \geq \tilde{\theta}(v)$.

Two cases must now be considered. If $v \leq \underline{v}$, full employment results from any positive value of the real interest factor, so that a sequence starting with any $\theta_0 \in]0,1[$ and determined by successive iterations of the function Φ_{FE} characterizes a full employment equilibrium (see figure 5.4a), provided those iterations can go on forever, that is provided $\lim_{\theta \to 0} \alpha(\theta) = 1$ (a condition implied by $\lim_{\theta \to 0} \sigma(\theta) > 1$).[31] This also applies to the constant returns case ($\beta = 1$), since full employment dynamics are not affected by technological characteristics.

If $v > \underline{v}$, we can associate with any $\theta_0 \in]0, \tilde{\theta}(v)[$ a sequence determined by successive iterations of the function Φ_{UE} (we show in the appendix that this function is indeed defined on the whole interval $]0, \tilde{\theta}(v)[)$, a sequence characterizing an underemployment equilibrium (see figure 5.5). But, if $v < \bar{v}$ and hence $\tilde{\theta}(v) < 1$, we can also take a sequence starting at $\theta_0 \in [\tilde{\theta}(v), 1[$, which will characterize an equilibrium with full employment in a finite number of periods and underemployment thereafter. As long as θ_t belongs to the interval $[\Phi_{FE}^{-1}(\tilde{\theta}(v)),1[$, dynamics are, as we know, described by the function Φ_{FE}. However, if $\theta_t \in [\tilde{\theta}(v), \Phi_{FE}^{-1}(\tilde{\theta}(v))[$ equilibrium involves a transition from the full employment to the underemployment regime. We show in the proof of Proposition 6, in the appendix, that such transition is described by an increasing function Φ_T which continuously extends both Φ_{UE} and Φ_{FE}, so that the dynamics are essentially the same in this case as in the pure regimes.

In the constant returns case, there is a single possible real interest factor $\theta_t = \tilde{\theta}(v)$, but there still exists a continuum of underemployment equilibria, associated with any initial price larger than the full employment price (the price corresponding to $\tilde{\theta}(v)$), as soon as $\tilde{\theta}(v) < 1$, i.e. $v < \bar{v}$. Of course, no deflationary equilibrium can exist, so that $\tilde{\theta}(v) > 1$ leads to the inexistence of an intertemporal equilibrium. We thus obtain the following proposition:

Proposition 6 (monotonicity): Consider the substitutability case, as characterized by Proposition 2, and assume non-increasing returns. Refer to $\underline{v}$, $\bar{v}$ and $\tilde{\theta}(v)$ as defined by (60), (45) and (59), respectively.

- If $0 \leq v \leq \underline{v}$, there only exist full employment intertemporal equilibria, the stationary one and, provided $\lim_{\theta \to 0} \alpha(\theta) = 1$, a continuum of equilibria with explosive inflation associated with any $\theta_0 \in]0,1[$.
- If $\underline{v} < v \leq \bar{v}$, there is a stationary full employment equilibrium and, for $\rho > 1$ or $v < \bar{v}$ a continuum of inflationary equilibria with eventually declining output. If $\beta > 1$, those equilibria, associated with any $\theta_0 \in]0,1[$, are characterized by explosive inflation. If $\beta = 1$, those equilibria have constant inflation at the rate $1/\tilde{\theta}(v) - 1$, and are associated with any initial price at least as large as the full employment price corresponding to $\tilde{\theta}(v)$.
- If $\bar{v} < v$ and $\beta > 1$, there only exist underemployment intertemporal equilibria, the stationary one and, associated with any $\theta_0 \in]0,1[$, a continuum of equilibria with explosive inflation and declining output. No equilibrium exists if $\bar{v} < v$ and $\beta = 1$.

Proof: See appendix.

It is again quite easy to establish that monotonic equilibria are Pareto ranked, at least if the aggregate demand elasticity Δ is a decreasing function (for instance, if U is CES). Indeed, real wage costs per unit of output are then a decreasing function of the real interest factor, since market power is positively related to this variable. Besides, U^* and, if returns are decreasing, underemployment equilibrium output are increasing functions of the real interest factor θ. Thus young consumers, as shareholders, unambiguously prefer in both regimes higher values of θ. As we know, they are indifferent, as workers, among underemployment equilibria, and prefer in full employment higher values of θ, since the favourable effect of an increase of the real interest factor on U^* (given by the elasticity $1-\alpha$) dominates, when substitutability prevails, the negative effect through market power of that increase on the real wage (given by the elasticity $\rho'\theta/\rho-1$). Finally, old consumers would object to lower values of θ. Indeed, the Keynesian multiplier is in this case a decreasing function of θ, so that higher prices must compensate for lower values of the real interest factor and higher associated values of the multiplier, if demand is to be kept at its full employment level. As for underemployment equilibria, the same relation prevails between prices and values of θ, as can be seen from (52). Thus, and because of monotonicity, equilibria in the decreasing returns case are positively related (according to the Pareto criterion) to the initial values of the real interest factor. The preceding argument also applies to the constant returns case, except that underemployment equilibria are negatively related to initial prices.

As we see, the multiplicity of Pareto ranked intertemporal equilibria may occur even in the substitutability case under non-increasing returns, a situation which excludes the multiplicity of temporary equilibria. This conclusion is however by no means related to market power, and is quite compatible with perfect competition. It is the overlapping-generations structure together with an infinite horizon of the economy, and not the imperfectly competitive nature of its markets, that explains the equilibrium indeterminacy, through the indeterminacy of potentially harmful yet self-fulfilling expectations, leading to Pareto inferior intertemporal equilibria.

Increasing returns

Substitutability and non-increasing returns together lead to monotonicity of intertemporal equilibria. A continuum of inflationary and declining equilibria may then exist, but fluctuations are excluded. This is no longer the case under increasing returns. Of course, as soon as multiple equilibria are non-monotonic, they will in general not be Pareto ranked. In any case, it is the possibility of occurrence of endogenous fluctuations in itself that will, from now on, hold our attention.

In order to focus on the role of increasing returns as a source of endogenous fluctuations, we shall first stick to the substitutability case (the combined influence of increasing returns and complementarity will be considered in the next subsection). No difference arises from the assumption of increasing returns in full employment dynamics, which are independent of technological specifications, because of stationarity in output. As for underemployment dynamics, by (56) and (57), the elasticities of both the left- and right-hand sides of the difference equation (53) will in general take negative values for θ close to 0, and positive values for θ large enough. Indeed, if we exclude the singular case in which $\lim_{\theta \to 0} \sigma(\theta) = 1$ (so that $\lim_{\theta \to 0} \alpha(\theta) = 1$) and taking into account the fact that, by (A20) and (A29) in the appendix, $\lim_{\theta \to 0} \rho'(\theta)\theta/\rho(\theta) = 1$), we obtain:

$$\begin{aligned} &\lim_{\theta \to 0} \frac{f'(\theta)}{f(\theta)}\,\theta = (\beta - 1) \lim_{\theta \to 0} \sigma(\theta) < 0, \\ &\lim_{\theta \to 0} \frac{g'(\theta)}{g(\theta)}\,\theta = (\beta - 1)\left(\lim_{\theta \to 0} \sigma(\theta) - 1\right) < 0, \end{aligned} \tag{61}$$

and, as $\lim_{\theta \to \infty} \alpha(\theta) = 0$ and $\lim_{\theta \to \infty} \rho'(\theta)\theta/\rho(\theta) = 1$,

$$\lim_{\theta \to \infty} \frac{f'(\theta)}{f(\theta)}\,\theta = \beta > 0, \; \lim_{\theta \to \infty} \frac{g'(\theta)}{g(\theta)}\,\theta = 1 > 0. \tag{62}$$

The functions f and g are thus non-monotonic, and the dynamics generated by (53) are not uniquely defined, either forwards or backwards.

To keep things straightforward, let us concentrate on the existence of a 2-cycle, the simplest form of persistent endogenous fluctuations. We look for a pair (θ^*, θ^{**}) of two different positive values of the real interest factor verifying the difference equation (53), with $\theta_t = \theta_{t+2} = \theta^*$ and $\theta_{t+1} = \theta^{**}$. Since this equation has the form

$$f(\theta_t) = f(\theta_{t+1})\theta_{t+1}^{1-\beta}, \tag{63}$$

we obtain:

$$f(\theta^*) = f(\theta^{**})\theta^{**\,1-\beta} = [f(\theta^*)\theta^{*\,1-\beta}]\theta^{**\,1-\beta}$$

and, since f takes only positive values, we get: $\theta^{**} = 1/\theta^* > 1$ (without loss of generality). Thus, the difference equation (63) admits a 2-cycle as a solution if and only if there exists $\theta^* \in]0,1[$ verifying the equation:

$$F(\theta) \equiv \frac{f(\theta)}{f(1/\theta)}\,\theta^{1-\beta} = 1, \tag{64}$$

where F has elasticity:

$$\frac{F'(\theta)}{F(\theta)}\,\theta = \frac{f'(\theta)}{f(\theta)}\,\theta + \frac{f'(1/\theta)}{f(1/\theta)}\,\frac{1}{\theta} + 1 - \beta. \tag{65}$$

As $F(1)=1$, a 2-cycle can exist only if F is non-monotonic, hence (for $\beta<1$) if f is somewhere decreasing, which is precisely *the condition required for multiple temporary equilibria*. By continuity of F, a sufficient condition for the existence of a solution $\theta^* \in]0,1[$ to (64) is that $F'(1)<0$ together with $\lim_{\theta\to 0} F(\theta)=0$. Let us assume that $\lim_{\theta\to\infty} \alpha(\theta)=0$ (which results simply from excluding the case of a Cobb–Douglas intertemporal utility function). Then, by (65), (56), and Lemma 4 in the appendix, and recalling that $\lim_{\theta\to 0} \sigma(\theta)\leq 1+\beta$, by the assumption of Proposition 2, we get:

$$\lim_{\theta\to 0} \frac{F'(\theta)}{F(\theta)}\,\theta \geq (\beta-1)\lim_{\theta\to 0}\sigma(\theta)+1\geq \beta^2>0,$$

and hence $\lim_{\theta\to 0} F(0)=0$. Let us now consider the sign of $F'(1)$. Using again (65) and (56), we obtain:

$$F'(1)=2\frac{f'(1)}{f(1)}+1-\beta=-2\left[(1-\beta)\left(\Delta(1)-\frac{1}{2}\right)+\alpha(1)-\frac{\rho'(1)}{\rho(1)}\right]. \quad (66)$$

Referring to this equation and to (58), one straightforwardly checks that $F'(1)<0$ if and only if the stationary equilibrium is unstable in the (forward) underemployment dynamics, i.e. if $|\Phi'_{UE}(1)|>1$ (where Φ_{UE} is now the implicit function defined by the difference equation (53) in a neighbourhood of the stationary state).

From (66) we see that favourable circumstances for the existence of a 2-cycle are: (1) a high degree of impatience (think of a CES intertemporal utility function with a small discount factor for future consumption) leading to a high value $\alpha(1)$ of the propensity to consume when young and leading also to a large elasticity $\Delta(1)$ of aggregate demand (indirectly and in conjunction with a high elasticity of intertemporal substitution); (2) significantly increasing returns (small β); (3) some variability of market power (expressed, when Δ is a decreasing function, as in the case of constant elasticity of intertemporal substitution, by a small value of $\rho'(1)/\rho(1)$).

Of course, in order that the 2-cycle $(\theta^*,1/\theta^*)$ (with $0<\theta^*<1$) satisfying the difference equation (63) be an intertemporal equilibrium, the usual admissibility conditions must be verified, both the non-negative profits condition: $\beta\geq 1-1/n\delta(1,\theta^*)$ (for a decreasing Δ, the condition prevails also for $1/\theta^*$), and the labour availability condition: either $\theta^*\geq\tilde{\theta}(v)$ (with $\tilde{\theta}(v)$ as defined in (59) and taking into account the fact that y_t, as given by (54), is now a decreasing function of θ_t) or else $v\leq \underline{v}$ (with $\underline{v}$ as defined by (60)). The first condition is verified, even for θ^* arbitrarily small, if $\beta\geq 1-1/ns$ (since $\delta<s$ if $\sigma\leq s$). As for the second condition, if v is larger than $\underline{v}$ but small

enough, so that $0 < \tilde{\theta}(v) \leq \theta^*$, the inflationary full employment equilibria cease to exist, but the existence of the 2-cycle as an underemployment equilibrium is preserved.

Complementarity and strong variability of market power

We will first show in this subsection that complementarity may contribute to the occurrence of 2-cycles, and then that, by leading to negative values of the marginal revenue for some range of the real interest factor (an extreme form of variability of market power), it may also generate persistent endogenous fluctuations in *all* intertemporal equilibria.

Existence of cycles of period 2 when the stationary equilibrium is unstable

The argument used in the earlier subsection to establish existence of a 2-cycle in the underemployment dynamics also applies to the complementarity case. By (65), (56), and (A16) and (A29) in the appendix, and assuming that $\lim_{\theta \to 0} \alpha(\theta) = 0$ (which is implied by $\lim_{\theta \to 0} \sigma(\theta) < 1$), and that $(1 - 1/m)s + 1/m \inf_{\theta \in]0,\infty[} \Delta(\theta) > 1/n$, we get:

$$\lim_{\theta \to 0} \frac{F'(\theta)}{F(\theta)} \theta \geq 1 + (\beta - 1) \lim_{\theta \to \infty} \sigma(\theta) \geq 1 \text{ if } \beta > 1$$
$$\geq \beta \text{ if } \beta < 1.$$

As the elasticity of F always tends to a positive limit when θ tends to 0, we have: $\lim_{\theta \to 0} F(\theta) = 0$. Thus, a sufficient condition for the existence of a 2-cycle in the underemployment dynamics is again $F'(1) < 0$ or, equivalently, the instability of the stationary state in the forward underemployment dynamics.

A simple inspection of (66) shows that a large variability of market power expressed, when Δ is a decreasing function, by a small value of $\rho'(1)/\rho(1)$, makes that condition easier to satisfy. Also, the lower is the value of β, for $\sigma(1) \geq 1/2$ (and hence $\Delta(1) > 1/2$), the easier is that condition to satisfy. Increasing returns are still favourable to the occurrence of a 2-cycle when complementarity is weak. In that case, the same conclusions apply to the (favourable) influence of a high degree of impatience (leading to a large $\alpha(1)$) and of a not too small elasticity of intertemporal substitution $\sigma(1)$. If returns are decreasing, it is on the contrary a small value of $\sigma(1)$ that favours occurrence of a 2-cycle.

We should point out that, as long as complementarity and variability of market power are both weak ($\sigma(\theta)\geq 1/2$ and $\rho'(\theta)\theta/\rho(\theta)\geq\alpha(\theta)$ for any θ), no 2-cycle can exist under non-increasing returns in either of the two regimes. Indeed, existence of a 2-cycle in underemployment dynamics requires (64) to have more than one solution, hence F to be non-monotonic. But, by (65) and (56), we obtain in that case:

$$\frac{F'(\theta)}{F(\theta)}\theta=(\beta-1)\left[\Delta(\theta)+\Delta\left(\frac{1}{\theta}\right)-1\right]+\frac{\rho'(\theta)}{\rho(\theta)}\theta-\alpha(\theta)$$
$$+\frac{\rho'(1/\theta)}{\rho(1/\theta)}\frac{1}{\theta}-\alpha\left(\frac{1}{\theta}\right)\geq 0.$$

Exactly the same argument applies to full employment dynamics, once $\Delta(\theta)+\Delta(1/\theta)-1\geq 0$.

If complementarity is strong enough, or more precisely if the elasticity of intertemporal substitution is allowed to take values smaller than 1/2, one can, however, obtain the existence of 2-cycles, either in the full employment regime ($\Delta(1)<1/2$ is then a sufficient condition) or in the underemployment regime, and even with decreasing returns and a constant, in fact nil, market power, i.e. even under perfect competition. In the latter, given high degrees of impatience and a large intertemporal complementarity, more complex dynamics, with cycles of higher periods, are possible. Thus, strong complementarity is as such, and quite independently of its effects through market power, a potential source of persistent endogenous fluctuations. We are however interested in effects that are specific to imperfect competition. Such effects work through the variability of market power, an extreme form of which is the emergence of negative values of the marginal revenue. That possibility not only favours occurrence of endogenous fluctuations, it can in fact threaten existence of stationary or, in conjunction with increasing returns, even monotonic intertemporal equilibria.[32]

Non-stationarity and non-monotonicity of all intertemporal equilibria

It is not difficult to find conditions implying that any intertemporal equilibrium is non-stationary, even non-monotonic. But a related question has to be considered: do such non-stationary (or non-monotonic) equilibria exist, or on the contrary do those conditions preclude existence of intertemporal equilibria in general? Indeed existence of a stationary equilibrium, either in full employment or underemployment, is clearly ruled out if the marginal revenue is negative at a real interest factor equal to one: $\rho(1,1)<0$. If in addition returns are increasing, no monotonic intertemporal equilibrium can exist. We know indeed that deflationary equilibria are impossible, since θ_t cannot converge to 1 without making the marginal revenue negative, so that prices would necessarily converge to zero in such an equilibrium, leading to a violation of the labour availability constraint. On the other hand, inflationary equilibria are also impossible in this case, since the fact that θ_t cannot converge to 1 has the further consequence that the sequence of prices would be unbounded, and hence that output and, under increasing returns, labour productivity would both converge to zero. The firm's reservation wage would then eventually become smaller than the labourers' reservation wage, which is incompatible with an equilibrium. So, if intertemporal equilibria exist, they cannot be monotonic.

Now, in order to get existence in the case of decreasing returns,[33] let us assume, together with the conditions of Proposition 1, that[34]

$$\Delta(1) < m/n - (m-1)s < 1 = \lim_{\theta \to 0} \Delta(\theta), \tag{67}$$

or equivalently, by (15) and (26), that the marginal revenue function takes negative values for θ close to 1, and positive values for θ close to 0. The marginal revenue function and consequently the firms' reservation wage function must then switch from positive to negative values at some $\underline{\theta} \in]0,1[$. Assuming regularity ($f'(\underline{\theta}) < 0$), the difference equation (53), of which $\underline{\theta}$ is clearly a stationary solution, defines in a neighbourhood of $\underline{\theta}$ an implicit function Φ_{UE} with $\underline{\theta}$ as a fixed point and such that $\theta_{t+1} = \Phi_{UE}(\theta_t)$. Also,

$$\begin{aligned} 0 < \Phi'_{UE}(\underline{\theta}) &= \left.\frac{d\theta_{t+1}}{d\theta_t}\right|_{\theta_{t+1}=\theta_t=\underline{\theta}} \\ &= \frac{f'(\underline{\theta})}{f'(\underline{\theta})\underline{\theta}^{1-\beta} + f(\underline{\theta})(1-\beta)\underline{\theta}^{-\beta}} = \underline{\theta}^{\beta-1} < 1, \end{aligned} \tag{68}$$

since $f(\underline{\theta}) = 0$. Thus, the stationary solution $\underline{\theta}$ is stable in the forward dynamics and attracts the trajectories starting with values of the initial real interest factor θ_0 in a neighbourhood of $\underline{\theta}$. Since the marginal revenue remains positive and the labour availability constraint is satisfied for θ_t smaller than, but sufficiently close to, $\underline{\theta}$ (by (54), y_t tends to 0 if θ_t tends to $\underline{\theta}$), we obtain existence of a continuum of equilibria converging to $\underline{\theta}$ and associated with different values of $\theta_0 \in]0, \underline{\theta}[$ close enough to $\underline{\theta}$. These equilibria are inflationary. Also, as θ_t is bounded away from 1 and the sequence of prices is consequently unbounded, whereas the multiplier is bounded, these equilibria must be such that demand, and hence output, converge to zero. We sum up the preceding result in a proposition.

Proposition 7 (non-stationarity): With decreasing returns, under the assumptions of Proposition 1, and also assuming that

$$\Delta(1) < m/n - (m-1)s < 1 = \lim_{\theta \to 0} \Delta(\theta),$$

there exists a continuum of inflationary intertemporal equilibria, such that output converges to zero. No stationary equilibrium exists.

Now observe that, if marginal revenue is negative for any θ larger than 1 (as will necessarily be the case if the demand elasticity Δ is a decreasing function), then *all* equilibria are of course inflationary and such that output converges to zero. Stationarity is excluded, but so in that case are endogenous fluctuations. Let us therefore assume that the marginal revenue function takes positive values when the real interest factor is either sufficiently smaller or sufficiently larger than 1:

$$\Delta(\theta) > m/n - (m-1)s \text{ iff } \theta \in]0, \underline{\theta}[\cup]\bar{\theta}, \infty[, \text{ with } \underline{\theta} < 1 < \bar{\theta}. \qquad (69)$$

Such behaviour of the function Δ is possible if we assume, for instance, that σ is an increasing function.[35] Existence of fluctuating equilibria is then possible in any one of the two regimes and under both decreasing and increasing returns. Assuming for instance that $\underline{\theta}\bar{\theta} > 1$, we immediately obtain existence of a 2-cycle in the underemployment dynamics. Indeed, using again the function F as defined in (64), recalling that

$$\lim_{\theta \to 0} F(\theta) = 0$$

if $\lim_{\theta \to 0} \alpha(\theta) = 0$, and noticing that

$$\lim_{\theta \to (1/\bar{\theta})^-} F(\theta) = +\infty$$

since $1/\bar{\theta} < \underline{\theta}$ and by (69), we obtain by continuity of F in the interval $]0, 1/\bar{\theta}[$ existence in that interval of a solution θ^* to (64). That solution satisfies: $0 < \theta^* < 1/\bar{\theta} < \underline{\theta} < 1 < 1/\underline{\theta} < \bar{\theta} < 1/\theta^*$, so that the cycle $(\theta^*, 1/\theta^*)$ lies in the admissible domain where a positive marginal revenue prevails. Of course, the non-negative profits condition and the labour availability constraint should also be satisfied. The former requires a large enough value of β; as $\Delta(\theta) \leq 1$, a sufficient condition is clearly:

$$\beta \geq 1 - \frac{1/n}{(1 - 1/m)s + 1/m}.$$

A sufficient condition for the latter is a large (small) degree v of labour disutility if returns are decreasing (increasing). Indeed, by (54), the outputs associated with each one of the values θ^* and $1/\theta^*$ (independent of v) of the real interest factor along the cycle tend to 0 when v tends to infinity (zero).

Now, persistent endogenous fluctuations, taking for instance, as in the preceding example, the form of a 2-cycle, or converging to it, or taking possibly more complex forms, will characterize *any* intertemporal equilibrium if returns are increasing, a condition that excludes monotonicity of the equilibrium trajectories. Also, if intertemporal complementarity is not too strong, those fluctuations will involve output as well as prices. Indeed, if $\Delta(\theta) > 1/2$ for any θ, no 2-cycle exists in the full employment regime, as we have seen. In fact, no full employment intertemporal equilibrium can then exist.[36] We sum up these conclusions in Proposition 8.

Proposition 8 (non-monotonicity): Under the assumptions of Proposition 1, and also assuming that $\lim_{\theta \to 0} \alpha(\theta) = 0$ and that, for some $\underline{\theta}$ and $\bar{\theta}$ such that $0 < \underline{\theta} < 1 < \bar{\theta}$ and $\underline{\theta}\bar{\theta} > 1$,

$$\Delta(\theta) > m/n - (m-1)s \text{ iff } \theta \in]0, \underline{\theta}[\cup]\bar{\theta}, \infty[,$$

there exists a cyclical underemployment equilibrium of period 2, either under decreasing returns and for a sufficiently high degree v of labour disutility or under increasing returns, large enough values of the elasticity β of the cost function and sufficiently low values of v. Also, under increasing returns, *any* intertemporal equilibrium is characterized by persistent fluctuations in prices and in the real interest factor and, if Δ takes only values larger than 1/2, also in output.

Conclusion

We have analysed an overlapping-generations economy with 'Cournotian monopolistic competition' in the produced goods markets. This generalizes and extends previous macroeconomic models with imperfect competition, either of the Cournotian or of the Chamberlinian variety. All prices have been assumed perfectly flexible and no adjustment costs were introduced. The labour market has been supposed perfectly competitive, in order to emphasize the effects of firms' oligopolistic market power, as opposed to the effects of union power.

Such effects work in our model through two main channels: increasing returns to employment, more precisely decreasing *marginal* costs, and variability of mark-ups over marginal costs, induced by changes in demand elasticity as prices vary. Both channels of influence have often been neglected in previous models, particularly the ones using the Dixit and Stiglitz monopolistic competition framework, by assuming both a constant (or at least non-decreasing) marginal cost and a constant mark-up over marginal cost. This practice cancels the effects we wanted to stress, making output market power somewhat unimportant.

Increasing returns and significant variability of market power, isolated or in conjunction, have been shown to be responsible for multiple Pareto ranked temporary or stationary equilibria on the one hand, and for the occurrence of endogenous equilibrium fluctuations, in an economy which is not subject to shocks, on the other. Perfectly competitive economies can certainly experience such endogenous fluctuations, related in particular to intertemporal complementarity, that is, to the dominance of income effects over substitution effects. But output market power, through the two channels of influence we have emphasized, not only makes endogenous fluctuations more plausible, but can actually exclude any equilibrium devoid of persistent fluctuations. This is the consequence of possible negative marginal revenue making the stationary state inadmissible, and of sharply increasing marginal costs as output diminishes along inflationary trajectories, excluding monotonicity. Variability of market power can however in this case entail positive marginal revenue along fluctuating equilibria, and the occurrence of such equilibria is clearly a possibility.

Appendix

Lemma 1: Consider the problem

$$\max_{y\in\mathscr{R}_+} \pi(y) \equiv \max_{y\in\mathscr{R}_+} \{\psi(y)y - wy^{\beta}\}, \tag{A1}$$

where $w\in\mathscr{R}_+$, $\beta\in\mathscr{R}_{++}$, and ψ: $\mathscr{R}_{++}\rightarrow\mathscr{R}_{++}$ is a C^2 decreasing function (continuously extended to the domain $\mathscr{R}_+$), such that, denoting $\varepsilon(y)\equiv -\frac{\psi'(y)y}{\psi(y)}$, for any $y\in\mathscr{R}_{++}$,

$$1-\beta\leq\varepsilon(y)\leq 1 \Rightarrow \varepsilon'(y)y > -[1-\varepsilon(y)][\varepsilon(y)-(1-\beta)]. \tag{A2}$$

Then, the following condition on $y^*\in\mathscr{R}_{++}$:

$$\psi(y^*)\beta \geq \psi(y^*)\,[1-\varepsilon(y^*)] = w\beta y^{*\beta-1} \tag{A3}$$

is necessary and sufficient for y^* to be a solution to the problem. If it exists, this solution is the unique positive solution.

Proof: *Necessity*: Suppose the problem has a positive solution y^*. Then y^* is a critical point of π:

$$\pi'(y^*) = \psi(y^*)\,[1-\varepsilon(y^*)] - w\beta y^{*\beta-1} = 0, \tag{A4}$$

and has a non-negative value:

$$\pi(y^*) = \psi(y^*)y^* - wy^{*\beta} \geq \lim_{y\to 0} \pi(y) \geq 0. \tag{A5}$$

Sufficiency: Suppose some positive y^* satisfies the first order and non-negative value condition (A3). Then, $1-\beta\leq\varepsilon(y^*)\leq 1$, so that

$$\pi''(y^*) = -\frac{\psi(y^*)}{y^*}\{\varepsilon'(y^*)y^* + [1-\varepsilon(y^*)][\varepsilon(y^*)-(1-\beta)]\} < 0, \tag{A6}$$

by assumption. Hence, $\pi(y^*)$ is a strict local maximum. Suppose π has another local maximum $\bar{y}$ in $[a,b]$, with a and b arbitrarily close to 0 and ∞, respectively. As π is C^2, the two local maxima must be separated by an interior minimum at say $\underline{y}$, so that $\pi'(\underline{y})=0$ (hence $\varepsilon(\underline{y})\leq 1$) and $\pi''(\underline{y})\geq 0$. Given assumption (A2) on ε, it must be the case that $\varepsilon(\underline{y})<1-\beta$. But that implies in fact: $\forall y\in]0,\underline{y}]$, $\varepsilon(y)<1-\beta$, since by the same assumption $\varepsilon'(y)>0$ whenever $\varepsilon(y)=1-\beta$. So, we must admit that $a\leq\bar{y}<\underline{y}<y^*$, with $\pi'(\bar{y})\leq 0$. Then (as $\beta<1-\varepsilon(\bar{y})$), we get $\pi(\bar{y})=(\bar{y}/\beta)\ [\psi(\bar{y})\beta - w\beta\bar{y}^{\beta-1}] < (\bar{y}/\beta)\pi'(\bar{y})\leq 0\leq\pi(y^*)$. We conclude that $\pi(y^*)$ is a global maximum and that y^* is the unique positive solution to the problem. □

Lemma 2: Consider the problem

$$\max_{y\in\mathscr{R}_+}\pi(y;\bar{Y})\equiv\max_{y\in\mathscr{R}_+}\{d^{-1}(y+\bar{Y})y - wy^{\beta}\}, \tag{A7}$$

where $w\in\mathscr{R}_+$, $\beta\in\mathscr{R}_{++}$, $\bar{Y}\in\mathscr{R}_+$, and d: $\mathscr{R}_{++}\to\mathscr{R}_{++}$ is a C^2 decreasing onto function such that, denoting $\delta(p)\equiv -\frac{d'(p)p}{d(p)}$, for any $p\in\mathscr{R}_{++}$,

$$\delta(p)\leq 1 \Rightarrow \frac{\delta'(p)p}{\delta(p)} > -[1-\delta(p)] \tag{A8}$$

$$1\leq\delta(p) \text{ and either } \beta\geq 1 \text{ or } \delta(p)\leq\frac{1}{1-\beta} \Rightarrow$$

$$\frac{\delta'(p)p}{\delta(p)} > -[\delta(p)-1][1-(1-\beta)\delta(p)]. \tag{A9}$$

Then, the following condition on $y^*\in\mathscr{R}_{++}$, with $p^*=d^{-1}(y^*+\bar{Y})$:

$$p^*\beta\geq p^*\left(1-\frac{y^*/(y^*+\bar{Y})}{\delta(p^*)}\right)=w\beta y^{*\beta-1} \tag{A10}$$

is necessary and sufficient for y^* to be a solution to the problem. If it exists, this solution is the unique positive solution.

Proof: First, observe that if we identify $\psi(y)\equiv d^{-1}(y+\bar{Y})$, we obtain:

$$\varepsilon(y)=\frac{\eta}{\delta(p)},\ \text{with}\ \eta=\frac{y}{y+\bar{Y}}\ \text{and}\ p=d^{-1}(y+\bar{Y}). \tag{A11}$$

Thus, the condition on y^* is just the transposition of the first order and non-negative value condition (A3) of Lemma 1. Hence, it suffices to apply that lemma to prove the present one, if only we can show that the assumption of the former is implied by the assumption of the latter. By a simple calculation we get:

$$\frac{\varepsilon'(y)y}{\varepsilon(y)} = 1 - \eta + \varepsilon(y)\frac{\delta'(p)p}{\delta(p)}. \tag{A12}$$

Consider first the case: $\delta(p) \leq 1$. By assumption,

$$\varepsilon'(y)y > \varepsilon(y)\{1 - \eta - \varepsilon(y)[1 - \delta(p)]\} = \varepsilon(y)[1 - \varepsilon(y)] \geq 0 \text{ if } \varepsilon(y) \leq 1. \tag{A13}$$

Thus, the assumption of Lemma 1 is fulfilled in this case. Now, consider the case: $\delta(p) \geq 1$. By assumption, if either $\beta \geq 1$ or $\delta(p) \leq \frac{1}{1-\beta}$ (otherwise, $\varepsilon(y) < 1 - \beta$, and ε is unconstrained),

$$\begin{aligned} \varepsilon'(y)y &> \varepsilon(y)\{1 - \eta - \varepsilon(y)[\delta(p) - 1][1 - (1-\beta)\delta(p)]\} \\ &= \varepsilon(y)[1 - \eta] - [\eta - \varepsilon(y)][\varepsilon(y) - (1-\beta)\eta] \\ &= (1-\eta)[2\varepsilon(y) - (1-\beta)(1 + \eta - \varepsilon(y))] \\ &\quad - [1 - \varepsilon(y)][\varepsilon(y) - (1-\beta)] \geq \\ &\quad - [1 - \varepsilon(y)][\varepsilon(y) - (1-\beta)] \text{ if } 1 - \beta \leq \varepsilon(y) \leq 1. \end{aligned} \tag{A14}$$

Again, we see that the assumption of Lemma 1 is fulfilled. □

Proposition 1 (complementarity): Assume intersectoral (non-strict) complementarity: $s \leq 1$. Also, assume that the elasticity of intertemporal substitution σ (a C^1 function) satisfies, for any $\theta \in \mathscr{R}_{++}$,

$$0 < \sigma(\theta) < 1 \text{ and } \frac{\sigma'(\theta)\theta}{\sigma(\theta)} \geq -[1 - \sigma(\theta)]. \tag{A15}$$

Then, the first order and non-negative value condition (21) is necessary and sufficient for y to be the unique positive solution to the producer's problem (17), with d as defined by (11) and (12).

Proof: By Lemma 2, it suffices to show that assumption (A15) implies assumption (A8). Denoting

$$q \equiv \frac{\partial P}{\partial p}\frac{p}{P} = \frac{1}{m}\left(\frac{p}{P}\right)^{1-s} \in]0,1[,$$

we have $\delta = (1-q)s + q\Delta$ (see (15)), with $\Delta = \alpha\sigma + 1 - \alpha$ (see (13)), so that

$\delta < 1$ for any values of p/P and θ, and hence δ should always satisfy $\frac{\delta' p}{\delta} > \delta - 1$. First notice that

$$\begin{aligned}\frac{\Delta'\theta}{\Delta} &= \frac{\alpha\sigma'\theta + \alpha'\theta(\sigma - 1)}{\Delta} = \frac{\alpha\sigma'\theta - \alpha(1-\alpha)(1-\sigma)^2}{\Delta} \\ &\geq \frac{-\alpha(1-\sigma)[\sigma + (1-\alpha)(1-\sigma)]}{\Delta} = -(1-\Delta).\end{aligned} \tag{A16}$$

Also, with elasticity $\frac{dq}{dp}\frac{p}{q} = (1-s)(1-q)$, we obtain:

$$\begin{aligned}\frac{\delta' p}{\delta} &= \frac{q(1-q)(1-s)(\Delta - s) + q^2\Delta'\theta}{\delta} \\ &\geq \frac{q(1-q)(1-s)(\Delta - s) - q^2\Delta(1-\Delta)}{\delta} > -(1-\delta)\end{aligned} \tag{A17}$$

if

$$\begin{aligned}&q(1-q)(1-s)(\Delta - s) - q^2\Delta(1-\Delta) + \delta(1-\delta) \\ &= [-\Delta(2-3s) + s(1-2s)]q^2 + (\Delta - s)(2-3s)q + s(1-s) \\ &\equiv f(q) > 0.\end{aligned} \tag{A18}$$

That is precisely the case, since f is quadratic, $f(0) = s(1-s) \geq 0$, $f(1) = 0$, and $f'(1) = -\Delta(2-3s) - s^2 < 0$. □

Proposition 2 (substitutability): Assume intersectoral (non-strict) substitutability: $s \geq 1$. Also, assume that the elasticity of intertemporal substitution σ satisfies, for any $\theta \in \mathscr{R}_{++}$,

$$1 \leq \sigma(\theta) \leq \min\{s, 1+\beta\} \text{ and } \sigma'(\theta) \geq 0. \tag{A19}$$

Then, the first order and non-negative value condition (21) is necessary and sufficient for y to be the unique positive solution to the producer's problem ($\bar{Y} > 0$, and hence $n > 1$, is required in condition (21) if $s = 1$).

Proof: Again, by Lemma 2, it suffices to show that assumption (A19) implies assumption (A9). We have, in this case:

$$\begin{aligned}\Delta'\theta &= \alpha\sigma'\theta - \alpha(1-\alpha)(\sigma - 1)^2 \geq -\alpha(1-\alpha)(\sigma - 1)^2 \\ &= -(\Delta - 1)(\sigma - \Delta),\end{aligned} \tag{A20}$$

and hence

$$\frac{\delta' p}{\delta} = \frac{q(1-q)(s-1)(s-\Delta) + q^2\Delta'\theta}{\delta}$$

$$\geq \frac{q}{\delta}[(1-q)(s-1)(s-\Delta)-q(\Delta-1)(\sigma-\Delta)] \\ > -(\delta-1)[1-(1-\beta)\delta] \tag{A21}$$

for $(1-\beta)\delta \leq 1$, if (since $q/\delta < 1$)

$$\begin{aligned}(1-q)(s-1)(s-\Delta)-q(\Delta-1)(\sigma-\Delta)+(\delta-1)[1-(1-\beta)\delta] \\ = -(1-\beta)(s-\Delta)^2q^2 \\ +[(s-\Delta)(s-1+(1-2s)\beta)-(\Delta-1)(\sigma-\Delta)]q \\ +(s-1)(1+\beta s-\Delta)\equiv g(q)>0.\end{aligned} \tag{A22}$$

This inequality is satisfied. Indeed, g is quadratic and, using the assumption $\sigma \leq 1+\beta$, we have for $s>1$,

$$g(0)=(s-1)(1+\beta s-\Delta)\geq \beta(s-1)^2>0, \tag{A23}$$
$$g(1)=(\Delta-1)(1+\beta\Delta-\sigma)\geq \beta(\Delta-1)^2\geq 0.$$

If $\beta \leq 1$, g is concave, so that $g(q)>0$ for $q\in]0,1[$. If $\beta>1$, using the assumption $\sigma \leq s$,

$$g'(1)=-(s-\Delta)[(\beta-1)(2\Delta-1)+s]-(\Delta-1)(\sigma-\Delta)\leq \\ -(s-\Delta)(\beta-1+s)-(\Delta-1)(\sigma-\Delta)<0, \tag{A24}$$

and we get the same result. In the case $s=1$, since $1\leq\sigma\leq s$, so that $\Delta(\theta)\equiv 1$, we have, by (A21), $\delta'(\theta)\equiv 0$ and, by (A11) and (A12), $\varepsilon'(y)y=(1-\eta)\varepsilon(y)>0$ (for $\bar{Y}>0$ and hence $\eta<1$). Thus, condition (A2) is always satisfied and we can directly apply Lemma 1 in order to complete the proof. □

Lemma 3 (symmetry in quantities): Any temporary equilibrium is symmetric in quantities, inside each sector $k=1,\ldots,m$, relatively to the active firms: $y^*_{ki}=y^*_{kj}$ for any i and j in the same sector k, such that $y^*_{ki}>0$ and $y^*_{kj}>0$.

Proof: Consider the first order necessary condition (A10), and notice that it must be verified in equilibrium by any active firm for the same price p^* and with $y^*+\bar{Y}=d(p^*)$. The left-hand side of the equation is decreasing in y^* and the right-hand side is non-decreasing in the same variable if $\beta\geq 1$. Thus, there is at most one value of y^* satisfying the condition, the same for all active firms. The argument does not however apply to the case $\beta<1$. Suppose in this case that the first order and non-negative value condition (A10) is verified by two outputs y and y', such that $0<y<y'$. Dividing both sides of the equality which expresses for y the first order condition by the corresponding sides of the y' equality, we obtain:

$$\frac{1-y/\delta^*d^*}{1-y'/\delta^*d^*}=\left(\frac{y}{y'}\right)^{\beta-1}, \tag{A25}$$

with $\delta^* = \delta(p^*)$ and $d^* = d(p^*)$. Non-negativity of the profit resulting from y imposes:

$$1-\beta \leq \frac{y}{\delta^* d^*} = \frac{(y'/y)^{1-\beta}-1}{(y'/y)^{2-\beta}-1} \tag{A26}$$

or

$$1-\beta-(y/y')+\beta(y/y')^{2-\beta} \leq 0. \tag{A27}$$

The derivative with respect to y/y' of the expression on the left-hand side of this inequality, $-1+\beta(2-\beta)(y/y')^{1-\beta}$, is negative since y/y' and $\beta(2-\beta)$ are both less than 1. Thus, as $y/y' < 1$, that expression is always positive, and we get a contradiction. □

Lemma 4: In the substitutability case, under the assumptions of Proposition 2, the elasticity of the symmetric marginal revenue function $\rho' p/\rho$ is larger than the propensity to consume α. As a consequence, in this case and with non-increasing returns, the elasticity of the firms' reservation wage function $\omega_f' p/\omega_f$ is always larger than the elasticity of the labourers' reservation wage function $\omega_l' p/\omega_l$.

Proof: From (28) and (30),

$$\frac{\omega_f'}{\omega_f} p = (\beta-1)\Delta + \frac{\rho'}{\rho} p > \alpha = \frac{\omega_l'}{\omega_l} p, \tag{A28}$$

if $\beta \geq 1$ and $\rho' p/\rho > \alpha$. But, using (15), (26) and (A20),

$$\frac{\rho'}{\rho} p = 1 + \frac{\Delta' \theta}{m\delta(n\delta-1)} \geq 1 - \frac{m\alpha(1-\alpha)(\sigma-1)^2}{[(m-1)s+\Delta][n(m-1)s+n\Delta-m]}. \tag{A29}$$

Now, observe that the denominator of the last term in this expression is equal to

$$[(m-1)(s-1)+(\Delta-1)+m][n(m-1)(s-1)+n(\Delta-1)+(n-1)m] > m(\Delta-1)(s-1). \tag{A30}$$

Indeed, all terms in the left-hand side are non-negative, so that the inequality results from the fact that the coefficient of the term with $(\Delta-1)(s-1)$, $2n(m-1)$ is not smaller than m (as $m \geq 2$). Strictness is a consequence of either $s > 1$ (so that $n(m-1)^2(s-1)^2 > 0$) or $n > 1$ (so that $m^2(n-1) > 0$). Since $\Delta - 1 = \alpha(\sigma-1)$ and $\sigma \leq s$ by assumption, we obtain from (A29) and (A30):

$$\frac{\rho'}{\rho} p > 1-(1-\alpha)\frac{m\alpha(\sigma-1)^2}{m\alpha(\sigma-1)^2} = \alpha \tag{A31}$$

if $\sigma > 1$, and

$$\frac{\rho'}{\rho} p \geq 1 > \alpha \tag{A32}$$

if $\sigma = 1$. □

Proposition 6 (monotonicity): In the substitutability case, as characterized by Proposition 2, and under decreasing returns, there only exist a stationary equilibrium and, possibly, a continuum of equilibria with explosive inflation, associated with any $\theta_0 \in]0,1[$. Let $\underline{v}$ and $\bar{v}$ be defined as follows:

$$\begin{aligned} \underline{v} &\equiv (1/\beta)(L/mn)^{1/\beta - 1} \lim_{\theta \to 0} \{[1 - 1/n\delta(1,\theta)]U^*(\theta)\} \\ \bar{v} &\equiv (1/\beta)(L/mn)^{1/\beta - 1}[1 - 1/n\delta(1,1)]U^*(1). \end{aligned} \tag{A33}$$

- If $0 \leq v \leq \underline{v}$, full employment prevails in all intertemporal equilibria. The inflationary equilibria exist if and only if $\lim_{\theta \to 0} \alpha(\theta) = 1$.
- If $\underline{v} < v \leq \bar{v}$, the stationary equilibrium has full employment. The inflationary equilibria exist and are characterized by eventually declining output.
- If $\bar{v} < v$, underemployment prevails in all intertemporal equilibria. The inflationary equilibria exist and are characterized by declining output.

Under constant returns, the proposition applies to the case: $0 \leq v \leq \underline{v}$. No equilibrium exists if $\bar{v} < v$. If $v = \bar{v}$, there exists a stationary full employment equilibrium and a continuum of stationary underemployment equilibria. If $\underline{v} < v < \bar{v}$, there exists a full employment stationary equilibrium, a continuum of equilibria with constant inflation and declining output, and a countable infinity of inflationary equilibria with increasing inflation and full employment in a finite number of periods, constant inflation and declining output thereafter.

Proof: We have seen on pp.113ff. that the dynamics can be defined forwards, in full employment and, if $\beta > 1$, in underemployment, by explicit functions, Φ_{FE} and Φ_{UE} respectively, which are increasing, have 1 as their unique fixed point, and have at this point derivatives larger than 1. Hence, a sequence $(\theta_t)_{t \in \mathcal{N}}$ associated with any $\theta_0 > 1$ and determined by successive iterations of Φ_{FE} or Φ_{UE} will monotonically diverge, whatever the regime, implying that the corresponding sequence of prices converges to 0. But this is incompatible with the upper bound imposed by labour availability on output. Thus, all non-stationary equilibria start at some $\theta_0 \in]0,1[$ and are necessarily inflationary, with θ_t converging to 0 (since the stationary equilibrium is unstable, and no fixed point exists in $]0,1[$).

Now, consider the different cases associated with the potential values of v. If $v \leq \underline{v}$, the firms' reservation wage is larger than the workers' reservation wage at any θ, since $[1-1/n\delta(1,\theta)]U^*(\theta)$ is increasing (by Lemma 4, it has elasticity $\rho'\theta/\rho-\alpha>0$). Hence, full employment prevails for all θ, so that for $\beta \geq 1$ and for any $\theta_0 \in]0,1[$, the sequence $(\Phi_{FE}^t(\theta_0))_{t\in\mathcal{N}}$ of the images of θ_0 by the successive iterates of Φ_{FE} clearly characterizes an equilibrium if $\lim_{\theta\to 0} \alpha(\theta)=1$. Otherwise, the function Φ_{FE} is not defined over the whole interval $]0,1[$ (as one can easily check by looking at the difference equation characterizing full employment dynamics), and the stationary equilibrium is the only intertemporal equilibrium.

For $v>\underline{v}$, we can use the value $\tilde{\theta}(v)$ of the real interest factor that equalizes the two reservation wages at full employment:

$$v=(1/\beta)(L/mn)^{1/\beta-1}[1-1/n\delta(1,\tilde{\theta}(v))]U^*(\tilde{\theta}(v)), \tag{A34}$$

with a function $\tilde{\theta}$ which is increasing and verifies $\tilde{\theta}(\bar{v})=1$. If $v>\bar{v}$, underemployment prevails for any $\theta \leq 1$, since the right-hand side of this equation is increasing in θ. We can express, for $\beta>1$ and using the equality of firms' and workers' reservation wages, the output y_t as an increasing function of θ_t:

$$y_t=\left[\frac{1}{\beta v}[1-1/n\delta(1,\theta_t)]U^*(\theta_t)\right]^{\frac{1}{\beta-1}}. \tag{A35}$$

If $\beta=1$, so that $\tilde{\theta}(v)>1$ is the only value of the real interest factor that is compatible with the equality of the two reservation wages, no equilibrium can exist. But if $\beta>1$, the sequence $(\Phi_{UE}^t(\theta_0))_{t\in\mathcal{N}}$ of the images of θ_0 by the successive iterates of Φ_{UE} characterizes an equilibrium with explosive inflation and declining output, for any $\theta_0\in]0,1[$. Contrary to Φ_{FE}, the function Φ_{UE} is indeed always defined on the whole interval $]0,1[$. This results from: $\lim_{\theta\to 0}\omega_f(\theta,1)\theta^{1-\beta}/\omega_l(\theta,1)=0$ (implying $\lim_{\theta\to 0}\omega_f(\theta,1)/\omega_l(\theta,1)=0$), or equivalently from: $\lim_{\theta\to 0}[(1-\alpha(\theta))^{\beta-1}U^*(\theta)]$, which is always true, because either $\lim_{\theta\to 0}\alpha(\theta)=1$ or $\lim_{\theta\to 0}U^*(\theta)=0$. To see this, suppose that $\lim_{\theta\to 0}[U^{*\prime}(\theta)\theta/U^*(\theta)]=\lim_{\theta\to 0}[1-\alpha(\theta)]\in]0,1[$ (implying $\lim_{\theta\to 0}\sigma(\theta)=1$), with $\lim_{\theta\to 0}U^*(\theta)>0$. Then we get $\lim_{\theta\to 0}U^{*\prime}(\theta)\theta\in]0,\infty[$ and, by applying L'Hospital's rule, $\lim_{\theta\to 0}U^{*\prime}(\theta)\theta=-\lim_{\theta\to 0}U^{*\prime\prime}(\theta)\theta^2$, so that $\lim_{\theta\to 0}U^{*\prime\prime}(\theta)\theta/U^{*\prime}(\theta)=-1$. But

$$\lim_{\theta\to 0}U^{*\prime\prime}(\theta)\theta/U^{*\prime}(\theta)=\lim_{\theta\to 0}[-\alpha'(\theta)\theta/(1-\alpha(\theta))-\alpha(\theta)]$$

$$= \lim_{\theta \to 0} [\alpha(\theta)(\sigma(\theta) - 2)] = -\lim_{\theta \to 0} \alpha(\theta) > -1,$$

and we obtain a contradiction.

Finally, if $\underline{v} < v \leq \bar{v}$, any value of θ_t will be associated with full employment in period t if $\theta_t \geq \tilde{\theta}(v)$ and with underemployment if $\theta_t < \tilde{\theta}(v)$. Also, if $\theta_t \leq \tilde{\theta}(v)$, we can apply the underemployment dynamics, since we then get: $\Phi_{UE}(\theta_t) < \theta_t \leq \tilde{\theta}(v)$. Full employment dynamics can however only be used if $\theta_t \geq \Phi_{FE}^{-1}(\tilde{\theta}(v))$, so that if $\tilde{\theta}(v) < \theta_t < \Phi_{FE}^{-1}(\tilde{\theta}(v))$, we must consider the transition from the full employment in period t to the underemployment regime in period $t+1$. From the equalities $L = (mn)^{1-\beta} D(p_t, p_{t+1})^{\beta}$ and $\omega_f(p_{t+1}, p_{t+2}) = \omega_l(p_{t+1}, p_{t+2})$ (see Proposition 3), we straightforwardly obtain the difference equation:

$$[1 - \alpha(\theta_t)]\theta_t = [1 - \alpha(\theta_{t+1})] \times \left[\frac{1}{\beta v} \left(\frac{L}{mn} \right)^{1/\beta - 1} \left[1 - \frac{1}{n\delta(1, \theta_{t+1})} \right] U^*(\theta_{t+1}) \right]^{\frac{1}{\beta - 1}}. \tag{A36}$$

Both the left- and the right-hand sides of this equation are increasing in the real interest factor, so that forward dynamics are well defined in the interval $[\tilde{\theta}(v), \Phi_{FE}^{-1}(\tilde{\theta}(v))]$, and can be described by a function Φ_T such that $\Phi_T(\tilde{\theta}(v)) = \Phi_{UE}(\tilde{\theta}(v))$ and $\Phi_T \circ \Phi_{FE}^{-1}(\tilde{\theta}(v)) = \tilde{\theta}(v) = \Phi_{FE} \circ \Phi_{FE}^{-1}(\tilde{\theta}(v))$, as can be easily checked from (A36) and the corresponding difference equations for the two pure regimes (see also (A34)). We thus obtain a continuous increasing function Φ, which coincides with Φ_{UE} in $]0, \tilde{\theta}(v)]$, with Φ_T in $[\tilde{\theta}(v), \Phi_{FE}^{-1}(\tilde{\theta}(v))]$, and with Φ_{FE} in $[\Phi_{FE}^{-1}(\tilde{\theta}(v)), 1]$ (see figure 5.5). Since this function has no fixed point in $]0,1[$ any sequence starting in this interval and determined by the successive iterations of Φ converges to 0 and characterizes an equilibrium with explosive inflation and declining output.

If $\beta = 1$, the only admissible value of θ in the underemployment regime is $\tilde{\theta}(v)$. A continuum of equilibria results from the possibility of taking $\theta_0 = \tilde{\theta}(v)$ with any p_0 at least as high as the full employment price corresponding to $\tilde{\theta}(v)$. A transition between regimes is still possible, if $\theta_t = \Phi_{FE}^{-1}(\tilde{\theta}(v))$, so that an equilibrium may start with full employment at any $\theta_0 = \Phi_{FE}^{-k}(\tilde{\theta}(v))$, for any $k \in \mathcal{N}$. We obtain inflationary equilibria, with full employment in a finite number $k+1$ of periods, and declining output afterwards. Inflation is constant once the underemployment phase is attained. □

Notes

We wish to thank Jean-Pascal Bénassy, Huw Dixon and Henri Sneessens for helpful comments and the Banco de Portugal for its hospitality. Support from the European Commission HCM programme and from the Belgian PAI Program (SPPS) are gratefully acknowledged.

1. See, for example, Gordon (1990) or Mankiw and Romer (1991).
2. See the recent surveys of Dixon and Rankin (chapter 2, this volume) and Silvestre (1993, 1994).
3. This is already the case with Hart (1982), where the presence of an oligopolistic labour market is a necessary condition for the occurrence of unemployment. But the same kind of remark applies for instance to the more recent work of Jacobsen and Schultz (1991) or Schultz (1992), transposing Hart's framework to overlapping-generations models.
4. Examples are Blanchard and Kiyotaki (1987) and Ball and Romer (1991).
5. See Akerlof and Yellen (1985b).
6. For a presentation of this concept, see d'Aspremont *et al.* (1991b).
7. The importance of changes in mark-ups, induced in particular by changes in aggregate demand, as a source of endogenous fluctuations and other macroeconomic phenomena, has been stressed by Rotemberg and Woodford (1991, 1992, 1993). See also Portier (1994).
8. We do not rely on fixed costs to incorporate increasing returns in the model, as much of the related literature (of which Weitzman, 1982, Blanchard and Kiyotaki, 1987 and Rotemberg and Woodford, 1992 are but a few examples). Neither do we consider free entry equilibria, as this literature in general does.
9. See, for instance, Cooper and John (1988). Multiplicity of equilibria related to demand conditions has already been considered by Heller (1986). Multiplicity due to increasing returns to scale of production or market participation is present, for example, in Kiyotaki (1988), Chatterjee and Cooper (1989) and Manning (1990).
10. See Grandmont (1985) and Guesnerie and Woodford (1992).
11. The influence on unemployment of a negative marginal revenue has been analysed, in the context of different models, by d'Aspremont *et al.* (1984, 1989a, 1989b and 1990), Dehez (1985), and Silvestre (1990). The main results concern 'cooperation', not 'coordination' problems, if we adopt the classification introduced by Silvestre (1993). Both problems appear however in d'Aspremont *et al.* (1990). The corresponding results in an overlapping generations model, which may be associated with endogenous fluctuations, are given in d'Aspremont *et al.* (1991a).
12. See d'Aspremont *et al.* (1991b, 1992). An alternative assumption, leading to the so-called *Cournot–Walras equilibrium* (Gabszewicz and Vial, 1972), would make the producer conjecture the quantities produced by all other firms, whether in the same sector or not, implying that the whole system of sectoral demand functions is used to compute the inverse demand function.
13. The Cobb–Douglas case ($s=1$) is discussed on pp.104–5 below.
14. By imposing, for instance, that, for any $\theta \in \mathscr{R}_{++}$, $s \leq \sigma(\theta) < 1$ (in the complementarity case), or $1 \leq \sigma(\theta) < s$ and $(1-\beta)s < 1$ (in the substitutability case).
15. This is an avowedly improper term, meaning in fact the wage that sustains p^* as an output market equilibrium. It appears however that transactions in the labour market proceed, as we should expect, until the firms' reservation wage (the buyers' reservation price) becomes smaller than the workers' reservation wage (the sellers' reservation price), or until labour endowment is exhausted (see Proposition 3).

16. There are other early and independent references in the second class, for instance Mankiw (1985) or Svensson (1986).
17. However, Mankiw (1985) considers monopolistic behaviour in the output markets, with a competitive labour market. Subsequent models of Ball and Romer (1990) or Dixon (1987) also stress, at least in a first stage, product markets imperfections (with suppliers setting prices of differentiated goods, or playing Cournot on the basis of conjectural variations, respectively).
18. A temporary monetary equilibrium framework with explicit intertemporal decisions and expectations formation is used, in the context of Hart's (1982) model, by d'Aspremont *et al.* (1984, 1989a, 1989b) and Rankin (1992, 1995) and, in the line of Weitzman's (1985) model, by d'Aspremont *et al.* (1990).
19. See d'Aspremont (1989a, 1989b, 1990, 1991a). Lack of separability or homotheticity in the consumers' preferences, just as inexistence of a representative consumer, should not prevent us from taking Ford effects into account, even if the construction of an objective demand appears more complex in that case: see, for instance, Bénassy (1987).
20. However, some recent macroeconomic overlapping-generations models introduce imperfect competition in output markets, whether in a Cournotian approach, as Schultz (1992), or under monopolistic competition, as Bénassy (1991a, 1991b), or both, as d'Aspremont *et al.* (1991a).
21. As already mentioned in the Introduction, we do not impose the number of products to be very large, as we should in a model of pure monopolistic competition. By considering Cournotian competition instead, close substitutes may be subsumed in the same sector, and the number of sectors may accordingly be kept relatively small.
22. Substitutability is of course not the only way to ensure uniqueness. Hart (1982) uses (together with non-increasing returns) the assumption of an increasing marginal revenue function ρ and the absence of labour disutility, so that $\omega_l' p/\omega_l = 0 < (\beta - 1)\Delta + \rho' p/\rho$.
23. As one can easily check, the constants in the reservation wage functions, are:
$$B = \frac{1}{\beta}\left[\frac{M}{(1-\alpha)mn}\right]^{1-\beta}\left[1 - \frac{m/n}{(m-1)s+1}\right] \text{ and } A = \frac{v\hat{P}^{1-\alpha}}{\alpha^{\alpha}(1-\alpha)^{1-\alpha}}.$$
24. The graphs in figure 5.2 have been plotted for the parameter values: $\alpha = 0.75$, $\beta = 0.25$, $A = 1$ (entailed for instance by $\hat{P} = 0.75$ and $v \simeq 0.612$), $B = 1$ (corresponding for instance to $n = 1$, $m = 20$, $s = 4/3$ and $M \simeq 5.265$), and $p^{**} = 0.5$ (resulting, given the other parameter values, from $L \simeq 24.093$).
25. By applying the same type of argument, based on an elasticity of the workers' reservation wage function larger than the elasticity of the firms' reservation wage function, at least for arbitrarily high prices, the previous result may be extended, under supplementary assumptions, to the complementarity case.
26. We are using the same type of assumption as Heller (1986).
27. In the two equalities, we use the fact that, for $\sigma < 1$, α increases with θ from 0 to 1, so that Δ tends to 1 when the price tends to 0, to the same limit as σ when the price tends to infinity. Only if σ converges to 1 for a price tending to zero (infinity) can the limit of α be larger than zero (smaller than 1), but Δ will still have the stated limit, 1 (the same as σ), in that case.

28. The graphs in figure 5.3 correspond to the case of a CES utility function with $\sigma = 0.5$ and a discount factor for future consumption $\gamma = 0.81$. Other parameter values are $s=0.15$, $v \simeq 0.0108$, $m=10$, $n=5$, $\beta=1$, $\hat{P}=0.5$ and $p^{**}=0.1$ (resulting from $M/L \simeq 0.0668$).
29. See (A16) in the appendix (proof of Proposition 1).
30. The price sequence (p_t) is monotonic in the substitutability case because θ_t cannot switch from $]0,1[$ to $]1,\infty[$ or vice versa. On the contrary in the complementarity case θ_t, when different from 1, takes alternatively values in each one of the two intervals, so that prices are alternatively increasing and decreasing.
31. Indeed, using the equality $\alpha'\theta = \alpha(1-\alpha)(1-\sigma)$, it is easy to calculate: $\alpha(\theta) = [1 + e(\theta)]^{-1}$, with $e(\theta) = \theta^{-1} \exp \int (\sigma(\theta)/\theta) d\theta$. Hence, from $\lim_{\theta \to 0} \alpha(\theta) \in]0.1[$, so that (using L'Hospital's rule) $\lim_{\theta \to 0} e(\theta) = \lim_{\theta \to 0} [e(\theta)\sigma(\theta)] \in]0,\infty[$, we obtain $\lim_{\theta \to 0} \sigma(\theta) = 1$.
32. The first possibility has been shown in d'Aspremont *et al.* (1991a).
33. Existence is also compatible with constant returns if marginal revenue is positive at $\tilde{\theta}(v)$, as defined by (59).
34. A simple example, with a decreasing demand elasticity, is given by a CES intertemporal utility function, with elasticity of substitution $\sigma < m/n - (m-1)s < 1$ and a discount factor for future consumption γ such that $\gamma^{\sigma} < [m/n - (m-1)s - \sigma]/[1 - m/n + (m-1)s]$.
35. A simple example is given in d'Aspremont *et al.* (1991a).
36. Indeed, the stationary state is unstable in the full employment dynamics if $\Delta(1) > 1/2$ (see (49)), and any sequence of real interest factors generated by (48) would be unbounded, since it cannot converge to a cycle. But (48) does not have a solution in θ_{t+1} for a too large θ_t since $\lim_{\theta \to \infty} [\theta(1-\alpha(\theta))]$ belonging to $]0,1[$ implies $\lim_{\theta \to \infty} \sigma(\theta) = 0$ (as can be seen using L'Hospital's rule and the equality $\alpha'\theta = \alpha(1-\alpha)(1-\sigma)$).

6 Macroeconomic externalities

Andrew John

Introduction

In recent years, researchers have made a great deal of progress in the quest for Keynesian models with rigorous microeconomic foundations. It is now well understood that macroeconomic models incorporating imperfect competition or search frictions can give rise to 'Keynesian features', such as inefficiency, multipliers and coordination failures, without recourse to unexplained price rigidities. The seminal contributions are Hart (1982) for models with imperfect competition, and Diamond (1982) for search models.[1]

The ability of both types of model to exhibit Keynesian features can be understood in terms of an underlying structural similarity: both imperfect competition and search in macroeconomic models lead to strategic complementarities.[2] But this similarity does not imply that the models are perfect substitutes, either in terms of interpretation or in terms of the questions that they can address. Search frictions are particularly plausible in labour markets, and allow for explicit discussion of unemployment. Imperfect competition, conversely, is appealing for analysis of output markets; some empirical work supports this view.[3]

Thus it comes as no surprise – indeed it is desirable – that researchers continue to investigate both paradigms. It *is* perhaps more surprising that these two branches of the literature still coexist in relative isolation from one another. One goal of the current work is to facilitate communication between these approaches by further identifying some of their formal similarities and substantive differences in a framework where both distortions can be understood as a source of externality. Noting that some studies in the literature discuss their results intuitively in terms of 'macroeconomic externalities', a second aim of the chapter is to suggest how this intuition might be formalized.[4]

Apart from the intellectual satisfaction of identifying similarities between apparently diverse areas, there may be more concrete payoffs. When mappings are made from one line of research to another, it is possible to import results, preventing the rediscovery of the wheel and perhaps generating new insights. Related to this, the identification of formal parallels improves dialogue among researchers: search and imperfect competition models appear dissimilar in part because they use different terminology to describe related phenomena. Perhaps most importantly, if both avenues of research have attractions, then models incorporating both effects may be desirable, and there is then a need for analysis applicable to both market and non-market interactions between agents.[5]

The basic framework is established in the second section, elaborated in the third section, and applied in the fourth section to strategic market games, imperfect competition under quantity-setting and price-setting, and search. The fifth section contains concluding comments.

The basic framework

Preferences and endowments

Consider the following characterization of an exchange economy. There are IN agents: I agents in each of N sectors. Denote the set of agents in sector n as I_n. There are $N+1$ goods, indexed by $n=0,\ldots,N$. Throughout the chapter, agents and the sectors they inhabit are identified by subscripts, while goods are identified by superscripts. Agents have initial endowments

$$\{\omega_{in}\}=\{\omega_{in}^0,\ldots,\omega_{in}^N\}$$
$$\omega_{in}^m=0,\ m\neq 0,\ n;$$
$$\omega_{in}^m>0,\ m=0,\ n;\ \text{all } i,\ n,\ m.$$

That is, agents are endowed with good zero and the good associated with their own sector. They therefore supply the good of their own sector and demand the goods of other sectors.[6] Good zero is taken as the numéraire; its main purpose is to facilitate the accounting. The presentation of the basic framework is simplified by a number of symmetry assumptions, although most of the results derived in the chapter do not rely on symmetry. Under symmetry, $\omega_{in}^m=\bar{\omega}$, all i, m, and $\omega_{im}^0=\bar{\omega}^0$.

All agents have preferences defined over their consumption of the $N+1$ goods which can be represented by a continuous, differentiable, and quasi-concave utility function:

$$U(c^0,\ldots,c^N).$$

Assume that this utility function is unchanged by a relabelling of all goods except good zero and good n for an agent in sector n. That is, the goods of all other sectors enter symmetrically into the utility function.[7] Also, assume this utility function is identical for all agents, up to a relabelling of sectors. Again, these symmetry assumptions are not essential for most of what follows.

More generally, it could be the case that agents' preferences are defined over the consumption of all goods by *all* agents, so that agent i in sector n has preferences given by

$$U(c_{in}^{0},\ldots,c_{in}^{N};\ c_{jn}^{0},\ldots,c_{jn}^{N},\ c_{km}^{0},\ldots,c_{km}^{N}),$$

$j \neq i$, $m \neq n$. The first $N+1$ arguments in this utility function represent the agent's own consumption, while the arguments after the semicolon are the consumption of other agents in sector n, and other agents in all other sectors. The arguments after the semicolon correspond to the familiar external effects of microeconomic theory and are a feature of the primitives of an economy: once the set of commodities is defined, it is possible to establish the presence or absence of such externalities simply by examining the preferences. That such externalities depend upon the commodity space was noted by Arrow (1969), who made the simple but elegant argument that it is possible to define a market for any given externality; if all external effects are priced, then the standard results of equilibrium theory apply.[8] The focus of the current work is explicitly *not* on such externalities. Instead, the emphasis here is on external effects that arise from the institutions of trade – again, once the commodity space is specified.

Allocation mechanisms

The reallocation of goods in this economy is governed by an allocation mechanism that maps from agents' strategies into reallocations of goods. Specifically, suppose that agent i in sector n controls an N-element vector of strategies $s_{in}=\{s_{in}^{m}\}$, $m=1,\ldots,N$. That is, there is a strategy associated with each good except for the numéraire. Sometimes, it is helpful to interpret these strategies as supplies and demands. An allocation mechanism is a collection of continuous and differentiable functions that determine the allocations of all goods to all agents, as functions of the strategies chosen by all agents. Thus, in general, a mechanism is a collection of $(N+1)$ functions per agent (or $NI \times (N+1)$ in total), each of which possesses IN^2 arguments. For example, for agent i in sector n,

$$\begin{aligned} a_{in} &= \{a_{in}^{m}\},\ m=0,\ldots,N \\ &= \{a_{in}^{m}(s_{in},\ s_{jn},\ s_{km},\ s_{hr})\}. \end{aligned}$$

This agent's allocation of good m depends, in general, on her own choice of strategies (s_{in}), on the strategies of other agents in her sector (s_{jn}), on the strategies of agents in sector m (s_{km}), and on all other strategies (s_{hr}). Assume that this allocation function is symmetric.[9]

Agent i's consumption is then given by

$$c_{in}^{m} = \omega_{in}^{m} + a_{in}^{m}.$$

At this point the precise nature of a mechanism is left intentionally vague; in particular, a mechanism might be a reduced form representation of a game that involves a number of stages and other strategic or non-strategic decisions.

An allocation is feasible if $\sum_i \sum_n a_{in}^{m} \leq 0$, all $m = 0, \ldots, N$. An allocation is just feasible if this condition holds with strict equality for all m. An allocation is attainable under a given mechanism if there is some set of strategies that result in that allocation. A mechanism is feasible if all allocations attainable under the mechanism are feasible.[10]

The allocation mechanism and utility functions can be combined to form continuous and differentiable payoff functions for each agent: $\sigma_{in}(s_{in}, s_{jn}, s_{km})$.[11] (The payoff function also obviously depends on the endowments of all agents; this dependence is suppressed for the sake of clarity.) The economies analysed below are thus fully described by endowments, preferences, and an allocation mechanism. Since the chapter focuses on non-cooperative behaviour, the equilibrium concept adopted is Nash.

Although an agent's payoff function possesses IN^2 arguments in general, the symmetry assumptions on the utility and allocations functions imply that these can be grouped into seven subsets, the members of which enter symmetrically into the payoff function. Write

$$\sigma_{in} = \sigma_{in}(s_{in}^{n},\ s_{in}^{m},\ s_{jn}^{n},\ s_{jn}^{m},\ s_{kn}^{n},\ s_{km}^{m},\ s_{km}^{r}),\ m \neq n,\ r \neq m, j \neq i.$$

Interpreting strategies as demands and supplies, this implies that the payoff of agent i in sector n depends upon: her own supply of good n (1 element); her demands for other goods ($N-1$ elements); her competitors' supply of good n ($I-1$) elements; her competitors' demand for all other goods ($(I-1)(N-1)$ elements); all other agents' demands for good n ($I(N-1)$ elements); all other agents' supplies of goods ($I(N-1)$ elements); and all other strategies ($I(N-1)(N-2)$ elements). In many of the examples below, matters are even simpler (for example, there is normally no need to distinguish between s_{jn}^{m} and s_{km}^{r}).

Externalities

Mechanism externalities

To keep the exposition simple, this section treats the strategy of an agent, s_{in}, as a scalar rather than a vector. Each equation derived below should more properly be viewed as a vector of equations.

Although the framework established thus far explicitly rules out the standard consumption externalities of microeconomic theory, whereby one agent's consumption directly affects another's utility, interdependencies between agents remain. Define a *mechanism externality* to arise whenever one agent's strategy affects the allocation of another agent:

Mechanism externality: $\partial a_{km}/\partial s_{in} \neq 0$, $m \neq n$, or $m = n$ and $k \neq i$.

Since consumption externalities are assumed absent, a mechanism externality must be present if one agent's strategy affects the payoff of another agent:

$$\partial \sigma_{km}/\partial s_{in} \neq 0.$$

Mechanism externalities capture at least some of the features of external effects in recent macroeconomic models. The idea behind such externalities is that the institutions governing exchange give rise to interdependencies absent in the primitives of the economy. There is of course no presumption that the presence of mechanism externalities implies any inefficiency in the allocation of resources. For example, interdependencies arise through the price system in a Walrasian economy. But in many recent macroeconomic models, the institutions of trade and exchange are a source of inefficiency. For mechanism externalities to be a useful analytical construct, it is therefore necessary to isolate when interdependencies give rise to inefficiency.

An allocation is efficient if no feasible allocation Pareto dominates it, and an allocation is constrained efficient if no attainable allocation Pareto dominates it. As this chapter is concerned with the properties of mechanisms of exchange, the main focus throughout is on constrained efficiency. A constrained-efficient allocation can be characterized as the solution to a planner's problem, where the planner faces the same mechanism as the agents, but can dictate all agents' strategies. The planner solves

$$\max_{\{s\}} \sum_i \sum_n \lambda_{in}\sigma_{in}(s_{in}, s_{jn}, s_{km}).$$

The set of constrained-efficient solutions is obtained by varying the weights, λ_{in}.

The first order conditions from this planner's problem are

$$\lambda_{in}\left(\frac{\partial \sigma_{in}}{\partial s_{in}}\right) + \sum_j \lambda_{jn}\left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right) + \sum_m \sum_k \lambda_{km}\left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right) = 0, \text{ all } i, n, j \neq i, m \neq n. \quad (1)$$

These equations constitute a set of necessary conditions for constrained efficiency of an allocation.[12]

Lemma: If (i) the planner weights all agents equally; or (ii) the utility functions, allocation functions and endowments are symmetric and the planner restricts attention to symmetric strategy choices; or (iii) the planner can effect transfers of numéraire and utility is normalized appropriately; then a necessary condition for a Nash equilibrium of this model to be constrained efficient is

$$\sum_j \left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right)+\sum_m \sum_k \left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right)=0, \text{ all } i, n, j \neq i, m \neq n. \tag{2}$$

Proof: The first order conditions for individual agents in Nash equilibrium are

$$\left(\frac{\partial \sigma_{in}}{\partial s_{in}}\right)=0, \text{ all } i, n,$$

so the necessary conditions for the planner's problem reduce to

$$\sum_j \lambda_{jn}\left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right)+\sum_m \sum_k \lambda_{km}\left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right)=0, \text{ all } i, n, j \neq i, m \neq n. \tag{3}$$

(i) If the planner treats all agents symmetrically, then $\lambda_{jn}=\lambda_{km}=\lambda$, all j, k, m, n, and the result is immediate.

(ii) Summing the planner's necessary conditions yields

$$\sum_i \sum_n \left[\sum_j \lambda_{jn}\left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right)+\sum_m \sum_k \lambda_{km}\left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right)\right]=0, j \neq i, m \neq n.$$

These terms can be rearranged to yield

$$\sum_i \sum_n \lambda_{in}\left[\sum_j \left(\frac{\partial \sigma_{in}}{\partial s_{jn}}\right)+\sum_m \sum_k \left(\frac{\partial \sigma_{in}}{\partial s_{km}}\right)\right]=0, j \neq i, m \neq n.$$

Given symmetric choices of strategies and given the symmetry of the utility function and the allocation mechanism,

$$\left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right)\equiv\left(\frac{\partial \sigma_{in}}{\partial s_{jn}}\right);$$

$$\left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right)\equiv\left(\frac{\partial \sigma_{in}}{\partial s_{km}}\right);$$

and the result follows.

(iii) Let t_{in} denote the planner's transfer of the numéraire to agent i in sector n, so $\sigma_{in}=\sigma_{in}(s_{in}, s_{jn}, s_{km}; t_{in})$.[13] The planner chooses these transfers subject to the constraint that $\sum_i \sum_n t_{in} = 0$, yielding the additional first order conditions

$$\lambda_{in}\left(\frac{\partial \sigma_{in}}{\partial t_{in}}\right) = \mu,$$

where μ is the multiplier on the constraint. The planner therefore chooses transfers such that the marginal utility of the numéraire for each agent is inversely proportional to the weight placed on that agent in the planner's social welfare function. Thus

$$\sum_j \left(\frac{1}{\partial \sigma_{jn}/\partial t_{jn}}\right)\left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right) + \sum_m \sum_k \left(\frac{1}{\partial \sigma_{km}/\partial t_{km}}\right)\left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right) = 0, \text{ all } i, n. \tag{4}$$

But, without loss of generality, utility functions can be normalized at the given allocation such that $\partial \sigma_{in}/\partial t_{in}$, and hence λ_{in}, is constant for all i, n. The result follows. □

The Lemma establishes

$$\sum_j \left(\frac{\partial \sigma_{jn}}{\partial s_{in}}\right) + \sum_m \sum_k \left(\frac{\partial \sigma_{km}}{\partial s_{in}}\right) = 0, j \neq i, m \neq n, \text{ all } i, n,$$

as a necessary condition for efficiency in a number of cases. The condition simply states that the *aggregate* mechanism externality associated with a change in agent i's strategy must equal zero. The Lemma essentially presents conditions under which it makes sense to sum the externalities in this way.

Case (i) is obvious because the planner, by assumption, treats all agents symmetrically. Case (ii) applies when there is sufficient symmetry in the model: if the planner can choose only symmetric allocations, and given sufficient symmetry, then the planner again ends up treating all agents symmetrically.[14]

The interpretation of case (iii) is a little more delicate. The fourth section of the chapter examines the equilibria of a number of different models and considers whether or not the necessary conditions for constrained efficiency (1) are satisfied. If, given an equilibrium allocation, there exist some $\{\lambda_{in}\}$ (the weights on utility) such that (3) hold, then (1) are satisfied. Case (iii) suggests that a natural set of weights to examine is given by the inverse of the marginal utility of the numéraire, as in (4). But, of course, even if (4) are not satisfied, there could still be other weights on utility such that (3) hold. In other words, (4) are sufficient for (1) to hold, but not necessary for (1) to hold. If, however, the planner has access to transfers of the numéraire, then (4) are necessary and sufficient. (One interpretation of (4) is that they identify conditions for the absence of *actual* Pareto improvements if the planner can effect compensating transfers, and for the absence of *potential* Pareto improvements if the planner cannot.) Finally, given an equilibrium allocation, one can without loss of generality normalize utility functions such that all agents have identical marginal utility of the numéraire *at that allocation*, in which case (4) is equivalent to (2) given in the Lemma.[15]

The key observation is that efficiency does not require that mechanism externalities should be absent, but merely that they net out in the aggregate.[16] The remainder of the chapter makes extensive use of this observation. Define

$$\text{Distributional externality: } \sum_j \left(\frac{\partial\sigma_{jn}}{\partial s_{in}}\right) + \sum_m \sum_k \left(\frac{\partial\sigma_{km}}{\partial s_{in}}\right) = 0, \text{ all } i, n,$$

$$\text{and } \sum_{k,m\in H} \partial\sigma_{km}/\partial s_{in} \neq 0, \text{ some } i, n, k, m;\ i, n \notin H;$$

$$\text{Real externality: } \sum_j \left(\frac{\partial\sigma_{jn}}{\partial s_{in}}\right) + \sum_m \sum_k \left(\frac{\partial\sigma_{km}}{\partial s_{in}}\right) \neq 0, \text{ some } i, n.$$

There is a real externality associated with an agent's strategy if it results in a net change in the utility of others. There is a distributional externality if the change in strategy does not lead to a net change in total utility, but does affect the aggregate utility of some subset (H) of other agents. The idea is simply that external effects result sometimes in efficiency gains or losses, and sometimes in redistribution (for example, movements along the Pareto frontier).[17]

The intuition behind real and distributional externalities is straightforward, and provides a motivation for the framework adopted here: when there are no real externalities, agents acting in their own self-interest behave as a planner would wish. The planner cannot obtain any net improvements in utility by adjusting strategies in the neighbourhood of equilibrium. Note that the result requires only that there be no real externalities in equilibrium.

Pecuniary externalities

The literature on externalities has long recognized that there is a danger in associating interdependency and externality because, as already noted, not all interdependency is a sign of inefficiency. In the literature, such topics are often discussed under the heading of 'pecuniary externalities'. The term is avoided for the most part in this chapter because there is no apparent professional consensus on its meaning.

If one reads, for instance, Bohm (1987) or Graaff (1987) in the New Palgrave Dictionary, one is left feeling that the term 'pecuniary externalities' has never had a clear, generally accepted meaning. (Silvestre, 1991)

Indeed, Silvestre identifies four different usages in the literature.

In particular, some authors apparently use the term 'pecuniary externality' for those interdependencies that are not 'real' – that is, not a source of inefficiency – while others apparently use the term for any interdependency operating through prices – which in some cases can be a source of inefficiency.[18] While these two notions are closely related, they are not the same. Greenwald and Stiglitz (1986) and Laffont (1987), among others, have correctly noted that externalities operating through prices will in general have real consequences in economies where prices play more than an allocative role (as can occur, for example, under imperfect information).

One interpretation of the current chapter is that it generalizes these ideas of pecuniary externality to both market and non-market settings. The fourth section examines when interdependencies through prices do or do not give rise to inefficiency and likewise examines when non-market interdependencies do or do not give rise to inefficiency. The discussion here is perhaps in the spirit of Graaff's (1987) suggestion that 'clarity would be served by . . . speaking instead of technological interdependence (of production functions) on the one hand, and market independence (via the price system) on the other'.

As the analysis thus far has been in terms of abstract mechanisms, it is necessary to be precise about the interpretation of a 'price' in this setting. For future reference, let p^m be the price of good m in terms of the numéraire if, for every agent, the mechanism allocates $-p^m$ units of the numéraire to the agent for every unit of good m. Thus, if all goods have associated prices, $a_{in}^0 = -\sum_m p^m a_{in}^m$, all i, n. A mechanism with this property forces agents to satisfy their budget constraints.[19]

Macroeconomic externalities

Some recent work on the microfoundations of macroeconomics has interpreted Keynesian phenomena in terms of a macroeconomic externality arising from imperfect competition or search. Interpreting Keynesian economics in terms of externality is not new.[20] Like 'pecuniary externality', however, the term 'macroeconomic externality' apparently has no clear definition in the literature (see Silvestre, 1991, for some discussion). Macroeconomic models with search or imperfect competition have at least two features in common: the underlying distortion in the economy comes from the institutions governing trade, be they markets or matching processes, and the models are general equilibrium.

The notion of mechanism externalities introduced in this chapter provides a framework for analysing models with such features. In particular, the following section uses the framework to show that models of search, imperfect competition under price-setting, and imperfect competition under quantity-setting, are all characterized in general by real mechanism externalities. Such externalities therefore provide a means to highlight both the similarities and the distinct features of these models. This chapter thus provides a framework for thinking in a rigorous way about macroeconomic externalities.

The chapter does *not* claim, however, that mechanism externalities and macroeconomic externalities are the same thing. The aim here is not to provide a definition of macroeconomic externalities, but it seems reasonable to suggest that any candidate definition should probably include both notions mentioned earlier: a macroeconomic externality is one that arises from the institutions of trade and has a general equilibrium element. Mechanism externalities as defined here, however, can and in some cases do include external effects that have no particular macroeconomic (that is, general equilibrium) implications. Indeed, it is shown below that such is the case in a general equilibrium model of imperfect competition under quantity-setting. A definition of macroeconomic externality might therefore include the provision that the external effect be mediated through some aggregate variable, such as aggregate demand (see, for example, Blanchard and Kiyotaki, 1987 or Shleifer and Vishny, 1988).

Finally, this chapter certainly does not advocate that mechanism externalities necessarily provide the best way of thinking about search, imperfect competition, or other market failures. The claim is much more modest: imperfect competition and search can sometimes usefully be interpreted in terms of externalities, and this may aid our understanding of certain macroeconomic models.

Externalities from search and imperfect competition

This section of the chapter shows that imperfect competition and search give rise to mechanism externalities. Five mechanisms are presented. The first, provided primarily for motivation, is a simple Walrasian mechanism in which no mechanism externalities arise. The second is a strategic market game, which does exhibit mechanism externalities in general. These are followed by two general equilibrium models of imperfect competition: in one, suppliers choose the amount of goods that they wish to supply to the market and prices are determined by demand in competitive markets; in the other, suppliers set prices and quantities are determined by demand in competitive markets. In the final mechanism, trading frictions imply that agents must expend resources to find a trading partner. They search and, if matched, engage in mutually beneficial exchange.

The remainder of this section sets out each mechanism and shows the presence or absence of mechanism externalities.[21] In each case, the analysis proceeds by characterizing an equilibrium allocation (that is, it is assumed that individuals' first order conditions are satisfied) and looking for the presence of mechanism externalities at that allocation. Without loss of generality, utility functions are normalized throughout so that all agents have marginal utility of the numéraire equal to one at the allocation under consideration.[22]

A Walrasian mechanism

Consider the following mechanism: for all i, n,

$$a_{in}^m = -s_{in}^m, \; m = 1, \ldots, N;$$

$$a_{in}^0 = \sum_{m=1}^{N} p^m s_{in}^m.$$

Here, the prices $\{p^m\}$ are taken as parametric by all agents, and the mechanism is constructed so that agents satisfy their budget constraints.

Suppose that the p^ms are the Walrasian prices – that is, the prices that would be called out by the Walrasian auctioneer. Then it is evident that the game with this mechanism has a Nash equilibrium that yields the Walrasian outcome.[23] The game does not exhibit mechanism externalities, because each agent's strategies have no effect on others' allocations. Unfortunately, this is not a feasible mechanism. While the Walrasian allocation is feasible, there are many other choices of strategies that would imply infeasible allocations.

In real economies, feasibility evidently cannot be violated. If the demands and supplies expressed by agents are not mutually compatible, then something has to give. There are at least three possibilities. First, given quantities, prices may adjust to ensure market-clearing. Second, quantities may adjust at given prices. Third, feasibility may be guaranteed by agents' trading on a one-on-one basis. All these possibilities arise in the following subsections.

Imperfect competition: strategic market games

Mechanism externalities can be illustrated simply in strategic market games of the type developed and analysed by Dubey, Shubik and others.[24] These games also provide a useful introduction to macroeconomic models of imperfect competition. In a strategic market game, agents' strategies are bids and offers, which can be interpreted as supplies and demands.

Assume that there are two sectors ($N=2$), denoted by 1 and 2. Strategies represent supplies: agent i in sector 1 supplies good 1 ($s_{i1}^1>0$) and demands good 2 ($s_{i1}^2<0$). The terms on which agents eventually trade depend upon the relative supplies and demands by agents on each side of the market. The allocation mechanism is:

$$\begin{aligned}
a_{i1}^1 &= -s_{i1}^1 \\
a_{i1}^2 &= -s_{i1}^2/p^2 \\
a_{i1}^0 &= p^1 s_{i1}^1 + s_{i1}^2;\ i \in I_1. \\
a_{k2}^1 &= -s_{k2}^1/p^1 \\
a_{k2}^2 &= -s_{k2}^2 \\
a_{k2}^0 &= s_{k2}^1 + p^2 s_{k2}^2;\ k \in I_2.
\end{aligned}$$

Strategic market games ensure feasibility by a suitable definition of prices. Specifically, prices are given by the ratio of total demand for the good to total supply of the good. Let S_m^n equal the total effective supply of good n by agents in sector m; $n, m=1, 2$. That is,

$$\begin{aligned}
S_1^1 &= \sum_i s_{i1}^1 > 0; \\
S_1^2 &= \sum_i s_{i1}^2 < 0; \\
S_2^1 &= \sum_k s_{k2}^1 < 0; \\
S_2^2 &= \sum_k s_{k2}^2 > 0.
\end{aligned}$$

Now define

$$\begin{aligned}
p^1 &= -S_2^1/S_1^1; \\
p^2 &= -S_1^2/S_2^2.
\end{aligned}$$

It is easily confirmed that this mechanism satisfies feasibility. Note also that p^1 and p^2 do satisfy the definition of prices from p. 147. This mechanism will in general imply the existence of real externalities, as the following proposition shows.

Proposition 1: In this strategic market game, in equilibrium, (1) an increase in an agent's strategy variable imposes negative externalities on agents in the same sector and positive externalities on agents in the other sector; (2) agents impose real externalities in general; (3) when the number of agents in each sector is large, agents impose distributional externalities in symmetric equilibrium.

Proof: (1) Suppose that agent i in I_1 changes her strategy with respect to good 1:

$$\frac{\partial \sigma_{j1}}{\partial s_{i1}^1} = \left(\frac{\partial \sigma_{j1}}{\partial p^1}\right)\left(\frac{\partial p^1}{\partial s_{i1}^1}\right)$$
$$= \left(\frac{\partial U}{\partial a_{j1}^0}\right)(s_{j1}^1)\left(\frac{-p^1}{S_1^1}\right)$$
$$= (s_{j1}^1)\left(\frac{-p^1}{S_1^1}\right) < 0;\ j \in I_1, j \neq i,$$

using the normalization of the utility function.

$$\frac{\partial \sigma_{k2}}{\partial s_{i1}^1} = \left(\frac{\partial \sigma_{k2}}{\partial p^1}\right)\left(\frac{\partial p^1}{\partial s_{i1}^1}\right)$$
$$= \left(\frac{\partial U}{\partial a_{k2}^1}\right)\left(\frac{s_{k2}^1}{(p^1)^2}\right)\left(\frac{-p^1}{S_1^1}\right)$$
$$= \left(\frac{s_{k2}^1}{p^1}\right)\left(\frac{-p^1}{S_1^1}\right) > 0;\ k \in I_2.$$

(This derivation uses the first order condition from agent k's optimization with respect to s_{k2}^1: $\partial U/\partial a_{k2}^1 = p^1(\partial U/\partial a_{k2}^0) = p^1$.)

(2) Summing the external effects across all agents yields

$$\sum_{j \neq i} \frac{\partial \sigma_{j1}}{\partial s_{i1}^1} = \left(1 - \frac{s_{i1}^1}{S_1^1}\right)(-p^1);$$
$$\sum_k \frac{\partial \sigma_{k2}}{\partial s_{i1}^1} = p^1.$$
$$\Rightarrow \sum_{j \neq i} \frac{\partial \sigma_{j1}}{\partial s_{i1}^1} + \sum_k \frac{\partial \sigma_{k2}}{\partial s_{i1}^1} = \left(\frac{s_{i1}^1}{S_1^1}\right)(p^1) > 0.$$

(3) In symmetric equilibrium,

$$\sum_{j \neq i} \frac{\partial \sigma_{j1}}{\partial s_{i1}^1} = \left(1 - \frac{1}{I}\right)(-p^1);$$
$$\sum_k \frac{\partial \sigma_{k2}}{\partial s_{i1}^1} = p^1.$$

In the limit as $I \to \infty$, it follows that

$$\sum_{j \neq i} \frac{\partial \sigma_{j1}}{\partial s_{i1}^{1}} = -p^{1};$$

$$\sum_{k} \frac{\partial \sigma_{k2}}{\partial s_{i1}^{1}} = p^{1}.$$

$$\Rightarrow \sum_{j \neq i} \frac{\partial \sigma_{j1}}{\partial s_{i1}^{1}} + \sum_{k} \frac{\partial \sigma_{k2}}{\partial s_{i1}^{1}} = 0.$$

Analogous results hold for changes in the other strategy variables. □

By increasing the supply of good 1, and therefore reducing its price, agent i in sector 1 imposes a negative externality on other agents in sector 1, and bestows a positive externality on agents in sector 2. These externalities do not in general sum to zero. The reason is that agent i possesses some market power, and so internalizes some of the effect of the change in the price of good 1. Imperfect competition therefore provides one example of a real externality that arises from changes in prices (a real pecuniary externality), and the inefficiency of imperfect competition can be interpreted as a consequence of this externality. Only when the number of agents in a sector becomes large do the externalities cancel out.

It is perhaps noteworthy that, even when the number of agents is large, so that each individual agent is of negligible size in the market, there are still distributional externalities. As the number of agents becomes large, agent i's actions have a vanishing effect on any given other agent, but agent i's actions also affect a larger number of agents. Thus an arbitrarily small agent still has a real effect on a subset of agents. Strategic market games yield an efficient allocation when the market power on both sides of all markets goes to zero.[25] From the perspective of this chapter, efficiency results not from the fact that small agents have negligible effects on prices, but rather from the fact that small agents impose equal and opposite externalities on both sides of the market.[26] Finally, note that externalities need not disappear if the number of sectors, rather than the number of agents within a sector, is increased.

Imperfect competition: quantity-setting

Imperfect competition with quantity-setting closely resembles the strategic market game just analysed. In the strategic market game, however, agents possess both monopoly and monopsony power, whereas here agents are assumed to have some market power in terms of the good that they supply, but no market power on the demand side of other markets. The simplest assumption that generates this result is to suppose that the number of sectors (N) is arbitrarily large, but that the number of agents within a sector (I) is small.[27] Following some of the imperfect competition models in the literature, the model is presented as a two-stage game. This aids the exposition, since it allows a focus on the supply decisions of agents with market power. It should be emphasized, though, that the assumption on timing is irrelevant for the results of the model.

In the first stage, agents in a given sector (that is, suppliers) choose their level of output (that is, the amount that they wish to sell), taking as given the output of others in the same sector and the inverse demand curve faced by the sector. Thus agent i in sector n chooses s_{in}^n, taking as given $p^n(s_{in}^n;s_{jn}^n, s_{km}^m)$, $i, j \in I_n, j \neq i$, $k \in I_m$, $m \neq n$. The presence of s_{km}^m in this demand function reflects the fact that the output supplied in other sectors determines the incomes of agents in those sectors and the price of goods in those sectors, which in turn influences spending on good n. In the second stage, agents demand the goods of other sectors in competitive markets, taking as given their income from the first stage ($p^n s_{in}^n$) and the prices of other goods. Prices adjust in the second stage to clear markets, given these demands and given the quantities supplied in the first stage. Second-stage market-clearing thus guarantees feasibility.

The allocation mechanism is, for all i, n,

$$a_{in}^m = -s_{in}^m, \; m = 1, \ldots, N;$$

$$a_{in}^0 = \sum_{m=1}^{n} p^m s_{in}^m;$$

$$p^n(s_{in}^n;\, s_{jn}^n,\, s_{km}^m),\; i, j \in I_n, j \neq i,\; k \in I_m$$

where $s_{in}^n > 0$ and $s_{in}^m < 0$, $m \neq n$. As mentioned above, the focus here is on the first-stage supply decisions (s_{in}^n), rather than the second-stage demand decisions (s_{in}^m).

Consider the second-stage maximization of agent i in sector n. She has already chosen s_{in}^n in the first stage. In the second stage she chooses s_{in}^m, $m \neq n$, to solve

$$\max U(\omega_{in}^0 + p^n s_{in}^n + \sum_{m \neq n} p^m s_{in}^m,\; -s_{in}^m,\; \omega_{in}^n - s_{in}^n).$$

Now write the solution to this problem as an indirect utility function (suppressing the endowments):

$$\sigma_{in} = V_{in}(p^n(.)s_{in}^n, -s_{in}^n, p^m(.)).$$

In the first stage, agent i in sector n chooses s_{in}^n to maximize this indirect utility function, taking as given the prices in the other sectors, the strategies of other agents in sector n, and the demand curve facing sector n.

The claim that agent i takes prices in other sectors as given actually requires elaboration. Define $V_{in,2}$ to be the partial derivative of the indirect utility function with respect to its second argument. The first order condition from the agent's problem is then:

$$p^n + (s^n_{in})\left(\frac{\partial p^n}{\partial s^n_{in}}\right) - V_{in,2} + \sum_{m \neq n} (s^m_{in})\left(\frac{\partial p^m}{\partial s^n_{in}}\right) = 0$$

(using the normalization of the utility function to set the derivative of indirect utility with respect to income equal to one, and using Roy's identity to set the derivative of indirect utility with respect to price equal to (minus) demand). The first three terms in this expression constitute the standard first order condition for imperfect competition: in the absence of the final term they could be rearranged to give price as a mark-up over marginal cost ($V_{in,2}$), where the size of the mark-up depends upon the elasticity of demand in the usual way.

The final term of the first order condition reflects general equilibrium effects: an increase in the supply of good n by agent i influences demands in the second stage of the game and so will affect market-clearing prices in other sectors. Changes in the prices of goods in other sectors in turn influence the welfare of agent i when that agent purchases goods in the second stage. The claim that firms take prices in other sectors as given when making their first-stage supply decisions is really equivalent to the claim that the final term of this first order condition can be set to zero.

Under symmetry, when there is a large number of sectors, a change in any given price has a negligible effect on agent i's welfare because that agent purchases only a small quantity of that good.[28] But because these effects must be summed over a large number of goods, it does not immediately follow that the final term in the first order condition is vanishingly small for large N. Rather, this term goes to zero if and only if a change in agent i's supply of good n has a negligible effect on other prices.

There are two linkages whereby agent i's supply of good n affects other prices. First, an increase in the supply of good n reduces the price of that good, and so affects the demand for other goods. The size and magnitude of this effect depends upon the elasticity of substitution between good n and other goods. Second, an increase in the supply of good n changes the total revenue earned in that sector, which in turn affects spending in other sectors. The size and magnitude of this effect depends upon the elasticity of demand for good n. Under the assumption that the number of sectors is large, and given symmetry assumptions on the utility function, the effects of a change in s_{in}^n are spread evenly across all sectors, and so the effect on the price in any single sector should indeed be negligible.[29]

Note, finally, that there are some circumstances under which agent i's actions have no effect on other prices even with a small number of sectors – for example, if utility functions are Cobb–Douglas over all goods except the good that an agent supplies. Then all demand curves are unit-elastic, and the demand for a given good depends only on income and the price of that good. In particular, the change in the price of good n resulting from agent i's actions does not affect the demand curve in other sectors. Moreover, since the demand curve for good n is also unit-elastic, the change in agent i's supply does not affect the total revenue earned within sector n, and so there is no income effect on demand curves in other sectors either. In this case, $\partial p^m/\partial s_{in}^n = 0, m \neq n$.[30] (In fact, a two-sector version of the model under these assumptions is substantively identical to the strategic market game considered in the previous subsection.)

Proposition 2: This quantity-setting game exhibits real mechanism externalities.

Proof: Suppose that agent $i \in I_n$ changes her strategy. Then

$$\frac{\partial \sigma_{jn}}{\partial s_{in}^n} = (s_{jn}^n)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{m \neq n} (s_{jn}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right);$$

$$\frac{\partial \sigma_{km}}{\partial s_{in}^n} = (s_{km}^n)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) + \sum_{h \neq m} (s_{km}^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right)$$
$$= \sum_h (s_{km}^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right),$$

again using the normalization of utility and the properties of the indirect utility function. Agent i's actions affect the welfare of other agents through two channels. First, and most obviously, an increase in her supply of good n reduces the price of good n, which affects the welfare both of other suppliers and of consumers of good n. Second, there are the general equilibrium effects noted earlier: an increase in the supply of good n will in general affect market-clearing prices in other sectors.[31]

Summing the external effects across all agents yields

$$\sum_{j \neq i} (s_{jn}^n)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{j \neq i} \sum_{m \neq n} (s_{jn}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) + \sum_{m \neq n} \sum_k \sum_h (s_{km}^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right).$$

Define

$$S_m^h = \sum_k s_{km}^h, \; k \in I_m, \text{ all } m, h.$$

That is, S_m^h represents the total supply of good h by agents in sector m. Note that $S_m^m > 0$ and $S_m^h < 0$, $h \neq m$. The total externality is therefore

$$(S_n^n - s_{in}^n)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{m \neq n} (S_n^m - s_{in}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) + \sum_{m \neq n} \sum_h (S_m^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right).$$

Now market-clearing implies that $S_m^m = -\sum_{h \neq m} S_m^h \Rightarrow \sum_m S_m^h = 0$, all h. So the total externality equals

$$\begin{aligned}
&\left(S_n^n - s_{in}^n + \sum_{m \neq n} S_m^n\right)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{m \neq n} (S_n^m - s_{in}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) + \sum_{m \neq n} \sum_{h \neq n} (S_m^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right) \\
&= \left(\sum_m S_m^n - s_{in}^n\right)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{m \neq n} (S_n^m - s_{in}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) + \sum_{h \neq n} \left(\sum_m S_m^h - S_n^h\right)\left(\frac{\partial p^h}{\partial s_{in}^n}\right) \\
&= (-s_{in}^n)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) + \sum_{m \neq n} (S_n^m - s_{in}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right) - \sum_{h \neq n} (S_n^h)\left(\frac{\partial p^h}{\partial s_{in}^n}\right) \\
&= (-s_{in}^n)\left(\frac{\partial p^n}{\partial s_{in}^n}\right) - \sum_{m \neq n} (s_{in}^m)\left(\frac{\partial p^m}{\partial s_{in}^n}\right).
\end{aligned}$$

But, from the first-order condition, this simply equals

$$p^n - V_{in,2}. \quad \square$$

As in the analysis of the strategic market game, an increase in agent i's supply to the market reduces the price and so bestows a positive externality on the consumers of the good (that is, agents in other sectors) and a negative externality on other producers of the good. These two externalities do not cancel out because agent i has some market power and so internalizes some of the effect of the price change. There is thus an external effect equal to the difference between price and marginal cost. (As in the strategic market game, the externality would disappear if the number of agents in a sector (I) became large.)

The striking feature of Proposition 2 is that there is no further component to the externality, even though agent i's actions affect prices in other sectors and hence affect the welfare of other agents. While the effect of agent i's actions on any given price may be negligible, these effects must be summed over both a large number of agents and a large number of goods, so one might have expected general equilibrium consequences. But there are none.[32]

It is worth reiterating that the result does *not* come from the fact that agent i's actions have a negligible effect on the prices in other sectors. While the effect on any individual price goes to zero at the rate $(1/N)$, the number of agents affected by a given price change increases at the rate N. Instead, the reason that agent i's actions have no general equilibrium consequences is that the changes in prices in other sectors impose distributional externalities only. If, for example, the price in sector m rises as a result of agent i's actions, this benefits the suppliers of good m but imposes an equal and opposite cost on the demanders of good m. Market-clearing conditions permit the cancelling of the general equilibrium external effects, leaving only the standard partial equilibrium distortion.[33]

Imperfect competition: price-setting

Now consider a model of price-setting with differentiated products, so that there are N sectors with one firm in each sector ($I=1$). Because the quantity supplied by this agent is *not* a strategic variable, let q^n (rather than s^n_{1n}) denote the supply of good n by the agent in sector n (for brevity, refer to this agent as agent n). As before, let p^n denote the price of good n; this is now the strategy variable of agent n.

In the first stage, each firm (agent) chooses its price, taking as given the prices of all other firms and the demand curve it faces. Thus agent n chooses p^n, taking as given $q^n(p^n;p^m)$, $m \neq n$.[34] In the second stage, agent n chooses her demands for other goods, $\{-s^m_n\}$. The allocation mechanism is, for all n,

$$a^m_n = -s^m_n,\ m \neq n$$
$$a^n_n = -q^n(p^n,p^m)$$
$$a^0_n = \sum_{m=1}^{N} p^m s^m_n + p^n q^n(.),\ m \neq n;$$

where $q^n(.) > 0$ and $s^m_n < 0$, $m \neq n$.

Consider the second-stage maximization of agent n. She has chosen p^n in the first stage. In the second stage she chooses s^m_n, $m \neq n$, to solve

$$\max\ U(\omega^0_n + p^n q^n + \sum_{m \neq n} p^m s^m_n,\ -s^m_n,\ \omega^n_n - q^n).$$

As before, write the solution to this problem as an indirect utility function:

$$\sigma_n = V_n(p^n q^n(.),\ -q^n(.);\ p^m).$$

In the first stage, agent n chooses p^n to maximize this indirect utility function, taking as given the prices in the other sectors and the demand curve facing sector n. The first order condition from this problem is

$$\left[q^n + p^n\left(\frac{\partial q^n}{\partial p^n}\right)\right] - V_{n,2}\left(\frac{\partial q^n}{\partial p^n}\right) = 0,$$

using the normalization of the utility function. Rewrite this as

$$p^n - V_{n,2} = -\left(\frac{q^n}{\partial q^n/\partial p^n}\right).$$

Proposition 3: This price-setting game exhibits real mechanism externalities.

Proof: Consider a change in agent n's strategy. Then

$$\frac{\partial \sigma_m}{\partial p^n} = [p^m - V_{m,2}]\left(\frac{\partial q^m}{\partial p^n}\right) + s_m^n,\ m \neq n,$$

again, using Roy's identity and the normalization of utility. Summing over all agents, and using the market-clearing condition $-\sum_{m \neq n} s_m^n = q^n$,

$$\sum_{m \neq n} \frac{\partial \sigma_m}{\partial p^n} = -q^n + \sum_{m \neq n} [p^m - V_{m,2}]\left(\frac{\partial q^m}{\partial p^n}\right). \quad \square$$

The first term is the negative externality from imperfect competition: if agent n charges a price that is one unit higher, buyers will lose income equal to the total amount of the good traded. The second term reflects the general equilibrium consequences of the change in the price of good n. As in the quantity-setting game, a change in one agent's strategy will in general cause demand curves to shift in other sectors. Under quantity-setting, these general equilibrium repercussions are changes in prices in other sectors that result in distributional externalities only. But under price-setting, the general equilibrium consequences take the form of changes in quantities, which in general have real effects. Thus, if a change in p^n causes q^m to rise, this leads to an external benefit equal to the difference between price and marginal cost in sector m. Summing these externalities over all m gives the general equilibrium component of the externality.[35] Since real general equilibrium effects are present under price-setting but not under quantity-setting, price-setting models may have better claim to the term 'macroeconomic externality'.

Under quantity-setting, as noted, agents impose externalities of different sign on other agents, depending on whether they are on the same or the other side of the market. That is, increased output of a given good benefits purchasers of that good and imposes cost on other suppliers. The analogous result is harder to identify in models of price-setting. In, say, a standard differentiated product model then, from one perspective, each agent is a monopolist, and so there are no other agents on the same side of the market. The direct effect of an increase in price is simply the imposition of externalities on purchasers. But, from another perspective, goods are substitutes and other agents are competitors, so an increase in the price of one good will benefit suppliers of substitutes. Similar issues actually also arise under quantity-setting with multiple sectors: with Cobb–Douglas preferences and sufficient symmetry, there is a clear distinction between competitors (in the same sector) and customers (in other sectors) but, more generally, agents producing different goods in other sectors may still be viewed as competitors.

The situation is further complicated by the fact that, in these general equilibrium models, all agents are both suppliers and demanders. For example, in the price-setting model of Blanchard and Kiyotaki (1987), all goods are substitutes, so agents, acting as producers, are competitors. But since they purchase each other's goods, all agents are in some sense simultaneously on the same and the opposite sides of the market. The distinction is perhaps not very meaningful in this setting, although the externalities of this chapter could be used to *define* whether agents were, on net, on the same or the opposite sides of a market.[36] Another approach to differentiated products is to assume that goods are located on an attribute circle.[37] The dividing line between those on the same and the other sides of the market would then arise endogenously; the former would be nearby on the circle and the latter far away.

Search

In search models, agents expend resources in the attempt to find trading partners. Return to a two-sector model and suppose that, in stage 1, agents in I_1 search for agents in I_2, and vice versa.[38] If two agents are successfully matched, they trade in the second stage; otherwise they simply consume their endowment. Assume that each agent controls a single strategy variable, called search intensity. The probability of matching depends upon the search intensities chosen by all agents. It is assumed that the costs of search are incurred in terms of the good that an agent supplies.[39] When two agents are successfully matched, Nash bargaining occurs, resulting in transactions that divide the gains from trade. Feasibility is thus assured by the fact that agents trade on an individual basis. Let $G_{i1,k2}$ denote the gain from trade of agent $i \in I_1$ if she is matched with agent $k \in I_2$; by Nash bargaining, $G_{i1,k2} \equiv G_{k2,i1}$.[40] Denote the probability that these two agents meet by $\pi_{i1,k2} \equiv \pi_{k2,i1}$ and define the expected gain from search for agent i as:

$$V_{i1} = \sum_k \pi_{i1,k2} G_{i1,k2}.$$

The allocation mechanism can be written as follows.

With probability $\pi_{i1,k2}$

$$a^1_{i1} = -t^1_{i1,k2} - s_{i1}$$
$$a^2_{i1} = -t^2_{i1,k2}$$
$$a^0_{i1} = -t^0_{i1,k2};$$

and with probability $(1-\sum_k \pi_{i1,k2})$,

$$a^1_{i1} = -s_{i1}$$
$$a^2_{i1} = 0$$
$$a^0_{i1} = 0.$$

In this mechanism, $t^m_{i1,k2}$ represents the amount of good m that agent i will agree to transfer to agent k in the event that they are matched (so $t^m_{i1,k2} = -t^m_{k2,i1}$). These are optimal values chosen so as to divide the gains from trade evenly. Since the search process consumes some of the goods that agents supply, and so changes agents' endowments at the time when they trade, these agreed transactions in general depend upon the level of search that is chosen by agents i and k. That is, the gains from trade depend upon the search intensities, $\{s_{i1}, s_{k2}\}$, providing one source of externality.[41]

It is a useful simplification, however, to assume that utility functions are linear and separable in the good that the agent supplies. Thus for agent i in sector 1,

$$U(c^0_{i1}, c^1_{i1}, c^2_{i1}) = u(c^0_{i1}, c^2_{i1}) + \theta c^1_{i1}.$$

The gains from trade are then given by

$$\begin{aligned} G_{i1,k2} &= u(\omega^0_{i1} - t^0_{i1,k2}, -t^2_{i1,k2}) + \theta(\omega^1_{i1} - s_{i1} - t^1_{i1,k2}) \\ &\quad - u(\omega^0_{i1}, 0) - \theta(\omega^1_{i1} - s_{i1}) \\ &= u(\omega^0_{i1} - t^0_{i1,k2}, -t^2_{i1,k2}) - u(\omega^0_{i1}, 0) - \theta t^1_{i1,k2}, \end{aligned}$$

and so are independent of search effort. The cost of search is simply θs_{i1}. If it is further assumed that there is symmetry of endowments and utility functions, the gain from search will be an identical constant for all agents (it is actually sufficient to assume symmetry within each sector, since Nash bargaining equates the gain across sectors). It follows that

$$\sigma_{i1} = V_{i1} - \theta s_{i1} = \sum_k \pi_{i1,k2} G_{i1,k2} = \pi_{i1} G - \theta s_{i1},$$

where $\pi_{i1} = \sum_k \pi_{i1,k2}$ and G is the constant gain from search. This formulation allows a focus on the external effects from the matching technology.

Externalities arise from search because an agent's probability of being matched depends upon the search activity of all agents:

$$\pi_{i1,k2}=\pi_{i1,k2}(s_{i1},s_{j1},s_{k2},s_{h2});\ i,j\in I_1;\ k,\ h\in I_2;\ j\neq i;\ k\neq h.$$

Search models in the literature generally assume that this function is increasing in s_{i1} and s_{k2}, and decreasing in s_{j1} and s_{h2}. That is, it is assumed that increases in own search effort and increases in the search intensity of potential partners serve to increase the likelihood of being matched, while increases in the effort of others on the same side of the market lead to a lower probability of matching. It is also assumed that an increase in any agent's strategy increases the expected number of matches.

Proposition 4: (1) Agents in this model impose negative mechanism externalities on others on the same side of the market and positive or negative externalities on those on the other side of the market. (2) If matching functions are symmetric ($\pi_{i1,k2}(.)=\pi(.)$, all $i\in I_1$, $k\in I_2$), then agents impose positive externalities on agents on the other side of the market. (3) Agents impose real externalities unless the marginal disutility of search equals twice the aggregate benefit imposed on agents on the other side of the market.

Proof: (1) Suppose that agent i in I_1 changes her search strategy. Then

$$\frac{\partial\sigma_{j1}}{\partial s_{i1}}=G\left(\frac{\partial\pi_{j1}}{\partial s_{i1}}\right)=G\left(\sum_k\left(\frac{\partial\pi_{j1,k2}}{\partial s_{i1}}\right)\right)<0,\ j\neq i;$$

$$\frac{\partial\sigma_{k2}}{\partial s_{i1}}=G\left(\frac{\partial\pi_{k2}}{\partial s_{i1}}\right)=G\left(\left(\frac{\partial\pi_{i1,k2}}{\partial s_{i1}}\right)+\sum_{j\neq i}\left(\frac{\partial\pi_{j1,k2}}{\partial s_{i1}}\right)\right)\gtrless 0,\ i,j\in I_1;\ k\in I_2.$$

The first term in the second equation is positive and the second term is negative, establishing (1). (2) The expected number of matches, which (by assumption) is increasing in s_{i1}, equals $\sum_k\sum_j\pi_{j1,k2}=\sum_k\pi_{k2}$. Given identical matching functions, it follows that π_{k2} must be increasing in s_{i1} for each individual $k\in I_2$. To prove (3), note that since $\pi_{j1,k2}\equiv\pi_{k2,j1}$ and $G_{j1,k2}\equiv G_{k2,j1}=G$, the aggregate expected gain from matching must be identically equal for agents in I_1 and I_2:

$$V_{i1}+\sum_{j\neq i}V_{j1}\equiv\sum_k V_{k2},\ i,j\in I_1;\ k\in I_2$$

$$\Rightarrow\ \pi_{i1}+\sum_{j\neq i}\pi_{j1}\equiv\sum_k\pi_{k2}$$

$$\Rightarrow \frac{\partial \pi_{i1}}{\partial s_{i1}} + \sum_{j \neq i} \left(\frac{\partial \pi_{j1}}{\partial s_{i1}} \right) = \sum_{k} \left(\frac{\partial \pi_{k2}}{\partial s_{i1}} \right).$$

Using this result and the first order condition $G(\partial \pi_{i1}/\partial s_{i1}) = \theta$, the aggregate externality associated with a change in s_{i1} is

$$\begin{aligned}
\sum_{j \neq i} \left(\frac{\partial V_{j1}}{\partial s_{i1}} \right) + \sum_{k} \left(\frac{\partial V_{k2}}{\partial s_{i1}} \right) &= G \left(\sum_{j \neq i} \left(\frac{\partial \pi_{j1}}{\partial s_{i1}} \right) + \sum_{k} \left(\frac{\partial \pi_{k2}}{\partial s_{i1}} \right) \right) \\
&= G \left(2 \left(\sum_{k} \left(\frac{\partial \pi_{k2}}{\partial s_{i1}} \right) \right) - \frac{\partial \pi_{i1}}{\partial s_{i1}} \right) \\
&= 2G \left(\sum_{k} \left(\frac{\partial \pi_{k2}}{\partial s_{i1}} \right) \right) - \theta.
\end{aligned}$$

Thus there are distributional externalities associated with a change in s_{i1} if

$$2G \left(\sum_{k} \left(\frac{\partial \pi_{k2}}{\partial s_{i1}} \right) \right) = \theta$$

and there are real externalities otherwise. □

Proposition 4 shows that, in search models as in imperfect competition models, agents impose negative externalities on those on the same side of the market, and (under symmetry) positive externalities on those on the opposite side of the market. These results derive directly from standard assumptions on matching functions. While many search models are two-sector models (often involving search by workers and firms), some models in the literature, such as Diamond (1982), have the feature that all agents in the economy are potential trading partners. These models thus resemble some of the price-setting examples discussed earlier (for example, Blanchard and Kiyotaki, 1987) in that all other agents benefit from an increase in one agent's search activity. Such a model will exhibit real positive externalities unless agent i's marginal return to search, excluding search costs, is zero, implying that there will be real externalities unless marginal search costs are zero.

There is a noteworthy parallel between search models and models of quantity-setting. To see this, define $\tilde{p}^{i1}$ to be the expected gain per unit of search:

$$\tilde{p}^{i1} = V_{i1}/s_{i1} = \frac{\pi_{i1} G}{s_{i1}}.$$

Agent i's payoff is thus

$$\sigma_{i1} = \tilde{p}^{i1} s_{i1} - \theta s_{i1}.$$

There is now an analogy with the quantity-setting model. Under quantity-setting, agents in sector 1 give up s_{i1}^1 units of good 1 and obtain $p^1(s_{i1}^1)s_{i1}^1$ units of the numéraire. In the search model, agents in sector 1 give up s_{i1} units of good 1 and obtain $\tilde{p}^{i1}(s_{i1})s_{i1}$. Under imperfect competition, agents' strategies influence the price at which they trade; in the search game, agents' actions influence their probability of trading. From the perspective of agent i, $\tilde{p}_{i1}$ is analogous to the price faced by a quantity-setting agent.

Note, however, that $\tilde{p}_{i1}$ is *not* a price in the sense defined on p.147: not all agents exchange good 1 for the numéraire at this rate. This distinction implies that there are differences as well as similarities between search and quantity-setting. Under quantity-setting, an increase in the strategy of agent i in sector 1 leads to a fall in p^1 that benefits agents in other sectors. The total benefit to agents in other sectors is matched by an equal and opposite loss to the agents (including i) in I_1. There is an externality to the extent that some of the effects of a price change are internalized by agent i, when that agent has market power. An increase in agent i's strategy in the search game similarly generates increases in π_{k2} ($\tilde{p}^{k2}$) that benefit agents in I_2. These, however, are matched by an equal *gain* to the agents (again including i) in I_1.

As a consequence, real search externalities do not disappear in general even when the number of agents in each sector is large. To illustrate this, consider the class of matching functions described by[42]

$$\pi_{i1}(s_{i1}, s_{j1}, s_{k2}) = (s_{i1}/S_1)f(S_1, S_2);\ S_1 = \sum_j s_{j1},\ S_2 = \sum_k s_{k2}.$$

It follows that $\tilde{p}^{i1} = Gf(S_1, S_2)/S_1$. If there is a large number of agents in each sector, then agent i has a negligible effect on $\tilde{p}^{i1}$, and so the return to search is linear in search effort. The first order condition for agent i is simply $\tilde{p}^{i1} = \theta$. Yet though this is reminiscent of the equating of price and marginal cost, there are still real externalities in general, as Proposition 5 demonstrates.[43]

Proposition 5: Assume that the matching function is of the form just specified and that there is a large number of agents in each sector. Then there are no real externalities if: (1) $f(S_1, S_2) = A(S_1S_2)^{1/2}$, where A is an arbitrary constant; or (2) $f(.)$ is constant returns and $S_1 = S_2$.

Proof: As a preliminary, note that the expected number of matches is given by

$$\sum_i \pi_{i1} = \sum_i (s_{i1}/S_1)f(S_1, S_2) = f(S_1, S_2),$$

and since $\sum_i \pi_{i1} = \sum_k \pi_{k2}$, it follows that $f(.)$ is symmetric:

$$f(S_1, S_2) \equiv f(S_2, S_1).$$

From Proposition 4, there are real externalities unless

$$\frac{\partial \pi_{i1}}{\partial s_{i1}} = 2\left(\sum_k \left(\frac{\partial \pi_{k2}}{\partial s_{i1}}\right)\right).$$

Under the assumption that I is large, agent i has a negligible effect on S_1, so

$$\frac{\partial \pi_{i1}}{\partial s_{i1}} = \frac{f(S_1, S_2)}{S_1}.$$

Also

$$\frac{\partial \pi_{k2}}{\partial s_{i1}} = \left(\frac{s_{k2}}{S_2}\right)\left(\frac{\partial f(S_2, S_1)}{\partial S_1}\right)$$

which implies there will be real externalities unless

$$\frac{f(S_1, S_2)}{S_1} = \frac{f(S_2, S_1)}{S_1} = 2\left(\frac{\partial f(S_2, S_1)}{\partial S_1}\right).$$

Solving this differential equation implies

$$f(S_1, S_2) = f(S_2, S_1) = A(S_1 S_2)^{1/2},$$

proving (1). To prove (2), note that since $f(.)$ is symmetric, $\partial f(S_2, S_1)/\partial S_1 = \partial f(S_2, S_1)/\partial S_2$ when these functions are evaluated at $S_1 = S_2$. The condition for no real externalities when $S_1 = S_2$ can thus be written

$$f(S_2, S_1) = S_1\left(\frac{\partial f(S_2, S_1)}{\partial S_1}\right) + S_2\left(\frac{\partial f(S_2, S_1)}{\partial S_2}\right). \quad \square$$

Absence of real externalities in this setting therefore requires either that the matching function be constant returns and that there be symmetry, or that the matching function is Cobb–Douglas, in which case there are no real externalities even when the equilibrium is not symmetric.[44] A simple illustration comes from the matching function

$$f(S_1, S_2) = (S_1^\rho + S_2^\rho)^{\alpha/\rho}.$$

Real externalities are absent in this case when $S_2^\rho = (2\alpha - 1)S_1^\rho$, which is satisfied for arbitrary S_1, S_2 only in the Cobb–Douglas case of $\rho = 0$ and $\alpha = 1$. Under symmetry ($S_1 = S_2$), however, the condition is satisfied provided only that $\alpha = 1$, which corresponds to constant returns.

The assumptions of Proposition 5 imply that individual agents are small and so have no effect on the gain from matching. By analogy with the quantity-setting model, it might be thought that this would imply efficiency. Moreover, since agent i is small relative to the economy, the effect of her search effort on the probability of any other agent's being matched goes to zero. As in the other examples, however, large I also means that more agents are affected by agent i's actions, so the aggregate effect is not zero. As with price-setting, and unlike the quantity-setting game, there are no market-clearing conditions guaranteeing that externalities cancel out in the limit.

Conclusion

Recent work on the microfoundations of macroeconomics interprets Keynesian phenomena in terms of 'macroeconomic externalities' arising from imperfect competition or search. Unfortunately, the term apparently has no clear definition in the literature. Rather than offering a precise definition, this chapter simply suggests that macroeconomic externalities provide a convenient way to characterize certain distortions that are associated with the general equilibrium implications of market or non-market mechanisms of exchange. Distortions arising from search and imperfect competition can be understood as externalities in a general framework capable of incorporating both market and non-market interactions.

The studies on search and imperfect competition in macroeconomics literature exist in relative isolation from one another, in part probably because they discuss similar phenomena using different terminology. One conclusion of this chapter is that models with search and imperfect competition resemble one another more closely than has previously been realized. The similarities of the two approaches are not readily apparent, partly because while it is natural to think about search as generating externalities, economists are less accustomed to analysing market power in such terms.

Since the primary concern of this chapter is the characterization of externalities caused by search and market power, the analysis paid little attention to the role of strategic complementarities. Specifically, the chapter considered only the impact effect of agents' strategies and did not explore the equilibrium responses of other agents in the economy. Such complementarities are relevant for comparative static analysis in these games, and perhaps warrant further study.

Notes

I would like to thank Russell Cooper, Carl Davidson, John Geanakoplos, Kenneth Koford, John Laitner, Rowena Pecchenino, Matthew Shapiro, Richard Startz and

Randall Wright for comments and discussion. I am particularly indebted to Joaquim Silvestre for very helpful discussion at the University of Warwick Macroeconomics Workshop (July 1993). The usual disclaimer applies with more than usual force.

1. Models with imperfect competition include Heller (1986), Blanchard and Kiyotaki (1987), Dixon (1987), Startz (1989), Kiyotaki (1988), Bénassy (1991a, 1991b) and many others. The literature is well surveyed by Silvestre (1993). Models of search externalities in general equilibrium include Howitt and McAfee (1988), Davidson *et al.* (1987, 1988, 1994), and Hosios (1990).
2. See Cooper and John (1988).
3. See, for example, Bils (1987), Shapiro (1987), Domowitz *et al.* (1988), Hall (1988), and Bresnahan (1989). Changes in market power over the business cycle are investigated by Rotemberg and Woodford (1991).
4. Koford *et al.* (1988) discuss policies to internalize such externalities.
5. There are as yet very few models that incorporate both search and imperfect competition. Drazen (1986) is an exception.
6. Although the model is set up as an exchange economy, it can easily be reinterpreted as a simple production economy. Agents in a sector could be endowed with labour and a technology for producing the good of their own sector using labour. Their preferences over the good produced in their own sector could then be interpreted as preferences over leisure.
7. A simple example is given by a variant of the Dixit and Stiglitz (1977) preferences:

$$U(.) = (c^0)^\alpha (C)^{1-\alpha} - v(c^n);\ C = (\sum_{m \neq n} (c^m)^\rho)^{1/\rho}$$

   for an agent in sector n. In this case, preferences are separable in the good supplied, and Cobb–Douglas over the numéraire good and a CES aggregator of all other goods. Blanchard and Kiyotaki (1987) use essentially this utility function.
8. Arrow's argument also led economists to *define* externalities in terms of missing markets rather than (as had previously been usual) in terms of interdependencies. Indeed, Heller and Starrett (1976) argue that viewing externalities as interdependencies 'is not a very useful definition, at least until the institutional framework is given'. The emphasis on interdependency in this chapter (taking as given the institutional framework) is thus a return to an earlier tradition (for example Meade, 1952; Buchanan and Stubblebine, 1962; Davis and Whinston, 1962; Turvey, 1963); the literature as it stood at the beginning of the 1970s was well surveyed by Mishan, 1971. The relationship between missing markets and the externalities of this chapter is considered in John (1988).
9. That is, assume that the arguments of these functions can be partitioned into a number of subsets, the members of which enter symmetrically. In general, there are many distinct subsets in each function, since each agent controls N strategy variables. In particular, in considering $a_{in}^m(s_{in}, s_{jn}, s_{km}, s_{hr})$, one must distinguish all strategies according to whether they pertain to good n, good m, or some other good, and also separate out strategies pertaining to the good in an agent's own sector.

10. In the terminology of Hurwicz (1979), a feasible mechanism is a 'balanced outcome function'.
11. Because the analysis here assumes that allocation mechanisms are continuous and differentiable, it excludes certain interesting models (such as the Leontief technology model of Bryant, 1983). Extending the ideas of this chapter to such settings would be a useful topic for future research but is well beyond the scope of the current work.
12. They are of course not sufficient conditions: sufficiency requires checking second order conditions and ensuring that the allocation is a global maximum. The focus in this chapter is purely on these necessary conditions.
13. The result does not depend on the assumption that transfers are executed through the numéraire good.
14. One could equivalently derive this result by supposing that the planner maximizes the utility of a representative agent.
15. Note that this normalization cannot be used in general to characterize the entire set of constrained-efficient allocations for a given mechanism, because the normalization would have to be different at different allocations. The normalization is appropriate for this purpose if utility functions are quasi-linear and separable in the numéraire. Such an assumption is common in much game theoretic analysis: see, for example, the discussion of 'u-money' in Shubik (1984). Invoking this assumption in case (3) amounts to giving the planner access to non-distortionary transfers: given an efficient allocation, the planner can then attain any desired distribution by reallocating the numéraire and leaving the allocation of other goods unchanged.
16. Noting that an agent's action may bestow positive externalities on some and negative externalities on others, Silvestre (1991) observes that internalizing externalities need not lead to Pareto improvements. He therefore argues against using externalities to characterize inefficiencies. By contrast, this chapter responds to the same observation by looking for ways to aggregate the external effects.
17. A similar distinction is made by Meade (1973); the definitions here can be understood as a generalization of Meade's approach. See also the discussion of pecuniary externalities in the following subsection.
18. Farrell (1988, p. 179), writing in the 'Puzzles' section of the *Journal of Economic Perspectives*, creates a fictional character who 'noticed pecuniary externalities, and thought they were real externalities'. Shleifer and Vishny (1988, p. 1221), by contrast, write of a 'pecuniary externality [that] makes a dollar of a firm's profit raise aggregate income by more than a dollar since other firms' profits also rise, and in this way gives rise to a "multiplier." Since such multipliers are ignored by firms making investment decisions, privately optimal investment decisions under uncertainty will not in general be socially optimal.' For other discussions of pecuniary externalities, see Baumol and Oates (1975), Van Huyck (1989), and Silvestre (1991).
19. Note also that if a feasible mechanism has this property it must be just feasible (i.e. $\sum_n \sum_i a_{in}^m = 0$, $m = 1, \ldots, N$). This is essentially Walras' Law.
20. For example, Lerner (1960, p.133) writes,

a small buyer may neglect the effect on *income* of his decisions to spend or not to spend. Although someone else's income must be reduced by a dollar when he spends a dollar less, he is not concerned with this. But as the buyer, or the group of buyers we are considering, becomes larger in relation to the economy, it becomes less and less appropriate to neglect the repercussion on income.

While Lerner does not use the term 'externality' explicitly, Colander (1986, p.354) quotes this passage and argues that 'Lerner used this "externality" argument – that individuals have no incentive to take into account the effect of their changes in spending – to justify monetary and fiscal policy'.

21. There are many other mechanisms that can be considered in this framework. In particular, it is shown in John (1988) that a fix-price mechanism (where quantities traded are determined by a rationing scheme) will give rise to real mechanism externalities unless all prices are at their Walrasian levels.
22. An alternative to this normalization would be symmetry assumptions that permit one to equate the marginal utility of the numéraire for all agents. All the results derived below hold in this case also, but the normalization has the advantage of simplifying the algebra.
23. The proof is as follows. Agent i in sector n can choose her strategies to yield the Walrasian allocation. Since the allocation mechanism forces agents to be on their budget constraints, no agent can do better than this allocation. Since the allocation obtained by this agent depends only on her strategy, the Walrasian outcome is, trivially, a Nash equilibrium.
24. See, for example, Dubey *et al.* (1980), Dubey (1980), Kahn (1979), Shubik (1984).
25. See Dubey *et al.* (1980) and Kahn (1979).
26. Greenwald and Stiglitz (1986) use similar intuition to explain why pecuniary externalities do not vanish when the number of agents is large unless real externalities are also absent.
27. There are other assumptions that would give market power to agents as suppliers only. For example, it could be assumed that there is a large number of agents in each sector, but that only a few agents are endowed with the good, or that agents in each sector acquire market power by forming coalitions. Alternatively, it could be assumed that there is another source of demand for goods that is not explicitly modelled.
28. That is, under symmetry, $s_{in}^{m} \leq \bar{\omega}/(N-1)$ since the total endowment is fixed at $I\bar{\omega}$ and there are $I(N-1)$ consumers of the good.
29. Under symmetry, any income effects will be spread over the other $(N-1)$ sectors, and so will disappear when N gets large, and the effect of a change in p^n on the demand for other goods will also be spread over the other $(N-1)$ sectors and so will also disappear when N gets large. Given the continuity assumptions on preferences it is then a reasonable conjecture that the claim in the chapter is true in most (all?) cases, although a complete proof is not offered here. It is difficult to derive more complete results without making more explicit assumptions on utility, and the argument is also more complicated in the absence of symmetry. As will be clear from Proposition 2 below, the result is not actually needed for any subsequent analysis.

30. A model with these features is considered in Cooper and John (1988). It should be emphasized that the previous discussion refers to the effect of a change in agent i's strategy, *holding other strategies constant*. A change in s_{in}^n will also affect other agents' incentives to produce through its effect on the price of good n. For example, if an increase in s_{in}^n induces agents in other sectors to increase their supply, then there is strategic complementarity. See Cooper and John (1988) for more discussion and analysis of this case.
31. The changes in prices will also induce changes in second-stage demands. By the envelope theorem, these can be ignored in the welfare calculations.
32. Again, it should be emphasized that the experiment here holds the strategies of other agents constant, and so does not explore the implications of strategic complementarities.
33. The intuition behind this result is thus related to the methodology of Greenwald and Stiglitz (1986). That study also examines cases where changes in prices have more than distributional consequences (for example, because prices also convey information). In order to isolate the efficiency consequences of these price changes, they net out the distributional effects.
34. The dependence of the demand for good n on the price of good m reflects both substitution possibilities across goods and income effects. That is, the demand for good n depends in part on the income earned in other sectors ($p^m q^m$).
35. Note that these external effects need not disappear when there is a large number of sectors. Whether or not they do in fact vanish depends upon how an increase in the number of sectors (goods) affects the elasticity of substitution between goods.
36. In Blanchard and Kiyotaki (1987), all other agents benefit if one agent lowers her price, so all agents would be viewed as being on the other side of the market. Analysis of Blanchard and Kiyotaki's model in terms of the framework of this chapter is complicated by the fact that they include money in the utility function, generating an extra channel whereby changes in prices affect the welfare of other agents. Detailed discussion is beyond the scope of this chapter.
37. Note that such a model would not satisfy the strong symmetry assumptions made on the utility function on p.140.
38. Baye and Cosimano (1990) investigate a search model where agents can choose their sector.
39. It does not matter for the analysis how search costs are incurred. If they are incurred in terms of the good that an agent supplies, one could interpret the good as labour effort (time) that is used first to search for trading partners and then to produce the good demanded by agents in the other sector. Alternatively, search effort could simply be another argument of the utility function, or search costs could be incurred in terms of the numéraire.
40. The analysis is easily extended to the case where agents do not divide the gains from trade evenly as in, for example, Hosios (1990).
41. In this two-stage game, it is reasonable to assume that the gains from trade that can be realized by matched agents are independent of the search activities of other agents. In a richer dynamic setting, this might not be true.
42. More precisely, let $\pi_{i1} = \min(1, (s_{i1}/S_1)f(S_1, S_2))$. An interior solution is assumed in what follows. The matching function has the property that a small agent

affects her own probability of matching but has a negligible effect on the total number of matches. It is of interest in part because it is similar to functions utilized in the literature on search and matching; for example, it is essentially equivalent to the matching function utilized by Davidson *et al.* (1987, 1988).

43. Proposition 5 is closely related to Hosios' (1990) result on entry–exit externalities. In essence, Hosios takes the matching function as given and identifies conditions on the sharing rule that imply efficiency, while Proposition 5 takes the sharing rule as given and looks for conditions on the matching function. See also Mortensen (1982, 1986) and Pissarides (1984).
44. Pissarides (1986) and Blanchard and Diamond (1989) have investigated the returns to scale properties of labour market matching functions in Britain and the USA, respectively. Pissarides finds evidence of constant returns, while Blanchard and Diamond find evidence of 'constant or mildly increasing' returns.

Part III

Labour market imperfections

7 Demand uncertainty and unemployment in a monopoly union model

Omar Licandro

Introduction

In unionized economies, nominal wages are normally set for a relatively long period of time, say one year, while employment fluctuates during the year depending on firms' particular situations. The standard 'right to manage' model, as in McDonald and Solow (1981), even if it assumes that wages are set before the firm decides employment, does not reflect completely this important sequence in the decision process, assuming that both decisions are taken under the same information concerning the environment. In this chapter, new information is revealed in between both decisions, allowing the firm to decide employment with richer information than the union has when deciding wages.[1] This sequence in the wage bargaining process leans on the assumption of nominal wage rigidities.

Many different types of uncertainty are relevant to the analysis of wage bargaining. Information concerning the aggregate price index, as in Lucas (1972) must generate some type of Lucas supply curve. Technological uncertainty or aggregate demand uncertainty could also be important to explain the behaviour of employment and wages over the business cycle, as is reported by Hansen and Wright (1992). However, we concentrate our attention on the effects of demand uncertainty coming from misinformation about individual preferences. As is frequently reported in the literature on marketing, firms are mainly concerned with forecasting their market shares.[2] However, macroeconomists seem to be more interested in the effects of technological shocks and aggregate demand shocks than in the effects that idiosyncratic demand shocks have on the aggregate equilibrium. Tobin (1972) argued, in his AEA presidential address, that random dispersion of demand in heterogeneous markets with wage rigidities is one of the main determinants of the 'natural rate of unemployment'.

The main structure of the model is taken from Licandro (1992) and Arnsperger and de la Croix (1993).[3] The economy is organized as an island economy. In each island, there is only one firm – which produces a differentiated good – a given number of households and a union, which represents households. The information structure of the model is crucial. It is a one period model, where decisions are made at two different moments in time. *Ex ante*, when preferences are not yet revealed, households decide to live and to work in a particular island and unions set the nominal wage. Unions are organized at the firm level, with decentralized negotiation. *Ex post*, when all relevant information is public, monopolistically competitive firms decide prices, employment and production and households decide consumption. The goods market is organized as in Dixit and Stiglitz (1977). Individual preferences are not symmetric, allowing for demand heterogeneity, i.e. some firms will have a high demand and some other firms a small demand. All the uncertainty is idiosyncratic (i.e. there is no aggregate uncertainty) and it is directly related to demand heterogeneity.

To stress the importance of information problems, we analyse a simple monopoly union model, in which there is full employment at the equilibrium with perfect information. However, when there are information problems, which take the form of firm-specific demand uncertainty, the nominal wage set by the union does not grant full employment at equilibrium. The existence of unemployment does not rely on the existence of union power, as in the standard 'right to manage' model. In the spirit of Tobin (1972), this chapter provides an explanation for the 'natural rate of unemployment', which is related to the existence of firm-specific demand uncertainty and wage rigidities.

The chapter is organized in the following way. The second section describes the general characteristics of the economy. In the third section the representative household problem is solved and the demand for goods is computed. In the fourth section we solve for the firm problem and the monopoly union problem. The fifth section gives the aggregate equilibrium, and conclusions are presented in the sixth section.

The economy

There are three types of economic agents: households, unions and firms. Each household supplies a given quantity of labour to a particular firm and demands goods. Households are represented by unions, which are organized at the firm level and set wages. Firms hire labour from households, produce differentiated goods and set prices.

A particular information structure is assumed: there are two times in the model, *ex ante* (before the revelation of individual preferences over goods) and *ex post* (when all relevant information is public). Households supply labour and unions decide wages without knowing with certainty the demand for the good produced by the firm. When wages are already set, households reveal their preferences and demand goods and firms set prices, hire workers and produce.

As in McDonald and Solow, there are two stages in the game. In the first stage the union sets the nominal wage and in the second stage the firm sets prices and hires workers in order to satisfy its demand. The main difference with the standard monopoly union model is that the firm information concerning the environment, when deciding employment, is richer than the union information when deciding wages. When the union sets the nominal wage the demand for the firm is not revealed yet, while the firm knows its own demand before deciding how many workers to hire. In this sense, the model imposes some type of wage rigidity.

The demand side

Households behave as in Dixit and Stiglitz (1977). Let us assume that all of them have the same utility function, hold the same initial money balances and supply the same given quantity of labour.

The representative household

There is a continuum of households in the interval $[0, n]$, each of them offering one unit of labour. There is also a continuum of goods in the interval $[0,1]$. Households are indexed by j and goods by i. Households are identical except for the fact that their labour incomes are not necessarily the same. The representative consumer optimization problem is

$$\max_{\{c(i)\}, M/P} C^{\gamma}\left(\frac{M}{P}\right)^{1-\gamma}$$

where

$$C=\left(\int_0^1 v(i)^{\frac{1}{\theta}}\, c(i)^{\frac{\theta-1}{\theta}}\, di\right)^{\frac{\theta}{\theta-1}}$$

$$\theta>1 \text{ and } \int_0^1 v(i)\, di=1;$$

subject to

$$\int_0^1 p(i)\, c(i)\, di=I;$$

I, p and $p(i) \forall i \in [0,1]$ are given. C is an index of consumption utility, M represents money holdings and p the aggregate price index. $c(i)$ and $p(i)$ are the consumption and the price of the good i, respectively. The parameters γ, θ and $v(i)$, $\forall i \in [0,1]$, are supposed given. I represents total nominal revenues of the representative consumer and it can be different from one household to another.

Optimal consumption and money holdings are[4]

$$C = \gamma \frac{I}{p}$$

and

$$M = (1-\gamma)I.$$

Notice that the 'indirect utility function', which can be derived by substituting both optimality conditions in the utility function, is proportional to real revenues.

The optimality condition for $c(i)$ is

$$c(i) = \left(\frac{p(i)}{p}\right)^{-\theta} C\, v(i), \tag{1}$$

where

$$p = \left(\int_0^1 v(i)\, p(i)^{1-\theta}\, di \right)^{\frac{1}{1-\theta}}$$

is the 'true price index' associated with the representative household utility function.[5]

Let us call $I(j)$ the total revenues of the household j:

$$I(j) = \frac{\bar{M}}{n} + w(j)\, l(j) + \int_0^1 \frac{1}{n}\, \pi(i)\, di,$$

where $\bar{M}$ represents aggregate initial money holdings, $w(j)$ is the nominal wage rate and $l(j) \in \{0,1\}$ represents employment. Profits, denoted by $\pi(i)$, are distributed among households. The share of the firm i is supposed to be the same for all households and equal to $1/n$. The only difference among households comes from the equilibrium value of labour incomes $w(j)l(j)$.

Let us define profits as

$$\pi(i) = p(i)y(i) - w(i)l(i).$$

Aggregating revenues over consumers, we have that

$$\int_0^n I(j)\,dj=\bar{M}+p\bar{y}.$$

Variable $\bar{y}$ represents aggregate production, and it takes the same functional form as the quantity index C.

Imposing the condition $\int_0^n C(j)dj=\bar{y}$ on the goods market, which must hold in equilibrium, from previous conditions we get

$$\bar{y}=\left(\frac{\gamma}{1-\gamma}\right)\frac{\bar{M}}{p}.$$

Let $c(i,j)$ represent the demand of good i from household j. Integrating (1) over households we have that total demand $d(i)$ for good i is

$$d(i)=\left(\frac{p(i)}{p}\right)^{-\theta}\bar{y}\,v(i). \tag{2}$$

Aggregate demand, $\bar{y}$, is distributed among the differentiated goods depending on relative prices and the $v(i)$ parameters.[6]

Demand heterogeneity is directly related to the distribution of the v parameter among firms. It plays a very important role in the model, because all uncertainty comes from the absence of perfect information concerning this parameter.

The supply side

In the supply side of this economy there is a continuum of monopolistically competitive firms, each of them producing a variety of the unique good. The index i also identifies firms. Each worker supplies one unit of labour to a specific firm.[7] Workers are uniformly distributed among firms and, from previous assumptions, the number of workers offering their labour to a particular firm is n. There is a continuum of unions, each of them representing the workers offering their labour to a particular firm. Unions are also indexed by i. At the firm level, unions set wages as a monopoly.

An important assumption is imposed to produce full employment capacities at the firm level: labour markets are segmented. Each worker is offering his labour to a specific firm and, if a firm decides not to hire a worker, this worker is unable to offer his labour to another firm. Labour market segmentation can be justified by differences in human capital, labour mobility costs, turnover costs, etc. Under this assumption each firm faces an upper bound on production, i.e. the 'full employment output'. This assumption is crucial to have unemployment at equilibrium.

Firm behaviour

Let us assume that the labour marginal productivity is constant and equal to one, i.e. there is a constant returns to scale technology

$$y(i)=l(i)$$

where $y(i)$ and $l(i)$ represent firm i's production and employment. Notice that the firm's employment, $l(i)\in[0,n]$, is different to the household's employment $l(j)\in\{0,1\}$. Full employment output is equal to n for all i.

Under the previous conditions, the representative firm must solve the following optimization problem:

$$\max_{p(i),y(i)} \pi(i)=(p(i)-w(i))\,y(i) \tag{3}$$

subject to

$$y(i)=\left(\frac{p(i)}{p}\right)^{-\theta}\bar{y}\,v(i), \qquad y(i)\leq n,$$

where $\bar{y}$, p, n and $w(i)$ are given. Parameters θ and $v(i)$ come from household preferences.

The first order condition for this problem is:
if $d(i)\leq n$,

$$p(i)=\left(1-\frac{1}{\theta}\right)^{-1} w(i); \tag{4}$$

if not,

$$p(i)=p\left(\frac{\bar{y}}{n}\right)^{\frac{1}{\theta}} v(i)^{\frac{1}{\theta}}. \tag{5}$$

Depending on the relation between demand $d(i)$ and full employment capacity n, the representative firm sets prices following two different rules. When demand is relatively small, the firm sets a price that satisfies the standard condition for a monopoly, i.e. marginal costs equal to marginal revenues (interior solution). When the optimal condition for a monopolistic competitive firm holds for a demand greater than full employment capacities, the firm sets a higher price in order to satisfy demand at the full employment level (corner solution).

Let us call

$$\bar{v}(i)=\left(1-\frac{1}{\theta}\right)^{-\theta}\left(\frac{w(i)}{p}\right)^{\theta}\frac{n}{\bar{y}} \tag{6}$$

the value of $v(i)$ at which the interior solution holds at the corner, i.e. both conditions (4) and (5) are satisfied simultaneously.

From the restrictions in problem (3), and the corresponding optimal condition (4) and (5), we can deduce the optimal employment of firm i:
if $v(i) \leq \bar{v}(i)$,

$$l(i) = l_d(i) = \frac{v(i)n}{\bar{v}(i)} \leq n, \tag{7}$$

if not,

$$l(i) = n. \tag{8}$$

In the interior solution, the firm is choosing employment subject to its unconstrained labour demand curve, denoted $l_d(i)$ in (7). In this case, the firm does not employ all the workers living in the island and there will be unemployment in this segment of the market. In the corner solution, the firm is constrained by the labour supply and producing at its full employment output.

Monopoly union behaviour

In each island, a trade union represents the workers offering their labour to the firm producing the corresponding variety. The union is assumed to behave as a monopoly in the labour market.

The objective function of the ith union is:

$$V(w(i), l(i)) = \left(\frac{w(i)}{p}\right) l(i),$$

where V is the sum of the indirect utility functions of the risk-neutral members after the deduction of the fallback level,

$$\frac{\bar{M}}{p} + \int_0^1 \frac{\pi(i)}{p}\, di,$$

i.e. the non-human revenues.

The standard monopoly union model

To give a better understanding of the results provided in this chapter, let us first solve the standard monopoly union model, in which wages and employment are decided under the same information set. Let us assume in this section (this assumption will be dropped in the next) that the union has full information, in particular that it knows the value of $v(i)$ faced

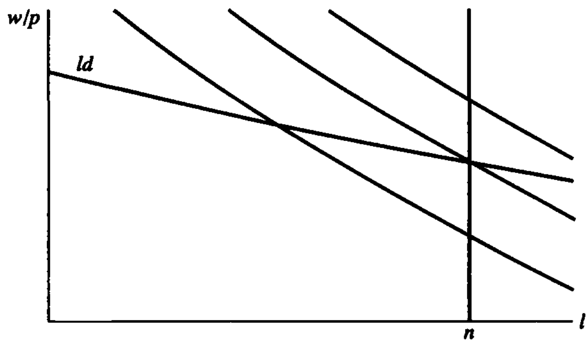


Figure 7.1 The standard monopoly union model

by the firm i. It can be easily shown that under this assumption the monopoly union is optimally setting nominal wages in order to have full employment at equilibrium. The main reason for that is that the inverse of the labour demand elasticity ($\frac{1}{\theta}$) is smaller than the elasticity of the indifference curves (which is equal to one), anywhere. In this case the union is interested in reducing wages until full employment is reached. As figure 7.1 shows, the optimal choice for the union is to set a nominal wage that induces the firm to optimally choose to produce at full employment. In other words, the union is choosing $w(i)$ in such a way that both (4) and (5) hold simultaneously.

Under these particular assumptions, if wages and employment are decided under the same information set, there is full employment at equilibrium in the standard monopoly union model.

Monopoly union behaviour under demand uncertainty

Let us assume that, when deciding wages, the ith union knows the 'demand function' assigned to the variety i, (2) and the distribution of the $v(i)$ parameters, denoted by $F(v)$. However, we assume that the representative union does not know with certainty the specific $v(i)$ faced by the ith firm. Since there is no aggregate uncertainty, the union can solve for the aggregate demand $\bar{y}$ and the aggregate price index p.

Notice that under these conditions all unions are *ex ante* identical, even if *ex post* the labour demand can be different from one island to another. Then, they set the same wage rate and they face the same $\bar{v}$. For this reason, we can drop the i index in what follows.

Since the objective of the union is linear in l, under demand uncertainty, the union is mainly concerned with the forecast of expected employment. From (7) and (8), expected employment can be written as[8]

$$E(l)=\frac{n}{\bar{v}}\int_{v\leq\bar{v}} v\ \mathrm{dF}(v)+n\int_{v\geq\bar{v}}\mathrm{dF}(v)\leq n, \tag{9}$$

where $\bar{v}$ is given by (6).

As stated before, the distribution function $\mathrm{F}(v)$ represents the distribution of parameter v among the different firms and it depends on household's preferences. The union knows that its specific v is drawn from this distribution. If there is a strictly positive probability of being in an unemployment equilibrium, expected employment will be strictly smaller than full employment.

Let us define the weighted probability of being in a full employment equilibrium as

$$\mathrm{P}_w(l=n)=\int_{v\geq\bar{v}}\frac{n}{\mathrm{E}(l)}\mathrm{dF}(v),$$

and the weighted probability of being in an unemployment equilibrium as

$$\mathrm{P}_w(l<n)=\frac{n}{\bar{v}}\int_{v\leq\bar{v}}\frac{n}{\mathrm{E}(l)}\mathrm{dF}(v).$$

The representative union problem is

$$\max_{w}\ \mathrm{E}(V)=\left(\frac{w}{p}\right)\mathrm{E}(l),$$

where $\mathrm{E}(l)$ is given by (9) and $\bar{v}$ is given by (6). Because there is not aggregate uncertainty, the aggregate variables p and $\bar{y}$ are perfectly forecast by the union.

The first order condition for this problem is

$$\frac{\mathrm{E}(l)}{n}\bar{v}=\theta\int_{v\leq\bar{v}} v\ \mathrm{dF}(v). \tag{10}$$

Condition (10) can be interpreted in the following way: the union is optimally choosing the weighted probability of being in an unemployment equilibrium, whose optimal value is equal to the inverse of the demand elasticity of labour ($\frac{1}{\theta}$). Notice that the weighted probability of being in a full employment equilibrium is equal to $(1-\frac{1}{\theta})$.

Figure 7.2 gives a graphic representation of this problem. The union maximizes its utility function over the expected employment curve. Because there is a positive probability of *ex post* unemployment, the union expected

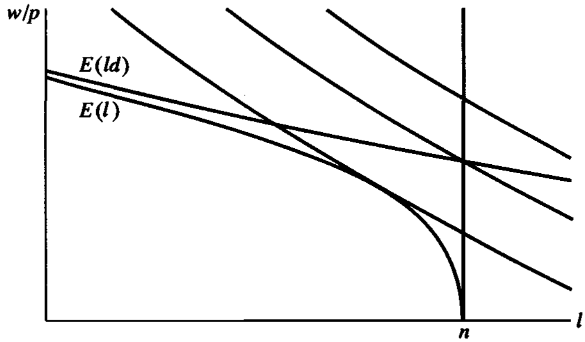


Figure 7.2 The monopoly union model under demand uncertainty

employment curve is always below full employment, in particular at the optimum. The expected employment locus is concave for any standard continuous distribution function, which is a sufficient condition for the existence of an interior solution.

Aggregate employment

Aggregate employment can be obtained by aggregation over firms' employment and it must be equal to union's expected employment given by (9).[9] Moreover, the optimality condition (10) must hold at equilibrium. Aggregate employment $\bar{l}$, aggregate production $\bar{y}$, aggregate real wages (which are equal over all islands) and $\bar{v}$ must satisfy at equilibrium the following conditions:

$$\left(\frac{\bar{l}}{n}\right)=\frac{1}{\bar{v}}\int_{v\leq\bar{v}} v\ \mathrm{dF}(v)+\int_{v\geq\bar{v}}\mathrm{dF}(v)\leq 1,$$

$$\left(\frac{\bar{l}}{n}\right)\bar{v}=\theta\int_{v\leq\bar{v}} v\ \mathrm{dF}(v),$$

$$\bar{v}=\left(1-\frac{1}{\theta}\right)^{-\theta}\left(\frac{w}{p}\right)^{\theta}\left(\frac{n}{\bar{y}}\right),$$

and

$$\left(\frac{\bar{y}}{n}\right)=\left[\bar{v}^{\frac{1-\theta}{\theta}}\int_{v\leq\bar{v}} v\ \mathrm{dF}(v)+\int_{v\geq\bar{v}} v^{\frac{1}{\theta}}\,\mathrm{dF}(v)\right]^{\frac{\theta}{\theta-1}}.$$

To better understand this result let us present an example, in which we assume a particular form for the distribution function $F(v)$.

Example

Let us assume that v follows a lognormal distribution, with unit mean and variance denoted by σ. In this case we can apply Lambert (1988) and approximate expected employment by the following function

$$E(l) = n(1 + \bar{v}^{\rho})^{-\frac{1}{\rho}}$$

where ρ is a decreasing function of σ. In particular, $\frac{\partial \rho}{\partial \sigma} < 0$, $\rho \to \infty$ when $\sigma \to 0$ and ρ is positive.

The representative union problem becomes

$$\max_{w} E(V) = \left(\frac{w}{p}\right) E(l)$$

where

$$E(l) = n(1 + \bar{v}^{\rho})^{-\frac{1}{\rho}}, \qquad \bar{v} = \left(1 - \frac{1}{\theta}\right)^{-\theta} \left(\frac{w}{p}\right)^{\theta} \frac{n}{\bar{y}}$$

and p, $\bar{y}$ and n are given.

Solving this problem as in the previous section and solving for the equilibrium value of aggregate employment $\bar{l}$, we have

$$\left(\frac{\bar{l}}{n}\right) = \left(1 - \frac{1}{\theta}\right)^{\frac{1}{\rho}} \leq 1. \tag{11}$$

Since workers are distributed homogeneously among firms and the marginal productivity is the same for all of them (it was normalized to one), full employment output n is equal across firms. Moreover, since unions are *ex ante* identical, they all set the same real wage. Under these conditions the marginal cost function is the same for all firms and it is constant and finite until full employment is reached, then it becomes infinitely elastic. Depending on their particular value for v, firms are setting prices and production either at the interior or at the corner solution. When demand is relatively small ($v < \bar{v}$) in an island, production is smaller than full employment output. When demand is relatively large ($v \geq \bar{v}$), the firm produces at full employment. In the aggregate there is unemployment.

The unemployment rate takes the following equilibrium value, denoted by u,

$$u = 1 - \left(1 - \frac{1}{\theta}\right)^{\frac{1}{\rho}}.$$

The unemployment rate at equilibrium depends on the elasticity of substitution θ and on the parameter ρ, which depends on the variance of the distribution of the ν parameter. If $\theta \rightarrow \infty$ all goods become perfect substitutes, and if $\rho \rightarrow \infty$ the parameter ν becomes the same for all goods. In both cases the heterogeneity of demand disappears and the unemployment rate goes to zero. The first derivative of u with respect to both parameters is negative, i.e. an increase in demand heterogeneity, coming from a greater elasticity of substitution or a greater dispersion on the ν parameters, always generates an increase in the unemployment rate.

Conclusions

The main concern of this chapter is to show the importance of demand uncertainty in the determination of the 'natural rate of unemployment'. Demand uncertainty is introduced in a monopoly union model where unions set wages at the first stage of the game, without knowing with certainty the demand for the good produced by the firm. Because the union assigns a positive probability to the event 'underemployment equilibrium', it sets an optimal nominal wage at which the expected employment is smaller than full employment. In an economy where all the uncertainty is firm-specific (i.e. there is not aggregate uncertainty), aggregate employment is equal to the union expected employment and then there is unemployment at equilibrium. In some islands the idiosyncratic demand shock is high and firms produce constrained by their full employment capacity, but at the same time in the other islands the idiosyncratic demand shock is low and firms optimally produce less than their full employment output.

The existence of unemployment depends crucially on the assumption of demand heterogeneity and demand uncertainty. The assumptions of nominal wage rigidity and labour market segmentation are not sufficient to generate this result. Moreover, the assumption of only one firm per island (monopolistic competition) is not critical for the existence of unemployment at equilibrium, and the result holds even if there is perfect competition on the goods market of each island. In this sense, the 'natural rate of unemployment', displayed by the model at equilibrium, relies more on the existence of 'information problems' than on the existence of 'coordination failures'.

Notes

I thank John Driffill, Neil Rankin, Rodolphe Dos Santos Ferreira, Joaquim Silvestre and the participants of the University of Warwick Macroeconomics Workshop (July 1993) for helpful comments. Financial support of the European

Economic Community (SPES-CT91-0079) and the Ministerio de Educación y Ciencia (CE92-0017) are gratefully acknowledged.

1. Manning (1987) and Espinosa and Rhee (1989) develop more general frameworks to analyse the question: are wage bargaining contracts efficient or does the union let the firm manage employment? In both studies it is assumed that, even if both decisions are taken sequentially at two different stages of the game, the information concerning the environment is the same in both stages. The existence of asymmetric information, or costly information search, or a costly bargaining process, could be useful to attempt an explanation for this particular sequence of the wage bargaining process.
2. See Lambin (1993).
3. It is an attempt at reconciling the 'fix-price' or 'quantity rationing approach' with the 'New Keynesian economics', in particular with the monopolistically competitive general equilibrium approach. We show in this chapter that the main results in Licandro (1992) and Arnsperger and de la Croix (1993) do not depend on the existence of 'quantity rationing' in the goods market. The essential element of the model is related to the sequence of decisions and the structure of information.
4. Since the utility function is concave in its arguments and the budget constraint is linear, the first order conditions are necessary and sufficient for a maximum.
5. The normalization condition imposed over the v parameters in problem (1), implies that $p=\bar{p}$ if $p(i)=\bar{p}\forall i\in[0,1]$.
6. The firm i market share is

$$\frac{p(i)d(i)}{p\bar{y}}=\left(\frac{p(i)}{p}\right)^{1-\theta}v(i),$$

which depends on both the relative price and the $v(i)$ parameter. At the symmetric equilibrium the market share is equal to $v(i)$.
7. Because the marginal disutility of labour is zero, the representative household is optimally willing to work the maximum feasible time, which is assumed to be one.
8. The expected value of the minimum condition has been largely analysed in econometric disequilibrium models, in particular in the context of 'aggregation over micro-markets in disequilibrium', as is reported by Quandt (1988).
9. The proposed definition of aggregate employment is the standard addition of employed workers, which does not take into account that the marginal value of workers is not necessarily the same in all islands. For this reason the employment index and the production index are different, even if production is equal to employment for each firm. We keep the standard definition to be consistent with the literature on employment and unemployment.

8 Efficiency wages as a persistence mechanism

Gilles Saint-Paul

Introduction

This chapter explores a dynamic extension of the efficiency wage model of Shapiro and Stiglitz (1984). The main result is that the efficiency wage model can generate hysteresis. This is an important result because it is usually thought that hysteresis can chiefly be explained by the insider–outsider theory of unemployment, as in Blanchard and Summers (1986). In part for this reason, it is often considered that the efficiency wage model is a good model of the natural rate of unemployment but that it has little to say about its dynamics. This chapter builds a model which shows that this assessment is incorrect. The intuition behind the hysteresis result is that when a firm's employment is expected to shrink, workers have a greater incentive to shirk. In order to prevent them from doing so, it is necessary for the firm to raise future wages. This will generate an incentive to limit the number of people it will lay off. Hence the firm's employment will have a tendency to stay where it is, because deviations from the current value are costly in terms of incentives. More specifically, we build a model where effort inducement generates a cost function which is kinked at the current level of employment, implying corridor effects in the same fashion as is obtained in models with linear adjustment costs.[1]

Previous work on business cycles and efficiency wages (including Strand, 1992a and Danthine and Donaldson, 1990) has focused on the effect of efficiency wages on employment *variability*. The present chapter is the first, to my knowledge, to focus on employment *persistence*.

The chapter is divided into three sections: in the first, the basic assumptions of the model are set up and the labour cost function is derived and discussed. The second section shows how such a cost function generates hysteresis. The third section deals with some general equilibrium problems.

Presentation of the model

We consider an infinite horizon firm that is subject to shocks to its product demand and must pay workers efficiency wages in order to avoid shirking.

The model can also be thought of as a model of an open economy where workers consume an imported good and the representative firm produces an exported good whose demand fluctuates over time. Hence we will interchangeably talk about 'the firm's labour force' or 'employment'.

In each period, for a given level of employment l_t, the total revenue of the firm is:

$$R_t = \theta_t f(l_t).$$

This formula can be interpreted in a competitive fashion, in this case θ_t is just the price of the firm's good in period t and f is the production function. It can also be interpreted as the total revenue of an imperfectly competitive firm facing a downward-sloping demand curve. In any case, f is a concave, increasing function and θ_t is a random variable with a cumulative density function $G(\theta)$, $G' = g$. We assume that the θs are i.i.d. Hence, the firm faces only temporary shocks.

We assume that the value of the shock in period $t+1$ is known at time t by both the firm and its employees. Hence l_{t+1} is known one period ahead.[2]

The wage equation

In this subsection, we derive the equation describing wage behaviour under the hypothesis that workers are paid efficiency wages. For this we use a discrete time version of the Shapiro–Stiglitz model. Hence we assume that if people shirk, they can be caught with some probability; that people are never caught erroneously; and that if somebody is caught shirking he is fired. Because the penalty from being caught shirking is losing one's job, employed people must enjoy rents compared to the unemployed. Otherwise the penalty from shirking would be equal to zero and nobody would provide effort. This prevents the unemployed from credibly underbidding the employed, thereby generating equilibrium involuntary unemployment.

Although this assumption sounds quite natural and realistic, it is surprisingly controversial. In particular, it has been argued that workers could post a bond when hired, on which they would earn some annuity while employed, and which they would lose if caught shirking. Bonding would thus allow the firm to make the loss from shirking large enough to induce effort without requiring a wage higher than the reservation wage. Hence involuntary unemployment would be eliminated. This argument is not very convincing. First, as Katz (1986) argues, bonding is never observed in practice and is subject to severe legal limitations. Second, the firm would have an incentive to pretend that the worker has shirked in order to collect the bond. This problem could be solved with the intervention of a third party. However, it is clear that such a contract would be costly to enforce, subject to frequent litigation, with the risk of collusion between the firm and the third party. Hence it is likely that paying efficiency wages is a preferable alternative.[3]

Once bonding is ruled out, it is clear that firing in case of shirking is the best alternative, at least in our model. This is because it imposes the highest possible punishment on a worker, while it is not accompanied by any sort of costs for the firm (in a more complex model, for example with match-specific investment and the possibility of erroneous detection of shirking, other punishment schemes might be preferable for the firm).

Workers have an infinite horizon. In any period t, a worker's instantaneous utility is equal to the difference between the amount of money he gets and the effort he is supplying:

$$u_t = w_t - e_t.$$

If the worker is employed and he 'shirks', $e_t = 0$, if he does not shirk $e_t = 1$. Let U_t (resp. V_t) be the present discounted utility of an unemployed (resp. employed) person at time t. Let w_t be the wage paid in period t.

Suppose that if a worker does not shirk, he loses his job with probability p_t at the end of period t, whereas if he shirks, he loses his job with probability $q_t > p_t$.

If γ is the discount factor, then a worker who decides to shirk in period t gets an expected utility equal to:

$$V_{st} = w_t + \gamma[(1 - q_t)E_tV_{t+1} + q_tE_tU_{t+1}]$$

where E_t is the expectations operator conditional on information available in periods t, which includes θ_{t+1}.

If he decides not to shirk, his expected utility is equal to:

$$V_{nt} = w_t - 1 + \gamma[(1 - p_t)E_tV_{t+1} + p_tE_tU_{t+1}]. \tag{1}$$

In order to avoid shirking in period t the firm must set wages in such a way that:

$$V_{nt} \geq V_{st}. \tag{2}$$

We know that workers that are paid efficiency wages typically enjoy rents. This implies that there are queues for those jobs (in the Shapiro–Stiglitz model, there is involuntary unemployment). Therefore, the firm can lower wages to the point where (2) is satisfied with equality while being still able to attract workers. The condition $V_{nt} = V_{st} = V_t$ can be written:

$$E_tV_{t+1} = E_tU_{t+1} + \frac{1}{\gamma(q_t - p_t)}. \tag{3}$$

Notice that the current wage plays no role in this equation. What is relevant for incentives is the cost of getting fired, which consists of forgone future wages embodied in V_{t+1}. Therefore, only future wages determine the worker's current effort.

Suppose now that there is a monitoring process such that people who shirk are caught with probability x while they can lose their jobs for other reasons with probability p_t, the two events being independent. Then:

$$q_t = p_t + x - xp_t. \tag{4}$$

Using (4), (3) can be rewritten:

$$E_t V_{t+1} = E_t U_{t+1} + \frac{1}{\gamma x(1-p_t)}. \tag{5}$$

If (5) is satisfied, then workers will not shirk and $V_t = V_{nt}$ in every period. Using this condition and (1) and (5) one gets:

$$E_{t-1} w_t = (E_{t-1} U_t - \gamma E_{t-1} U_{t+1}) + \left(1 - \frac{1}{x}\right) + \frac{1}{\gamma x(1-p_{t-1})}. \tag{6}$$

This is the equivalent of the 'No shirking condition' in discrete time. In (6), $E_{t-1}(U_t - \gamma U_{t+1}) + 1$ is the alternative wage and the remaining term is the rent. Given that l_t is known at time $t-1$, (6) implies that the expected cost of labour is predetermined. Hence there is no reason for the firm to put noise in the determination of w_t. We will thus assume that w_t is non-random. Furthermore, we will restrict ourselves to the case where $E_{t-1}U_t - \gamma E_{t-1}U_{t+1}$ is roughly constant. What this essentially means is that we neglect the effect of employment changes on an unemployed person's expected utility. This assumption can be taken as an approximation, i.e. an assumption that this quantity does not fluctuate too much. This will be true if the expected utility from getting a job is a small enough contribution to an unemployed person's utility. This assumption can be exactly true if we assume that unemployment compensation is adjusted each period so as to maintain U_t constant. This is not completely unrealistic since unemployment benefits tend to be adjusted upwards when joblessness becomes more of a problem (we relax this assumption in the third section below). (6) can therefore be written:

$$w_t = a + \frac{b}{1-p_{t-1}} \tag{7}$$

where $a = E_{t-1}U_t - \gamma E_{t-1}U_{t+1} + 1 - 1/x$; and $b = 1/\gamma x$.

(7) implies that the higher the probability that a worker be fired at the end of period $t-1$, the higher his future wage must be in order to induce him to work during period $t-1$.[4]

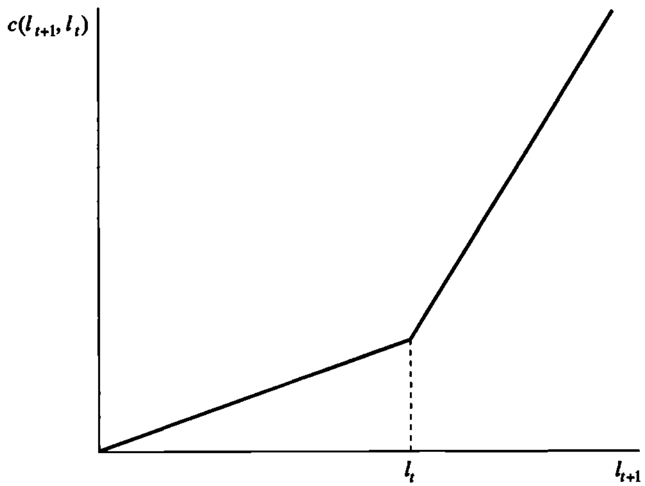


Figure 8.1 The total cost of labour in period $t+1$

The cost of labour

We now study the total cost of labour for a firm whose workers must be paid according to (7). We assume that it will choose a uniform rule, with everybody having the same probability of being fired.[5] Thus, each worker has a probability:

$$p_t = \max\{(l_t - l_{t+1})/l_t, 0\}$$

of being fired at the end of period t. Therefore, the wage that the firm must offer in period $t+1$ is, according to (7):

$$w_{t+1} = a + b\max\{l_t/l_{t+1}, 1\}.$$

The cost of employing l_{t+1} workers at date $t+1$ is then:

$$c(l_{t+1}, l_t) = al_{t+1} + b\max\{l_t, l_{t+1}\}. \quad (8)$$

Properties of cost function (8)

Figure 8.1 depicts the cost of employing l_{t+1} workers in period $t+1$. One can see that this function exhibits a kink at the current labour force. The marginal cost of labour is lower for $l_{t+1} < l_t$ than for $l_{t+1} \geq l_t$. Hence, the model is formally similar to a model of labour demand with linear adjustment costs (see Oi, 1962; Nickell, 1986).

The fact that the kink is at the *current* level of employment implies that this level will influence the future levels of employment, even if no adjustment costs are embodied in the firm's optimization programme. Furthermore, because of this discontinuity, small shocks will have no effect at all on the level of employment.

The next section studies the dynamic optimization programme of a firm that is subject to shocks and whose labour cost function is described by (8).

The dynamics of labour demand

In this section, we study the behaviour of a firm whose labour cost function is described by (8) above.

Let $V(l_t, \theta_{t+1})$ be the expected discounted value of the firm's profit as of period t, with δ the discount rate.

$$\begin{aligned} V(l_t,\theta_{t+1}) &= \max_{l_{t+1}} \theta_{t+1} f(l_{t+1}) - c(l_{t+1}, l_t) + \delta E_t V(l_{t+1}, \theta_{t+2}) \\ &= \max_{l_{t+1}} \theta_{t+1} f(l_{t+1}) - a l_{t+1} - b\max\{l_t, l_{t+1}\} \\ &\quad + \delta E_t V(l_{t+1}, \theta_{t+2}) \end{aligned} \tag{9}$$

where E_t denotes expectations at the end of period t. Notice that we omitted profits and wages in period t since they are determined in period $t-1$.

Since θ_{t+2} is independent of θ_{t+1} we can write:

$$E_t V(l_{t+1}, \theta_{t+2}) = H(l_{t+1}).$$

At the beginning of time t, the firm chooses l_{t+1} in order to solve (9). This problem is formally equivalent to a stochastic labour demand problem with linear adjustment costs as treated in Bentolila (1988) and Bentolila and Saint-Paul (1994). We briefly derive the solution. One can distinguish three cases:

1. The optimal solution satisfies $l_{t+1} < l_t$. In this case, the first order condition can be written:

$$\theta_{t+1} f'(l_{t+1}) + \delta h(l_{t+1}) = a \tag{10}$$

where $h = H'$. This happens whenever θ_{t+1} is lower than

$$\theta_m(l_t) = \frac{a - \delta h(l_t)}{f'(l_t)}.$$

2. The optimal solution satisfies $l_{t+1} > l_t$. The first order condition is then:

$$\theta_{t+1} f'(l_{t+1}) + \delta h(l_{t+1}) = a + b. \tag{11}$$

This happens if θ_{t+1} is higher than

$$\theta_M(l_t) = \frac{a + b - \delta h(l_t)}{f'(l_t)}.$$

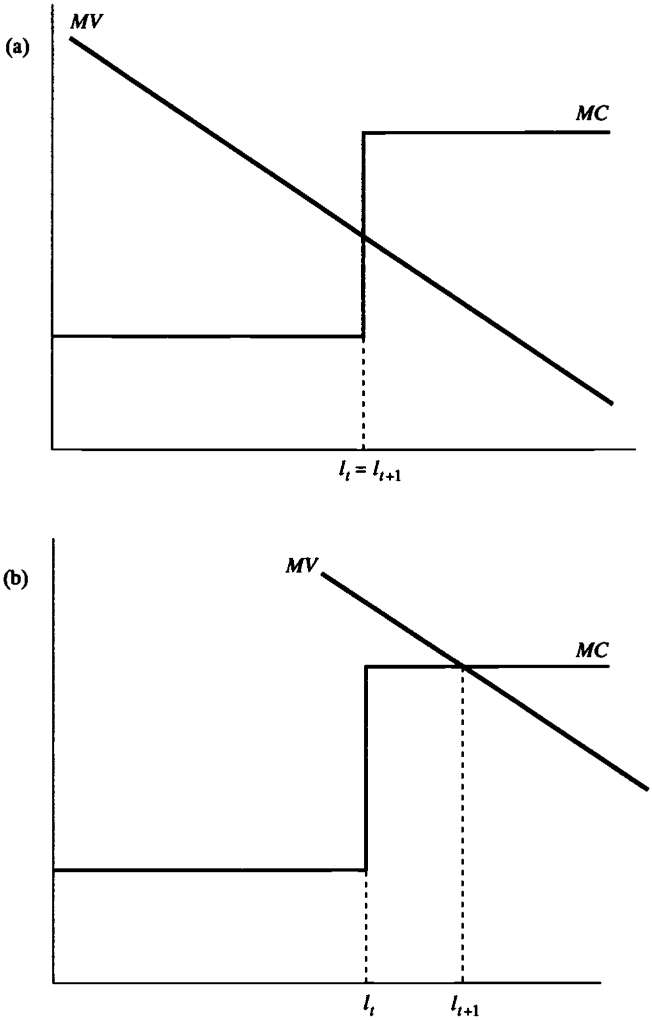
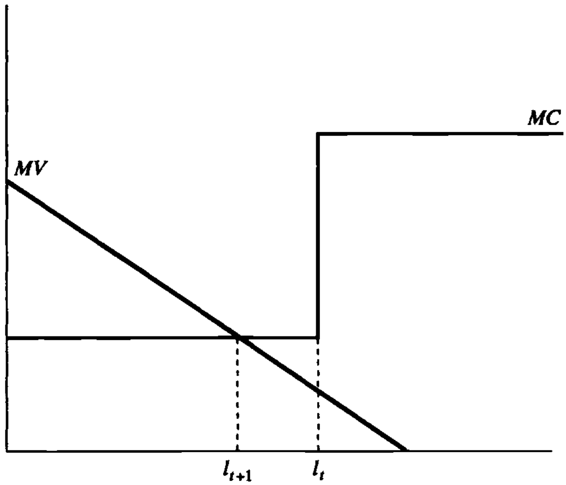


Figure 8.2 Determination of employment in period $t+1$
(a) No change
(b) Expansion
(c) Contraction

(c)



3. If $\theta_m \leq \theta_{t+1} \leq \theta_M$, then $l_{t+1} = l_t$.

(10) and (11) can be easily interpreted. The left-hand side is the marginal value of labour. The right-hand side is its marginal cost, which is $a+b$ when the firm is expanding but only a when the firm is contracting. When the marginal value of labour at $l_{t+1} = l_t$ is between a and $a+b$, it is profitable not to alter the labour force (case 3).

Furthermore, it is shown in appendix 1 that h is negative and decreasing and satisfies the following equation:

$$h(l) = -f'(l) \int_{\theta_m(l)}^{\theta_M(l)} G(\theta)d\theta. \tag{12}$$

This implies that the marginal value schedule $MV = \theta f'(l) + \delta h(l)$ is downward-sloping. Because h is negative, the marginal value of labour is lower than its marginal product in the current period. This is due to the fact that each additional worker increases the potential firing costs that the firm would incur if it were to lay off workers in the future. This is embodied in the gap between θ_m and θ_M that is larger, the larger is b. Hence, if the firm did not have to pay efficiency wages, it would hire more people at any given wage.

Figure 8.2 shows that l_{t+1} is determined by the intersection of the MV schedule and the marginal cost schedule MC which is derived from (8). Depending on where the MV schedule is located, one can distinguish three cases 8.2a, 8.2b, and 8.2c, that are the geometric counterpart to the three cases studied above.

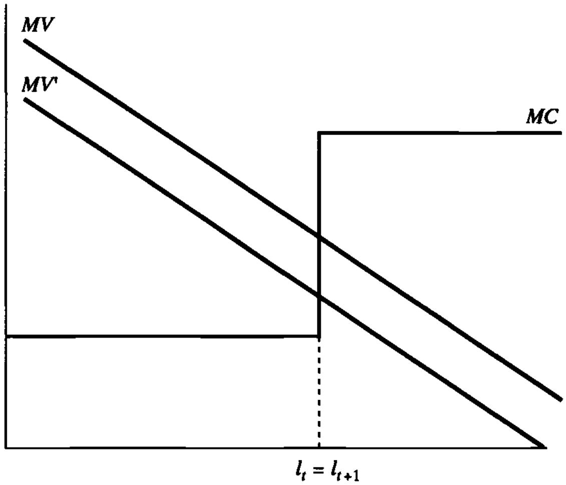


Figure 8.3 Effects of a small shock

Now suppose that figure 8.2a shows how the economy is 'normally' operating, say if θ is equal to its expected value $E\theta$. Suppose that there is a small shock (i.e. $\theta_{t+1} \neq E\theta$) that shifts the MV schedule. Then it is clear from figure 8.3 that if the shock is small enough, the size of the labour force is unchanged. Hence a first implication of the model is that it exhibits corridor effects:

Small shocks have no effect on the level of employment

Consider now the impact of a large shock, for instance a contraction in period $t+1$. Then, as is shown in figure 8.4, the MV schedule shifts downwards, and this shift is large enough to imply a lower level of employment in period $t+1$. but the relevant marginal cost schedule for period $t+2$ will also shift since it has now a step at the new level of employment, $l_{t+1} < l_t$ (see figure 8.4). Let MC′ be the new schedule.

Now assume that in period $t+2$, θ is back to its 'normal' value. Then the new level of employment is given by the intersection of the MV schedule and the MC′ schedule. It is clear from figure 8.4 that $l_{t+2} = l_{t+1}$: the level of unemployment does not return to its 'normal' value, and if θ stays at its normal value, it will never do so. If the MV schedule were steeper than that drawn in figure 8.4, then employment would rise in period $t+2$, but not enough to offset the drop in period $t+1$. Hence there is at least a part of the temporary shock that persists for ever.

Hence another implication of the model is that:

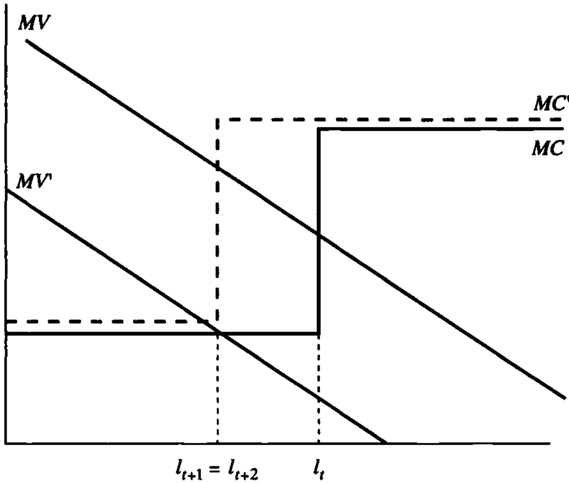


Figure 8.4 Persistence of large shocks

Large transitory shocks have permanent effects on the level of employment

Notice, however, that this effect is smaller than the transitory effect that would be obtained if the firm was not paying efficiency wages (i.e. if the MC schedule was flat). Furthermore, the effect of the shock will persist only until a new shock *changes* the labour force.

Hence it is possible to obtain hysteresis in the context of an efficiency wage model. The pattern of employment obtained in our model differs from the Blanchard and Summers study in that the size of the shock matters: small shocks have no effect at all, whereas big shocks have permanent effects. Another difference is that if the shock is really big, hysteresis will be only partial, whereas in the Blanchard–Summers model, employment follows a random walk. Therefore, our model is more similar to the 'linear adjustment cost' or 'firing cost' view of hysteresis.

Joint behaviour of wages and employment in general equilibrium

The above results were derived under the assumption that the alternative wage, denoted by a, is a constant. Although this can hold in general equilibrium for a given rule for the determination of unemployment benefits, this assumption prevents us from analysing the implications of the fact that the alternative wage is likely to rise in booms. What would be needed, ideally, is to derive the joint behaviour of real wages and employment under the assumption that the wage is determined by (6). This is an almost intractable problem. In order to get some insight we work out a 'linearized' version of the model. This will allow us to derive a closed form solution for the joint behaviour of employment and the real wage. The signs turn out to be in general ambiguous, and we do some simulations to study how various parameters affect the system.

We first 'convexify' (8) by assuming that the wage in period $t+1$ is equal to:

$$w_{t+1} = a_{t+1} + b\psi(l_{t+1}/l_t) \tag{13}$$

where ψ is a differentiable, decreasing function such that $\psi(0) = +\infty$ and $\psi(+\infty) = 1$. In the remainder of the section we shall assume that the total contribution of ψ to labour costs is not decreasing in l, implying $\psi(x) + x\psi'(x) \geq 0$. a_{t+1} is defined by:

$$a_{t+1} = (E_t U_{t+1} - \gamma E_t U_{t+2}) + 1 - 1/x.$$

(13) is consistent with (5) and (6) if, for example, we assume that the firm's headquarters set total employment, but do not perfectly control how employment changes are carried out. More specifically, assume that the firm is divided in I 'units' and that the evolution of the labour force in unit i is determined by:

$$\log l_{it+1} = \log l_{it} + \log l_{t+1} - \log l_t + \varepsilon_{it+1}$$

where l_{it} is employment in unit i at time t, l_t is the firm's aggregate employment, and ε is a unit-specific shock normally distributed with mean zero and variance σ^2. If the number of units is large enough then

$$\sum_{i=1}^{I} l_{it} = l_t$$

in every period provided this holds in period 0.

We can then compute the probability of losing one's job which is given by:

$$p_t = \int_{-\infty}^{-\log(l_{t+1}/l_t)} \frac{1}{\sqrt{2\pi}\sigma} \exp(-\varepsilon^2/2\sigma^2)(1 - (l_{t+1}/l_t).e^{\varepsilon})d\varepsilon = p(l_{t+1}/l_t).$$

It is clear that p_t is a decreasing function of l_{t+1}/l_t, equal to zero at $l_{t+1}/l_t = +\infty$ and 1 at $l_{t+1}/l_t = 0$. According to (6), the wage must follow an equation like (13), with $\psi(l_{t+1}/l_t) = 1/(1 - p(l_{t+1}/l_t))$.

This convexification has a cost: we will no longer get hysteresis, but only a positive serial correlation of employment in face of i.i.d. shocks. This correlation can however be made arbitrarily close to one by having σ close enough to zero.

The problem of the firm is then:

$$\max V(l_t, Z_t) = \theta_{t+1} f(l_{t+1}) - a_{t+1} l_{t+1} - b l_{t+1} \psi(l_{t+1}/l_t) + \delta E_t V(l_{t+1}, Z_{t+1})$$

where Z is a set of exogenous variables which describe the state of the economy at date t, and hence determine the current and future behaviour of the variable a_{t+1}.

The first order condition can be written:

$$\theta_{t+1} f'(l_{t+1}) - a_{t+1} - b\psi(l_{t+1}/l_t) - b l_{t+1} \psi'(l_{t+1}/l_t)/l_t + \delta E_t \partial V/\partial l_{t+1} = 0.$$

The envelope theorem implies:

$$\partial V/\partial l_t = b(l_{t+1}^2/l_t^2)\psi'(l_{t+1}/l_t) < 0.$$

Importing this into the first order condition one gets:

$$\theta_{t+1} f'(l_{t+1}) - a_{t+1} - b\psi(l_{t+1}/l_t) - b l_{t+1}/l_t \psi'(l_{t+1}/l_t) + \delta E_t b(l_{t+2}^2/l_{t+1}^2)\psi'(l_{t+2}/l_{t+1}) = 0.$$

We now derive the behaviour of a_t. We first compute the expected utility of an unemployed worker. Let w_u be unemployment compensation, assumed constant, and n be the labour force. Then:

- The number of people seeking work in period $t+1$ is: $n - l_t + p_t l_t$
- The number of people being hired is: $l_{t+1} - l_t + p_t l_t$
- Therefore, the probability of getting a job in period $t+1$ is $(p_t l_t + l_{t+1} - l_t)/(n_t - l_t + p_t l_t)$.

Hence, one has:

$$U_t = w_u + \frac{p(l_{t+1}/l_t)l_t + l_{t+1} - l_t}{n - l_t + p(l_{t+1}/l_t)l_t} \gamma E_t V_{t+1} + \frac{n - l_{t+1}}{n - l_t + p(l_{t+1}/l_t)l_t} \gamma E_t U_{t+1}. \tag{14}$$

Using (14) and (5) in the definition of a_t one gets:

$$a_t = 1 - 1/x + w_u + E_{t-1} \frac{l_{t+1}\psi(l_{t+1}/l_t) - l_t}{x(n - l_t/\psi(l_{t+1}/l_t))}.$$

Hence the equations describing the behaviour of the system are:

$$\theta_{t+1} f'(l_{t+1}) - a_{t+1} - b\psi(l_{t+1}/l_t) - b l_{t+1}/l_t \psi'(l_{t+1}/l_t) + \delta E_t b(l_{t+2}^2/l_{t+1}^2)\psi'(l_{t+2}/l_{t+1}) = 0 \tag{15}$$

$$a_t = 1 - 1/x + w_u + E_{t-1}\frac{l_{t+1}\psi(l_{t+1}/l_t) - l_t}{x(n - l_t/\psi(l_{t+1}/l_t))}.$$

along with (13), which describes the behaviour of the real wage.

Our strategy is to loglinearize this system around a 'steady-state'.

Let $l_t = \bar{l}(1+y_t)$, $\theta_t = \bar{\theta}(1+\eta_t)$, $a_t = \bar{a}(1+z_t)$, $w_t = \bar{w}(1+s_t)$ where y, z, s and η are small. The steady-state levels can be derived from the formulae:

$$\bar{\theta}f'(\bar{l}) - \bar{a} - b\psi(1) - b\psi'(1) + \delta b\psi'(1) = 0$$

$$\bar{a} = \frac{\bar{l}(\psi(1)-1)}{x(n-\bar{l}/\psi(1))} + 1 - 1/x + w_u$$

$$\bar{w} = \bar{a} + b\psi(1).$$

Since we are interested in dynamics rather than steady-state levels, we will not discuss them.

The dynamics of the system around $(\bar{l}, \bar{a})$ are described by the following equations:

$$c_0\eta_{t+1} + c_1 y_t + c_2 y_{t+1} + c_3 E_t y_{t+1} = 0 \qquad (17)$$

$$z_{t+1} = d_0 y_{t+1} + d_1 E_t y_{t+2} \qquad (18)$$

where:

$$d_0 = \frac{\bar{l}}{\bar{a}x(n\psi(1)-\bar{l})}\left[-\psi'(1)(\psi(1)-1) - \frac{(n-\bar{l})\psi(1)[\psi(1)+\psi'(1)]}{n\psi(1)-\bar{l}}\right]$$

$$d_1 = \frac{\bar{l}}{\bar{a}x(n\psi(1)-\bar{l})}\left[\psi(1)(\psi(1)+\psi'(1)) - \frac{\bar{l}\psi'(1)[\psi(1)-1]}{n\psi(1)-\bar{l}}\right]$$

$$c_0 = \bar{\theta}f'(\bar{l})$$

$$c_1 = b(2\psi'(1)+\psi''(1))$$

$$c_2 = \bar{\theta}\bar{l}f''(\bar{l}) - d_0 - b(1+\delta)(2\psi'(1)+\psi''(1))$$

$$c_3 = \delta b(2\psi'(1)+\psi''(1)) - d_1$$

and:

$$\bar{w}s_{t+1} = [\bar{a}d_0 + b\psi'(1)]y_{t+1} + \bar{a}d_1 E_t y_{t+2} - b\psi'(1)y_t.$$

It can be shown (appendix 2) that (17) has one root between zero and one. If we assume that the other root is outside the unit circle, then (17) has a unique solution and employment follows an AR1 process:

$$y_{t+1} = \lambda y_t + \Phi\eta_{t+1}$$

where λ is the stable root of $c_1 + c_2\lambda + c_3\lambda^2 = 0$, and $\Phi = c_0\lambda/c_1$. It is shown in appendix 2 that $\Phi > 0$. The 'sunspot' case where both roots are inside the unit circle and (17) has an infinity of solutions cannot be ruled out *a priori*. However, it never appears in the simulations that we present below.

Hence what efficiency wage implies here is a positively correlated behaviour of employment in face of i.i.d. shocks. We now turn to the behaviour of the real wage. First, the alternative wage follows the process:

$$z_{t+1} = (d_0 + \lambda d_1)y_{t+1}.$$

Hence the alternative wage is perfectly correlated with employment. As shown by the simulations below, this correlation can be either positive or negative. The real wage follows the process:

$$s_{t+1} = g_0 y_t + g_1 \Phi \eta_{t+1}$$

with:

$$g_0 = (\bar{a}d_0 + b\psi'(1) + \bar{a}d_1\lambda)\lambda - b\psi'(1)$$
$$g_1 = \bar{a}d_0 + b\psi'(1) + \bar{a}d_1\lambda.$$

Once again, the signs are ambiguous and the simulations will yield all kinds of results. The only thing that can be said is that if g_1 is positive, g_0 is positive. Hence one has three cases:

- A positive response of the real wage to past employment and the innovation in employment
- A positive response to past employment and a negative response to the innovation
- A negative response to past employment and to the innovation.

The main conclusion is that the persistence mechanism due to the effort inducement problem does not necessarily imply a countercyclical real wage, as opposed to the partial equilibrium model studied in the previous sections. The downward pressure on wages due to a slump can well offset the upward pressure due to incentives while maintaining some persistence. However, as will appear from the simulations, the degree of persistence consistent with a procyclical behaviour of wages is limited.

Simulation results

The values of $\lambda, d_0 + \lambda d_1, g_0$ and g_1 were computed for a 'plausible' set of parameters implying an average unemployment level of 5 per cent. The interest rate was taken equal to 2 per cent for both firms and people, implying a typical period of about a quarter. The results are parametrized according to three values:

1. x, the probability of being caught shirking. The higher is x, the more efficiency wage considerations are important.
2. $\mu = -\bar{l}f''(\bar{l})/f'(\bar{l})$, the curvature of the revenue function.

Table 8.1. *Response of* $\lambda, d_0+\lambda d_1, g_0, g_1$ *to* μ

$\sigma=0.01$; $x=0.5$

μ	λ	$d_0+\lambda d_1$	g_0	g_1
0.1	0.89	0.85	0.22	−0.17
0.2	0.88	0.70	0.19	−0.20
0.5	0.87	0.27	0.10	−0.31
1.0	0.85	−0.40	−0.02	−0.47
2.0	0.81	−1.56	−0.23	−0.75

Table 8.2. *Response of* $\lambda, d_0+\lambda d_1, g_0, g_1$ *to* x

$\sigma=0.01$; $\mu=0.5$

x	λ	$d_0+\lambda d_1$	g_0	g_1
0.4	0.871	1.13	0.12	−0.39
0.5	0.871	0.27	0.10	−0.31
0.6	0.868	0.10	0.10	−0.28
0.7	0.863	0.001	0.004	−0.27

Table 8.3. *Response of* $\lambda, d_0+\lambda d_1, g_0, g_1$ *to* σ

$\mu=0.5$; $x=0.5$

σ	λ	$d_0+\lambda d_1$	g_0	g_1
0.1	0.53	1.68	0.48	0.36
0.05	0.65	1.66	0.48	0.23
0.025	0.76	1.02	0.31	−0.06
0.01	0.87	0.27	0.10	−0.31
0.005	0.92	−0.16	0.001	−0.42

3. σ, the 'diffusion parameter'.

The results are reproduced in tables 8.1, 8.2 and 8.3 above. The main findings are the following:

- When x increases, λ decreases: when monitoring is more efficient, it is less necessary for the firm to promise higher wages in slumps, which lowers 'pseudo-adjustment costs', thereby diminishing the amount of serial correlation.

- When μ increases, λ decreases: the more curved the revenue function, the more costly it is to deviate from the long-run optimal level of employment. In addition to that, the simulations show that both the wage and the alternative wage tend to be less procyclical when μ increases. For example, one gets the following results for $\sigma = 0.01$ and $x = 0.5$. For low values of μ, the alternative wage is positively correlated with employment and the real wage positively correlated with lagged employment. For high values of μ, these correlations are negative. In addition to that, the response of the wage to the innovation is always negative, all the more so since μ is big. Hence, in order to get a high serial correlation and at the same time a procyclical behaviour of the real wage, it is necessary to assume a low curvature of the firm's revenue function.
- The effect of σ is also interesting. Not surprisingly, when σ decreases, λ increases. This is because when σ decreases, one gets closer to the previous section's case where employment reallocation is fully concentrated and hysteresis holds. However, when σ decreases, the real wage becomes less procyclical. Thus in our example, the real wage is positively correlated with past employment and the innovation in employment for $\sigma = 0.1$ and $\sigma = 0.05$, but the latter correlation becomes negative for σ smaller than 0.025 and the former becomes negative at $\sigma = 0.005$. Hence there is a trade-off between the amount of serial correlation one is able to get and the procyclicality of the real wage. Given that i.i.d. shocks at the firm level were assumed, and that only a small amount of real wage procyclicality is needed in order to match the data, we conclude this section by saying that the model can generate an important amount of serial correlation in employment while implying a procyclical real wage. Hence the previous section's results do not hinge on the convenient assumption that a_t is a constant. The fact that shrinking firms must increase compensation to induce effort from their workers does not necessarily imply a countercyclical real wage at the aggregate level.

Conclusion

The main conclusion of this chapter is that a great deal can be learned from extending the efficiency wage model in order to account for dynamic phenomena.

The main result is that the efficiency wage model can generate hysteresis in unemployment. Hence it can compete with the insider–outsider theory of unemployment in order to explain this persistence. This is an important result in light of the fact that a lot of countries exhibit very persistent unemployment in spite of a very low rate of unionization (USA, France).

Furthermore the amount of persistence generated by insider models is likely to be small, both theoretically and empirically (see Layard, Nickell, and Jackman, 1991).

A natural extension of the model would be to include the calibration exercise of the previous section in a full real business cycle model to calibrate the joint behaviour of output, consumption, investment, unemployment and real wages.

Appendix 1: solution of the firm's optimization problem

The problem is:

$$V(l_t,\theta_{t+1})=\max_{l_{t+1}} \theta_{t+1}f(l_{t+1})-(al_{t+1}+b\max\{l_t,l_{t+1}\}) + \delta E_t V(l_{t+1},\theta_{t+2}).$$

Let us define $E_tV(l_{t+1},\theta_{t+2})=H(l_{t+1})$ (θ_{t+2} is independent of θ_{t+1}); $E_t\partial V/\partial l(l_{t+1},\theta_{t+2})=h(l_{t+1})=H'(l_{t+1})$.

It is clear that the solution must be of the following form:

- for $\theta_{t+1}<\theta_m$, $l_{t+1}<l_t$
- for $\theta_m\leq\theta_{t+1}\leq\theta_M$, $l_{t+1}=l_t$
- for $\theta_M<\theta_{t+1}$, $l_{t+1}>l_t$.

In order to solve this problem we must determine the critical values θ_m and θ_M and the value of l_{t+1} when it differs from l_t.

(a) If $l_{t+1}<l_t$, the first order condition (FOC) is:

$$\theta_{t+1}f'(l_{t+1})+\delta h(l_{t+1})=a.$$

θ_m must be the value of θ such that this condition holds for $l_{t+1}=l_t$:

$$\theta_m(l_t)=(a-\delta h(l_t))/f'(l_t).$$

In this case, the envelope theorem can give us the value of $\partial V/\partial l(l_t,\theta_{t+1})$:

$$\partial V/\partial l=-b$$

(b) If $l_{t+1}>l_t$, the FOC is:

$$\theta_{t+1}f'(l_{t+1})+\delta h(l_{t+1})=a+b.$$

θ_M must be such that this condition is true for $l_{t+1}=l_t$:

$$\theta_M(l_t)=(a+b-\delta h(l_t))/f'(l_t).$$

The envelope theorem implies in this case:

$$\partial V/\partial l=0.$$

(c) Suppose now that θ_{t+1} is between θ_m and θ_M. Then $l_{t+1}=l_t$. As the firm's choice of l_{t+1} is at the kink of the total cost function, we cannot apply the envelope theorem, but we can compute directly $\partial V/\partial l_t$:

$$\frac{\partial V}{\partial l_t}=\theta_{t+1}f'(l_{t+1})\frac{dl_{t+1}}{dl_t}-a\frac{dl_{t+1}}{dl_t}+b\frac{d}{dl_t}\max\{l_t,l_{t+1}\}$$
$$+\delta h(l_{t+1})\frac{dl_{t+1}}{dl_t}.$$

Since in this zone $l_{t+1}=l_t$ we have:

$$dl_{t+1}/dl_t=1;\ d\max\{l_t,l_{t+1}\}/dl_t=1.$$

Hence:

$$\partial V/\partial l_t=\theta_{t+1}f'(l_{t+1})-(a+b)+\delta h(l_{t+1})=\theta_{t+1}f'(l_t)$$
$$+\delta h(l_t)-(a+b)$$

which is between $-b$ and 0 if θ_{t+1} is between θ_m and θ_M.

Let us now compute the h function. Let G be the cumulative density function of θ, $g=G'$:

$$h(l_t)=E_{t+1}\partial V/\partial l_t(l_t,\theta_{t+1}).$$

The above calculations imply that:

$$h(l_t)=-bG(\theta_m)+\int_{\theta_m}^{\theta_M}[\theta f'(l_t)+\delta h(l_t)-(a+b)]g(\theta)d\theta. \qquad \text{(A1)}$$

This implies that $h(l)$ is always lower than 0 and greater than $-bG(\theta_M)$.

Furthermore, after integrating by parts, one gets:

$$h(l_t)=-f'(l_t)\int_{\theta_m(l_t)}^{\theta_M(l_t)}G(\theta)d\theta. \qquad \text{(A2)}$$

Appendix 2: properties of (21)

We have the following results:

1. $\psi'(1)+\psi(1)>0$. This is because we have assumed that the contribution of ψ to total labour costs was increasing in employment.
2. $\psi(1)>1$. This is because $\psi=1/(1-p_t)$.
3. $d_0+d_1>0$. To see this, compute it using the formulae of the text:

$$d_0+d_1=\frac{n\psi(1)(\psi(1)-1)}{\bar{a}x[n\psi(1)-\bar{l}]^2}(-\psi'(1)+\psi(1)[\psi(1)+\psi'(1)])$$

which is positive, since $\psi' < 0$, $\psi(1) + \psi'(1) > 0$ and $\psi > 1$.

4. $d_1 > 0$.
5. $2\psi'(1) + \psi''(1) > 0$. This comes from the second order conditions which must imply that V be locally concave. One has $\partial^2 V/\partial l^2 = -b(2\psi'(1) + \psi''(1)) < 0$.
6. $c_1 > 0$ derives from the previous result.
7. $c_1 + c_2 + c_3 < 0$, since it is equal to $\tilde{\theta}\bar{l}f''(\bar{l}) - (d_0 + d_1) < 0$.
8. That $c_1 + c_2\lambda + c_3\lambda^2$ has one root between 0 and 1 is a corollary of (6) and (7).
9. That $c_0\lambda/c_1 > 0$ is straightforward.

Notes

This chapter was prepared for the University of Warwick Macroeconomic Workshop (July 1993). I thank Olivier Blanchard, Athanasios Orphanides, Michael Piore and Julio Rotemberg for helpful comments and suggestions

1. The model has the feature that wages rise when employment goes down. One might consider this feature as unattractive, since it is observed that the real wage is slightly procyclical. However, countercyclicality of the real wage is not implied in a strict sense by our model. Rather, what is needed is just that total compensation be countercyclical. In particular, in Saint-Paul (1990), we extend the model by allowing the firm to make severance payments, contingent on being fired for a reason other than shirking. It is shown that this generates the same kinked labour cost function, but that the real wage can go either up or down in response to positive increases in employment. It will also be shown below that even if all compensation is in the form of wages, procyclicality of the real wage holds in partial equilibrium, but not necessarily in general equilibrium.
2. More precisely the timing can be described as follows: each period t is divided in three subperiods. In the first subperiod, the firm determines employment l_t for period t. In the second subperiod, the value of θ_{t+1} is known. In the third subperiod, workers produce and are paid wages w_t. What is crucial is that θ_{t+1} is known before period t's production is undertaken. Another assumption is that no firing can occur in the second and third subperiod. This assumption implies that w_t cannot be used to induce workers to work in period t, which simplifies the analysis considerably. Although this structure might seem a little tricky, what is essential in the model is the interaction between the expected change in employment and the incentives to shirk. The particular timing we use will allow us to translate the intuition into a very simple form for the labour cost function, but the results do not rest crucially on the particular set-up of the model.
3. The purpose of this chapter is not to analyse in detail the microfoundations of the efficiency wage model from the point of view of contract theory. For this we refer the reader to Carmichael (1985) and MacLeod and Malcomson (1993). Rather, we take the efficiency wage model as granted and derive its implication for macroeconomic dynamics.
4. An important issue is whether the firm will have an incentive to reveal its private

information concerning tomorrow's value of θ. In Saint-Paul (1990), it is shown that this is indeed the case.

5. In Saint-Paul (1990), it is shown that under this cost structure the distribution of firing probabilities across workers has no effect on the total cost of labour. The assumption of a uniform firing rule is therefore, innocuous.

9 Efficiency, enforceability and acyclical wages

Christian Schultz

Introduction

One of the puzzles of modern macroeconomics is the apparent low volatility of the real wage over the business cycle. Employment tends to vary too much compared with the real wage, at least if the data are to fit standard macroeconomic models (see, for example, Blanchard and Fischer, 1989, for a quick empirical overview). The purpose of this chapter is to present a small model which can add to the understanding of why the real wage is so flat over the cycle.

We will study a simple example of a labour market where a firm employs a number of workers organized in a union. The union decides on the wage, the firm on the employment. The parties engage in a long-term relationship, and (as is well known) this gives them the opportunity to realize better outcomes than the one-shot non-cooperative equilibrium. The firm's revenue is subject to stochastic shocks: in some periods the (linear) demand curve is more favourable to the firm than in other periods. We focus on the efficient outcome sharing the surplus over the non-cooperative outcome equally. Here the wage is lower than in the non-cooperative outcome, and in return the firm employs more workers than it would like to do at this wage. When the parties' discount factor is sufficiently high, this outcome can be implemented in each period in a subgame perfect equilibrium of the infinitely repeated game. Although the firm in each period would like to employ fewer workers at the efficient wage, it restrains from dismissing workers, since the union will punish it for doing so in future periods by setting higher wages.

However, since revenue is fluctuating there are some periods where it is particularly tempting for the firm to deviate. In the chapter's model these are the good periods. If the discount factor is only moderately high, then this temptation may become too big for the firm, and equilibria where the surplus sharing efficient outcome is implemented in each period do not exist. The firm would deviate in the good periods. On the other hand there are no enforcement problems in the bad periods. Evidently, it is best for the parties to implement efficient equilibria. Since the temptation to deviate for the firm is increasing in the wage and it is only in the good periods that this temptation is too high, the best the parties can coordinate on is an equilibrium where the wage is lower in the good periods than in the surplus sharing efficient outcome, and where the wage in return is higher in the bad periods than in the surplus sharing efficient outcome. The enforcement problems therefore induce a wage which is more flat over the cycle than is the case for the surplus sharing efficient outcome. In fact, it turns out that the more flat is the wage, the smaller can be the discount factor without the enforcement problems becoming too serious. If the expected payoffs of the efficient outcome can be implemented, it can always be implemented by a constant wage. Evidently, this wage also fluctuates less than the Stackelberg equilibrium wage.

Another possible explanation of the flat wage over the cycle is offered by the theory of labour contracts (see, for example, the overview in Hart, 1983). If workers are more risk-averse than firms, workers and firms have incentives to construct labour contracts insuring the workers by having (relatively) flat wages over the cycle. In order not to replicate the results of this literature, we consider risk-neutral workers.

This chapter draws on Schultz (1993) where the same issues are considered in a more general setting. Rotemberg and Saloner (1986) were the first to consider the best surplus sharing equilibrium in a model with fluctuating revenue and a moderate discount factor. They consider an oligopoly. Espinosa and Rhee (1989) consider a union and a firm in an infinitely repeated game. They show that the efficient outcomes can be sustained as subgame perfect equilibria for a sufficiently high discount factor. Furthermore, for the case of fluctuating revenue and a moderate discount factor they show that there are equilibria where inefficient outcomes are implemented in each period. They do not consider whether there also are efficient equilibria in this case, and they do not try to characterize the behaviour of wages.

The organization of the chapter is as follows. The second section presents the model, and subgame perfect equilibria of the infinitely repeated game for high discount factors are studied in the third section. The fourth section deals with subgame perfect equilibria of the infinitely repeated game when discount factors are moderate, and the fifth section concludes.

The model

We consider a firm and a union in a long-term relationship. Time evolves from $t=0,\ldots,\infty$. In each period, the firm produces output q by means of labour input L. We assume that the firm has a constant returns to scale technology, so $q = L$. The firm faces a downward-sloping demand curve for its output, the inverse demand curve is $p = \alpha - q$, where p is the price. The intercept α is a stochastic variable which can take on two values α_g and α_b where $\alpha_g > \alpha_b > 0$. The demand shock is independently and identically distributed over time, the probability of the good demand shock α_g is ψ, $0 < \psi < 1$. In each period both the firm and the union learn the shock before any actions are taken. There is symmetric information. This is an important assumption for our results, but also a reasonable one. We conceive of the length of time in which labour contracts are fixed as short relative to the length of the business cycle. After having learned the shock, the union sets the wage w for the period, and the firm then decides on the period's employment L. Before the next period, the union sees how many were employed in the period.

The total revenue of the firm is $T(L,\alpha) \equiv (\alpha - L)L$ and the profit in a period is $\pi(w,L,\alpha) \equiv (\alpha - w - L)L$. The firm seeks to maximize the discounted sum of expected profits, it has a discount factor δ, $0 < \delta < 1$. The union's per period utility function is $U(w,L) \equiv wL$. It seeks to maximize the discounted sum of expected utilities, it has the same discount factor as the firm.

Given the wage set by the union w, the profit maximizing employment for the firm is $L^d(w,\alpha) \equiv \frac{\alpha - w}{2}$. If the parties act non-cooperatively – as in the Stackelberg equilibrium to the one shot game – the union takes this labour demand function as given and chooses the wage which maximizes its utility, i.e. w which solves $\max_w wL^d(w,\alpha)$. The solution is $w^s(\alpha) \equiv \frac{\alpha}{2}$. The resulting employment in the Stackelberg equilibrium is $L^d(w^s(\alpha),\alpha) \equiv \frac{\alpha}{4}$. The union's utility then is $U^s(\alpha) \equiv w^s(\alpha)L^d(w^s(\alpha),\alpha) = \frac{\alpha^2}{8}$, and the firm's profit is $\pi^s(\alpha) \equiv (\alpha - w^s(\alpha) - L^d(w^s(\alpha),\alpha))L^d(w^s(\alpha),\alpha) = \frac{\alpha^2}{16}$.

It is well known that the Stackelberg equilibrium is inefficient. The total surplus to be divided among the parties in a period is $\pi(w,L,\alpha) + U(w,L) = T(L,\alpha)$. The maximal possible surplus is $\max_L T(L,\alpha) = \frac{\alpha^2}{4} > \frac{\alpha^2}{8} + \frac{\alpha^2}{16} = U^s(\alpha) + \pi^s(\alpha)$. The corresponding employment is $L^*(\alpha) \equiv \frac{\alpha}{2}$, higher than the employment of the Stackelberg equilibrium $\frac{\alpha}{4}$. That there is a unique efficient employment level is due to the special utility function assigned to the union. While there is only one efficient employment level, there are many efficient wage rates, the union wishes high wages, the firm low. In the sequel we will concentrate on the efficient outcome where the parties share the extra surplus over the Stackelberg equilibrium equally. This is obviously not the only conceivable outcome, but seems a natural one to focus on. We will call it the efficient surplus sharing outcome. The results we will get will not change qualitatively, if we instead concentrate on some other outcome sharing the surplus in some (other) fixed proportion.

The total surplus to be divided is $\frac{\alpha^2}{4}$. In the Stackelberg equilibrium, the union's utility is $U^s(\alpha)=\frac{\alpha^2}{8}$, and the firm's profit is $\pi^s(\alpha)=\frac{\alpha^2}{16}$, therefore if the parties share the extra surplus the union's utility is $U^*(\alpha)\equiv\frac{5}{32}\alpha^2$, and the firm's profit is $\pi^*(\alpha)\equiv\frac{3}{32}\alpha^2$. With the efficient employment $L^*(\alpha)=\frac{\alpha}{2}$, this means that the wage of the efficient surplus sharing outcome is $w^*(\alpha)=\frac{5}{16}\alpha$.

The repeated game, high discount factors

In this section we will derive conditions under which the surplus sharing efficient outcome can be implemented in a subgame perfect equilibrium of the infinitely repeated game. The per period expected utility for the union in such an equilibrium is $EU^*\equiv\psi U^*(\alpha_g)+(1-\psi)U^*(\alpha_b)$, the expected per period profit for the firm is $E\pi^*\equiv\psi\pi^*(\alpha_g)+(1-\psi)\pi^*(\alpha_b)$. At time t, the history h_t of the game consists of all previous shocks, wages and employments. The union can condition its choice of wage in period t on this history h_t and the periods shock α_t. When the firm chooses employment it can condition on h_t, α_t, and the wage set by the union w_t. A strategy for the union is a sequence of functions w_t: $(h_t, \alpha_t)\rightarrow\mathbb{R}$, similarly a strategy for the firm is a sequence of functions L_t: $(h_t, \alpha_t, w_t)\rightarrow\mathbb{R}$.

We will now construct an equilibrium in which the players in the normal phase play the wages and employments belonging to the surplus sharing efficient outcomes. Since the firm has a short-sight incentive to deviate from the employment level $L^*(\alpha)$, the equilibrium has to specify a punishment for the firm if it deviates. We will take the punishment to be infinite reversal to the one shot Stackelberg equilibrium. This is not the hardest conceivable punishment. Since we are interested in conditions under which efficient outcomes can be implemented, one could argue that we should use the hardest possible punishment for defections, cf. Abreu (1988). However, as argued by the literature on renegotiation – see, for example, Farrell and Maskin (1989) – one may wonder how reasonable mutually devastating punishment phases are. Such punishment phases will, although they are part of a subgame perfect equilibrium, very likely be renegotiated by the players. Since this is not a chapter on the deep issues of renegotiation proofness we will simply content ourselves with looking at equilibria where the punishment consists of trust breaking down and players reverting to non-cooperative play, i.e. reverting to infinite repetition of the Stackelberg equilibrium.

We claim that for a sufficiently high discount factor the following prescribe the strategies of a subgame perfect equilibrium.

- If $t=0$ or only $w^*(\alpha)$, $L^*(\alpha)$ has been played so far, play $w^*(\alpha)$, $L^*(\alpha)$.
- If the union in a period deviates to $w\neq w^*(\alpha)$, the firm in that period plays $L^d(w,\alpha)$. In future periods, play $w^*(\alpha)$, $L^*(\alpha)$ unless somebody deviates.

- If the union plays $w^*(\alpha)$ and the firm deviates from $L^*(\alpha)$, or the union plays $w \neq w^*(\alpha)$ and the firm deviates from $L^d(w,\alpha)$, play $w^s(\alpha)$, $L^s(\alpha)$ in all future periods.

With these strategies, the union can never gain from a deviation. If it deviates, the firm immediately responds with employment on the labour demand curve and this gives the union a utility which is less than or equal to the Stackelberg equilibrium utility, which is less than the per period utility in the equilibrium above. If the union deviates, the equilibrium prescribes the best possible actions for the firm, so the firm will not deviate from these. Evidently, no one can gain by deviating in the punishment phase where the Stackelberg equilibrium is played in each period.

Finally, consider the firm in the normal phase of the equilibrium. If it deviates, it should deviate to the best possible employment $L^d(w^*(\alpha),\alpha)$, this gives the profit $\pi^d(\alpha) \equiv \pi(w^*(\alpha), L^d(w^*(\alpha),\alpha),\alpha) = (\frac{11}{32}\alpha)^2$. If the firm does not deviate in this period, it will receive profit $\pi^*(\alpha) = \frac{3}{32}\alpha^2$ in the period. Thus, the temptation to deviate is $\pi^d(\alpha) - \pi^*(\alpha) = (\frac{5}{32}\alpha)^2$. Notice, that this temptation *depends on the state, and it is highest in the good state.*

A deviation by the firm is punished by reversion to the Stackelberg equilibrium in all future periods. The expected loss of profits induced by this is

$$\sum_{t'=t+1}^{\infty} \delta^{t'-t}[\psi(\pi^*(\alpha_g) - \pi^s(\alpha_g)) + (1-\psi)(\pi^*(\alpha_b) - \pi^s(\alpha_b))]$$
$$= \frac{\delta}{1-\delta}\left(\psi\frac{\alpha_g^2}{32} + (1-\psi)\frac{\alpha_b^2}{32}\right).$$

Notice, that this punishment *is independent of the state.* Clearly, this is due to the assumption that the shock is i.i.d. If the firm should not deviate in a state, the temptation to deviate should be less than the punishment, i.e. it is necessary that

$$\frac{25}{32}\alpha^2 < \frac{\delta}{1-\delta}(\psi\alpha_g^2 + (1-\psi)\alpha_b^2). \tag{1}$$

The right-hand side is independent of the prevailing state, whereas the left-hand side is higher in the good than in the low state. Furthermore, the right-hand side approaches infinity when δ approaches one, it approaches 0 as δ approaches 0, and $\frac{\delta}{1-\delta}$ is increasing in δ. Therefore there are two critical δs, δ_b and δ_g, where $0 < \delta_b < \delta_g < 1$, such that, if $\delta_b < \delta$, then the firm will not deviate in the bad state, and if $\delta_g < \delta$, then the firm will not deviate also in the good state. The higher is the difference between α_g and α_b and the higher is ψ, the higher is δ_g and the lower is δ_b.

The repeated game, moderate discount factors

Suppose now that the discount factor is not sufficiently high to prevent the firm from deviating in the good state, i.e. assume that $\delta_b < \delta < \delta_g$. Then the surplus sharing efficient outcome cannot be enforced in each period. The problem is that the wage in the good period is too high when the firm is to employ the high number of workers $L^*(\alpha_g)$. To see this, notice that at the efficient level of employment, the firm's temptation to deviate as a function of the wage is

$$\begin{aligned} &\pi(w, L^d(w,\alpha),\alpha) - \pi(w, L^*(\alpha),\alpha) \\ &= \left(\alpha - w - \frac{\alpha - w}{2}\right)\frac{\alpha - w}{2} - \left(\alpha - w - \frac{\alpha}{2}\right)\frac{\alpha}{2} = \frac{w^2}{4}. \end{aligned} \tag{2}$$

This temptation is increasing in the wage. When $\delta_b < \delta < \delta_g$, the wage in the good state is too high. On the other hand, the firm's gain from a deviation in the bad state from $L^*(\alpha_b)$ (where the wage is $w^*(\alpha_b) = \frac{5}{16}\alpha_b^2$) is strictly smaller than the punishment. Therefore the firm will not wish to deviate in the bad state even though the wage is increased somewhat in this state. This means that the enforcement problem can be relieved by lowering the wage in the good state, which makes the union worse off, and compensating the union in the bad state by raising the wage there. This can be done without the firm wishing to deviate in the bad state. Therefore, if the discount factor is below δ_g, but not too much below, there is an equilibrium where the players in this way achieve the same expected utility and profit as in the case where the surplus sharing efficient outcome is implemented in each period. Furthermore, this is the only way the players can achieve this outcome. In this case the wage is flatter over the cycle than the wage of the surplus sharing efficient outcome itself.

We now show that if the difference between states is not too large there is a discount factor $\tilde{\delta}$ where $\delta_b < \tilde{\delta} < \delta_g$, such that if $\tilde{\delta} < \delta < \delta_g$, then there is an equilibrium to the infinitely repeated game where the expected payoffs are the same as if the surplus sharing efficient outcome is implemented in each period and where the wage is the same in both states.

Theorem 1: There is a $\tilde{\delta}$ where $\delta_b < \tilde{\delta} < \delta_g$, such that if $\tilde{\delta} < \delta < \delta_g$, then the surplus sharing efficient outcome cannot be implemented in each period in a subgame perfect equilibrium with Stackelberg punishment. If furthermore, $\frac{1}{4}\psi\alpha_g^2 + (1-\psi)\alpha_b(\frac{5}{4}\alpha_b - \alpha_g) > 0$, then there is a subgame perfect equilibrium with Stackelberg punishment with expected average payoffs EU^* and $E\pi^*$, where the wage is constant.

Proof: It follows from the previous section that the surplus sharing efficient outcome cannot be implemented in each period when $\delta < \delta_g$.

Provided employment is efficient in each period, a constant wage giving the union the same expected utility as if the surplus sharing efficient outcome is implemented in each period fulfils

$$\psi w L^*(\alpha_g) + (1-\psi) w L^*(\alpha_b) = \frac{5}{32}(\psi\alpha_g^2 + (1-\psi)\alpha_b^2),$$

implying that

$$w = \frac{10}{32}\left(\frac{\psi\alpha_g^2 + (1-\psi)\alpha_b^2}{\psi\alpha_g + (1-\psi)\alpha_b}\right).$$

Then $w^*(\alpha_g) > w > w^*(\alpha_b)$, as $\alpha_g > \alpha_b$. Let the strategies be as in the equilibrium of the previous section with the only difference that the union in the normal phase plays w rather than $w^*(\alpha)$.

If the union deviates in a period, it can maximally get the Stackelberg utility in that period. If $wL^*(\alpha) > U^s(\alpha)$, then the union will never deviate. This is fulfilled in both states if $w > \frac{\alpha_g}{4}$. This condition can be rewritten $\frac{1}{4}\psi\alpha_g^2 + (1-\psi)\alpha_b(\frac{5}{4}\alpha_b - \alpha_g) > 0$, which is the condition of the Theorem. (This condition is fulfilled for α_g and α_b sufficiently close or ψ sufficiently large.)

Clearly, just as in the previous section, the firm is interested in punishing the union for a deviation.

If $L^*(\alpha)$ is the employment and w the wage, the firm's expected profit is the same as if the surplus sharing efficient outcome were implemented in each period. Hence, since deviations are punished by Stackelberg reversion, the expected punishment is the same as in the previous section, i.e. equal to $\frac{\delta}{1-\delta}\left(\psi\frac{\alpha_g^2}{32} + (1-\psi)\frac{\alpha_b^2}{32}\right)$. Again this is increasing in δ. The temptation to deviate is given by (2). As $w^*(\alpha_b) < w < w^*(\alpha_g)$, the temptation to deviate in the present equilibrium is greater than it was in the equilibrium of the previous section in the bad state but smaller than it was in the good state. As the punishment is increasing in the discount factor the theorem is proved. □

Notice that the condition on the αs and ψ is much stronger than needed in order to ensure that the union will not deviate. It ensures that the union will not even get a short-term gain from a deviation. If it was not fulfilled one could still construct equilibria where the union would not deviate by making longer punishment phases for the union. As the above is just an example, we did not find it worthwhile to introduce these extra complications.

If the discount factor is exactly equal to $\tilde{\delta}$ of the theorem, then the only equilibrium which implements expected payoffs equal to those of the surplus sharing efficient outcomes is the one with the wage of the equilibrium. If the discount factor is higher, there are equilibria giving the same expected payoffs where the wage varies a little over states, but as long as the discount factor is less than δ_g, then the wage rate has to vary less than that of the surplus sharing efficient outcomes.

Concluding remarks

One may wonder how shocks to the slope of the demand function or to the productivity of the firm affect the implementable wages over the cycle. In Schultz (1993) it is shown that for a class of examples with linear demand and constant returns to scale productivity shocks and shocks to the slope of the demand function give the same result as we got here. The temptation to deviate for the firm is greatest in the good state. If the discount factor is moderate, then implementable outcomes will have a flatter wage over the cycle than the surplus sharing efficient outcome.

Another interesting question is how the results are affected if the shocks are not i.i.d., but follow a Markov process with persistence. In Schultz (1993) it is shown (in an example) that this does not change the qualitative nature of the results.

The Stackelberg equilibrium wage is $w^s(\alpha) \equiv \frac{\alpha}{2}$, it is higher in the good than in the bad state. It therefore follows directly that the wage implementing the surplus sharing efficient outcome for moderate discount factors is flatter than the Stackelberg equilibrium wage over the cycle.

10 Business fluctuations, worker moral hazard and optimal environmental policy

Jon Strand

Introduction

The main purpose of this chapter is to study the relationship between environmental and employment variables in an economy subject to worker moral hazard, and the way this relationship is affected by environmental policy. We consider two models. In model 1 (dealt with in the second section) firms face a stationary environment, while in model 2 (the third section) the (exogenous world market) prices of firms' products fluctuate randomly between a high and a low level, p_{OH} and p_{OL}. The main focus of the chapter is model 2, model 1 serving mainly as benchmark for the discussion of model 2.

In both models we assume that there is a worker moral hazard problem in production, in a way analogous to Shapiro and Stiglitz (1984), as workers must have incentives to put up a given total effort level, enforced through a wage exceeding the level clearing the labour market. This implies that employment in the economy in general is inefficiently low. In our set-up total effort has two components, namely output related effort, and effort that serves to reduce pollution discharges from the firm (which are assumed to be perfectly observable by outsiders). The firm can decide how a total worker effort x is to be composed of effort affecting output, y, and that affecting environmental care, z. The firm is subject to a constant unit pollution tax t, set by the government. The firm's output may also be subsidized or taxed by the government, at a given constant rate.

For model 1 (dealt with on pp.217–24) we show that the level of employment (for the representative firm and for the economy as a whole) in general is inefficiently low at a solution with no output subsidies. Assuming that some unemployment always remains as part of an optimal solution, first best can be implemented by subsidizing output sufficiently, and letting t exceed the (constant) marginal social damage cost from pollution, v. With no output subsidies, the government may, as part of a constrained optimal solution, choose to set t either higher or lower than v. $t < v (t > v)$ when an increase in t also affects employment, N, negatively (positively). With no effect of t on N, $t=v$. We also study balanced budget output subsidies financed by pollution taxes. Under some assumptions these may implement a first best solution. When they do not, such policies are still more efficient than using pollution taxes above. There is then always a tendency for t to exceed v, at the solution chosen by the government.

In model 2 (pp.224–30) we extend model 1 to a context of fluctuating demand, by assuming that firms' output prices in the world market jump stochastically between a high and a low level, p_{OH} and p_{OL}, according to a Poisson process. In Strand (1991, 1992a) various aspects of such an economy have been explored. I have shown there that for small price variations, output and employment in each firm are fixed over time, while for larger price variations employment varies, but there is an extra inefficiency to employment in high-demand periods. In the context of the present model our conclusions now differ substantially, depending on whether $N(H)=N(L)$, or $N(H)>N(L)$. When $N(H)=N(L)$, a small increase in $N(L)$ now automatically also raises $N(H)$. This implies that policies which increase $N(L)$ now become more advantageous than in the one-state case. In particular, when only pollution taxes are used, t should now exceed v by more for a given positive $\partial N(L)/\partial t_L$, and both the L-period employment subsidy rate and pollution tax rate should be higher given a balanced government budget. In H periods employment subsidies are now useless (as long as $N(H)$ remains unchanged), and the government prefers to set $t=v$. When we impose an intertemporal budget balance restriction, revenues from pollution taxes in both states should now be used to subsidize L-period employment, i.e. the government budget will be underbalanced in L periods, and both t_L and t_H should exceed v.

When $N(H)>N(L)$ at relevant solutions, the government's policy in each individual state is now much as for model 1. The only difference is that since the degree of employment distortion is greater in H periods than in L periods, the employment subsidy rate should be higher in H periods, and the government budget underbalanced in H periods, given an intertemporal budget balance restriction. Also here, t_L and t_H should both exceed v.

The chapter provides a new approach to the analysis of employment determination under fluctuating demand, by integrating work on environmental policy when workers' efforts are hard to enforce (Strand, 1992b, 1992c) with work on business cycles under worker moral hazard (Strand, 1991, 1992a). The basic idea is that firms' pollution to a significant degree depends on the care taken by employees to avoid it, and that this care is part of an effort subject to enforcement by the firm. These higher demands on workers' environmental care will affect firms' behaviour in other respects, through changes in the required non-shirking wage and/or other types of effort. Here we have simplified by assuming that total effort is constant, and that required increases in environmental effort thus do not affect the wage but only reduce productive effort requirements. This framework still permits us to derive several interesting conclusions, as referred to above. The most interesting of these can be restated briefly. First, when the government can only use a proportional pollution tax, this tax will be set higher or lower than marginal damage cost from pollution, depending on whether the tax affects employment positively or negatively. Secondly, the government revenue created by pollution taxes also provides room for more efficient policies by subsidizing employment, and gives the government incentives to increase the pollution tax more in excess of damage costs. Thirdly, with demand fluctuations, we find for small variations in output prices that employment subsidies are more advantageous in low-demand periods, and that the pollution tax then should deviate more from marginal damage cost, while for larger fluctuations the opposite may be the case. For the conduct of environmental policy over the business cycle we thus reach no simple and clearcut conclusion: sometimes the environmental tax should be raised going from a high- to a low-demand period, and at other times it should be lowered.

The amount of other literature directly related to the current chapter is limited. Besides the author's own studies already referred to (Strand 1991, 1992a, 1992b, 1992c), we may mention a recent study by Gabel and Sinclair-Desgagné (1993), which studies managerial incentives for environmental policy in a setting where managers may allocate effort among different tasks, among which is environmental care. That study uses a more traditional principal–agent framework, whereby the managers' compensation system trades off productive efficiency and insurance optimally for risk-averse managers, but does not consider consequences for environmental policy. There also exists an earlier literature discussing efficient environmental policy under different types of market imperfections, e.g. Buchanan (1969), Lee (1975), Sandmo (1975) and Barnett (1980).[1] These studies generally show that market imperfections create incentives for the government to set environmental taxes different from marginal damage cost. We reproduce this basic result in a setting of worker moral hazard, but the scope of the present chapter is wider, compared both to these and the Gabel and Sinclair-Desgagné study, integrating environmental policy in a macroeconomic and employment fluctuation framework. We may finally mention the recent work on constrained-efficient environmental policy, by Bovenberg and van der Ploeg (1992, 1993) and Strand (1993a). Bovenberg and van der Ploeg show that employment may decline with higher government preferences for environmental care, even when the employment tax is reduced (and the natural resource tax rate increased). This is generally in contrast to the main results from the present model. Strand (1993a) shows that in a distorted economy where individuals' behaviour is affected by the pollution level, environmental taxes may be either above or below marginal damage cost, and that low taxes generally are conducive to high employment. Note however that none of these studies specifies any microeconomic model whereby labour market imperfections are explicitly introduced, in contrast to our model.

Model 1: the stationary solution

In this first model we study a stationary economy, with no changes over time in the parameters facing the firms, nor in their behaviour. Assume that the number of firms in the economy is constant and normalized to one, and their output given by

$$Q=f(yN), f'>0, f''<0 \tag{1}$$

where N is employment, $\bar{N} \geq N$ the labour force, and y is work effort put up by workers. Pollution from the firm is given by

$$h=h(N,z),\ h_N>0,\ h_Z<0,\ h_{NN}\geq 0,\ h_{ZZ}>0,\ h_{NZ}<0, \tag{2}$$

where z measures workers' 'environmental effort' to avoid pollution. Increases in z reduce pollution, but at a decreasing rate while pollution increases with N, possibly at an increasing rate. $h_{NZ}<0$ implies that increases in z for all workers reduce pollution more, the more workers there are.[2]

We assume that firms have limited capacity to monitor workers' efforts, and that both y and z must be enforced through an 'efficiency wage' in a manner analogous to that of Shapiro and Stiglitz (1984). Workers are controlled simultaneously for both regular work effort and environmental effort, at random intervals which are exponentially distributed with parameter q. Define $EV(N,N)$ as the worker's expected lifetime value from non-shirking with respect to both y and z, and $EV(S,S)$ as the corresponding value when shirking with respect to both. We then have

$$rEV(N,N)=w-y-z+b[EV_u-EV(N,N)] \tag{3}$$

$$rEV(S,S)=w+(b+q)[EV_u-EV(S,S)]. \tag{4}$$

Here r is the common discount rate, b an exogenous turnover rate among workers, and EV_u the value of being currently unemployed. (4) implies that the worker is always fired when caught shirking with respect to $x = y + z$, where x is total effort. It can then be shown that the relevant single non-shirking constraint for workers with respect to x is $EV(N, N) \geq EV(S, S)$.[3] EV_u is defined by

$$rEV_u = s + c[EV(N, N) - EV_u], \tag{5}$$

given that all firms always enforce efforts and behave identically, where s is the opportunity cost of labour (taken to be the value of non-market activities), and c the rate of jobfinding. We will assume that the firm can decide on y and z individually, at non-negative levels, but that their sum $y + z = x$ is exogenous and constant. This implies that the total on-the-job effort of a worker who works efficiently is given, but that it can be distributed between productive effort and environmental effort in any way possible decided (and, we may assume, dictated) by the firm.

At an optimal solution for the firm, the worker non-shirking constraint $EV(N, N) \geq EV(S, S)$ is assumed to be binding.[4] This leads to the following constraint on the wage:

$$w \geq \frac{r + q + b}{q} x + rEV_u \equiv w^*, \tag{6}$$

or alternatively using (5), when all firms behave identically,

$$w \geq s + \frac{r + q + b + c}{q} x. \tag{6a}$$

Since EV_u is exogenous to the individual firm, however, (6) is the relevant condition on which to base the firm's decision on what levels of y and z to enforce. Here w^* is defined as the non-shirking constrained wage. Note that since at equilibrium, flows into and out of unemployment must be equal, and that there is no shirking (assuming homogeneous labour), $c \cdot (\bar{N} - N) = b \cdot N$, and thus $b + c = [\bar{N}/(\bar{N} - Nb)]b = b/u$, where u is the rate of unemployment. We here see that as $u \to 0$, i.e. full employment, $w^* \to \infty$, implying that full employment in this sense can never be reached. We also see that for any (less than full) employment level, $w^* > s + y + z = w_c$, where w_c is the standard competitive wage in this case, for the given efforts y and z. This implies some inefficiency in the allocation of labour in the unregulated market solution. We will, however, assume that $f'(\bar{N}) < s + y + z$, given that y and z are also socially optimal. As seen below, this assumption implies the possibility for the government in principle to eliminate these labour market inefficiencies, through appropriate taxes and subsidies.

The firm is assumed to maximize profits given by

$$\Pi = pf(yN) - wN - th(N, z), \tag{7}$$

where p is the output price facing firms, and t is a unit tax on pollution, whose level is assumed perfectly observable by the government.[5] From the assumptions made above, $w \geq w^*$ will be a strictly binding constraint on firms in maximizing profits with respect to employment, N, and the required levels of work and environmental effort, y and z. Above we have seen that only the sum $y+z=x$ is of importance for the determination of w. Assuming x constant implies $dz=-dy$.[6] The first order conditions for the firm with respect to N and z are then

$$\frac{\partial \pi}{\partial N}=pyf'(yN)-w-th_N(N,z)=0 \tag{8}$$

$$\frac{\partial \pi}{\partial z}=-pNf'(yN)-th_z(N,z)=0. \tag{9}$$

The effects of changes in p and t on N and z are now found as follows:

$$\frac{\partial N}{\partial p}=\frac{f'}{D}[(yh_{zz}+Nh_{Lz})\cdot t+pN] \tag{10}$$

$$\frac{\partial N}{\partial t}=\frac{1}{D}[ph_z f'+pNf''(Nh_N+yh_z)-t(h_N h_{zz}-h_z h_{Nz})] \tag{11}$$

$$\frac{\partial z}{\partial p}=-\frac{f'}{D}[(Nh_{NN}+yh_{Nz})\cdot t+py] \tag{12}$$

$$\frac{\partial z}{\partial t}=\frac{1}{D}[ph_N f'+pyf''(Nh_N+yh_z)-t(h_z h_{NN}-h_N h_{Nz})]. \tag{13}$$

Here

$$D=(py^2f''-th_{NN})(pN^2f''-th_{zz})-(pf'+pyNf''+th_{Nz})^2, \tag{14}$$

which is positive when the firm's second order conditions for maximum profits are fulfilled. A necessary condition for an internal solution with respect to both N and z is here that $yh_{zz}+Nh_{Nz}>0$, and thus $\partial N/\partial p>0$.[7] The other derivatives cannot be signed unambiguously from our assumptions. However, when $h_z h_{NN}<h_N h_{Nz}$ and $Nh_N+yh_z<0$ (which both seem reasonable), $\partial z/\partial t>0$. Intuitively, z increases in t since a higher environmental tax makes it worthwhile for the firm to require more effort for pollution avoidance, and relatively less for production activities. Reasonably also, $\partial N/\partial t$ and $\partial z/\partial p$ are both ambiguous. An increase in t thus lowers general profitability but increases the marginal productivity of N since y generally drops. This yields counteracting forces on the demand for N. Similarly, an increase in p increases N, but it is unclear how y, and thus z, are affected. If the marginal efficiency in lowering pollution from increasing z is improved when L increases, z is likely to increase as well, while in other cases z could decrease. We will here make no specific assumptions about the signs of $\partial N/\partial t$ and $\partial z/\partial p$.

Consider finally the effect of p and t on z when N is fixed. We then find, differentiating (9):

$$\frac{dz}{dp}=\frac{Nf'}{pN^2f''-th_{zz}}<0 \tag{15}$$

$$\frac{dz}{dt}=\frac{h_z}{pN^2f''-th_{zz}}>0. \tag{16}$$

Here, consequently, environmental effort is always lowered when p increases, since increasing p makes it more profitable to increase y. As before, there is a positive effect of increasing t on z, which may be weaker or stronger than in the case of variable N (it is generally stronger here when N then decreases).

Define now the government's utility from firms' activity by

$$G=p_0f(yN)-(s+y+z)N-vh(N,z). \tag{17}$$

Here p_0 is the demand determined (e.g. world market) price of output, and thus $p-p_0=\rho$, a government subsidy rate on output. v is the (constant) marginal social damage cost of pollution. The government is considered risk-neutral and has no preferences for own revenues *per se*, and $s+y+z=s+x$ is the social opportunity cost of labour given that not all workers in the economy are employed.

Consider first the unconstrained optimal solution for the government, i.e. that chosen if the government could choose N and z directly. This is found maximizing (17) with respect to N and z (noting that $dy=-dz$), yielding

$$\frac{\partial G}{\partial N}=p_0yf'(yN)-(s+x)-vh_N\geq 0 \tag{18}$$

$$\frac{\partial G}{\partial z}=-p_0Nf'(yN)-vh_z=0. \tag{19}$$

Here (18) can be fulfilled with inequality if and only if $N=\bar{N}$, and

$$p_0yf'(y\bar{N})-vh_N(\bar{N},z)>s+x \tag{18a}$$

for the levels of y and z then found from (19). In such a case the first best implies that all workers are employed. The same holds when (18) is fulfilled with equality for $N=\bar{N}$. In other cases we have

$$p_0yf'(yN)-vh_N(N,z)=s+x, \qquad N<\bar{N}. \tag{18b}$$

In such cases it is never socially optimal for all workers to be employed, since their marginal return would then be higher in other activities.

Consider now the possibility for the government to implement a first best solution, using the instruments $\rho = p - p_0$ and t, and assuming no government budget restrictions. Given that (18b) holds, (18)–(19) are identical to (8)–(9) given that

$$\frac{p-p_0}{p_0} = \frac{t-v}{v} = A, \tag{20)–(21}$$

where A is a positive constant, given by

$$A = \frac{r+b+c}{q} \cdot \frac{x}{s+x}. \tag{22}$$

(20)–(21) imply that implementation of a first best in this case requires a positive subsidy rate on output, and a pollution tax in excess of marginal social damage cost.[8] Moreover, the relative subsidy rate on output is to equal the relative pollution tax rate in excess of marginal social damage cost, and both are to equal A. The constant A expresses the degree of imperfection in the labour market due to the efficiency wage structure. We see that A may be sizeable in particular when q is relatively small and x large relative to s. Note from above that when all firms behave identically, A may be written as

$$A = \frac{1}{q}\left(r + \frac{b}{u}\right)\frac{x}{s+x}. \tag{22a}$$

Thus when $f'(\bar{N}) > s + x$, efficiency would require $u = 0$ and thus $A = \infty$. We find the following expression for the firm's profits in an implemented first best solution:

$$\Pi_F = (1+A)[p_0 f(yN) - (s+x)N - vh(N,z)], \tag{23}$$

which tends to infinity as u tends to zero.

The implication of this is that whenever $f'(\bar{N}) \geq s+y+z$, a first best is generally unattainable, since an infinite amount of subsidy is necessary for its implementation. This is the same conclusion as was reached in a pure work effort context (with no externalities) by Shapiro and Stiglitz (1984). Note here that when $f'(N) < s+x$, a first best may be attainable in our model, even in the absence of profit taxes or net subsidies of firms. Net subsidies S are namely given by

$$S = Ap_0 f(yN) - (1+A)vh(N,z). \tag{24}$$

Here S is always negative whenever A is sufficiently small, and positive when A is sufficiently large. The idea behind the implementation of an efficient solution when A is small is of course that the revenue from efficient pollution taxes gives room for a sufficient amount of output subsidy for the first best to be implemented.[9]

In the following we will generally assume that a first best is *not* implementable in a budget balancing way (ruling out profit taxes considered by Shapiro and Stiglitz), by assuming that A is above its critical value yielding $S=0$ in (24). We will instead consider two different constrained-efficient solutions, namely (a) the case with only pollution taxes, and (b) the case of budget balancing output subsidies and pollution taxes.

Environmental taxes only

In this case the government maximizes (17) with respect to t, taking into consideration that N and z are affected by t via (11) and (13). We find the following expression:

$$\frac{dG}{dt}=[p_0yf'(yN)-(s+x)-vh_N]\frac{\partial N}{\partial t}-[p_0Nf'(yN)+vh_z]\frac{\partial z}{\partial t}=0. \tag{25}$$

Here the first square bracket and $\partial z/\partial t$ must both be positive. $\partial N/\partial t$ is generally ambiguous from the arguments above. When $\partial N/\partial t<0$ (>0), the term in front of $\partial z/\partial t$ must be positive (negative), implying (using (9)) that $t<v(t>v)$ at the constrained-optimal solution for the government. Intuitively, when $\partial G/\partial z=0$, $t=v$. When G is affected negatively through N (which is the case when $\partial N/\partial t<0$), G must be affected positively through z. This implies that z is below its optimal value, which is the case when $t<v$. The opposite ($t>v$) holds when G is affected positively through N ($\partial N/\partial t>0$). In general the level of t chosen by the government can then be either below or above marginal damage cost.

Consider next the case where N is unaffected by changes in t. In such a case the government always sets $t=v$, i.e. selects a 'first best' with respect to z alone.

Balanced budget changes in t *and* p

We now assume that the government may subsidize firms' output price p_0 at a rate $\rho=p-p_0$, but in a balanced budget way, i.e. such that total output subsidies cannot exceed total revenues from taxation of pollution. We will assume that A in (24) above exceeds the level yielding $S=0$, i.e. the first best is not implementable in this way for the government. This generally implies that the budget balance restriction will be binding for the government, and the budget always balanced at equilibrium. This implies the condition

$$(p-p_0)f(yN)=t\cdot h(N,z). \tag{26}$$

When t changes, p thus changes according to

$$dp=\frac{h(N,z)}{f(yN)}\,dt=\frac{p-p_0}{t}\,dt. \tag{27}$$

We thus have in this case (letting subscripts BB denote budget balance)

$$\left(\frac{\partial N}{\partial t}\right)_{BB}=\frac{p-p_0}{t}\frac{\partial N}{\partial p}+\frac{\partial N}{\partial t} \tag{28}$$

$$\left(\frac{\partial z}{\partial t}\right)_{BB}=\frac{p-p_0}{t}\frac{\partial z}{\partial p}+\frac{\partial z}{\partial t}. \tag{29}$$

The optimality condition for the government becomes identical to (25) except that $\partial N/\partial t$ and $\partial z/\partial t$ from (11) and (13) are replaced by (28)–(29). From the discussion above of the partial derivatives of N and z we now find that $(\partial z/\partial t)_{BB}$ is still most probably positive, and that $(\partial N/\partial t)_{BB}$ is now much more likely to be positive than in (11), in particular when $(p-p_0)/t$ is not too small. Since $\partial G/\partial N>0$ as before, $(\partial z/\partial t)_{BB}$ and $(\partial N/\partial t)_{BB}$ both positive implies $\partial G/\partial z<0$. Comparing $\partial G/\partial z$ to (9) now yields that $t>v$ even when $\partial G/\partial z=0$ (since $p>p_0$). In this case we thus in most likelihood have $t>v$, using the results derived for case (a).

When a pollution tax is accompanied by a balanced budget output subsidy, it is thus more likely that the tax will be above marginal damage cost from pollution, and more so when it is above v in both cases.

Consider next a balanced budget increase in p and t when N is fixed. Note that in this case there is nothing to gain in terms of efficiency by increasing both p and t, since as noted for case (a), a first best can be implemented for z using t alone. We may still derive the optimal t in this budget balancing case. We find the following relationship between t and v:

$$t=\frac{1}{1-\dfrac{vh(L,z)}{p_0f(yL)}}\cdot v>v. \tag{30}$$

t must consequently also in this case exceed v. The reason for this is that p now is raised above its socially efficient level. This must be counteracted by a similar increase in t, to make environmental effort and thus pollution constrained socially efficient.

Consider now the effects of a (permanent) increase in p_0. For given t, and $\rho=0$, these effects on N and z are given by (10) and (12) with $p-p_0$. In general such an increase in p_0 will also induce a change in t. We will however argue that the effect through t is small compared to that through p_0 directly.

A possible effect is likely to work via $\partial G/\partial z$ in (25), where the term p_0Nf' increases and makes $\partial G/\partial z$ lower, thus increasing t. Such an effect will in case (b) be present both with a change only in t and with balanced budget changes in t and ρ.

We may also study the effect of a change in p_0 when N is fixed. Then as before $t=v$ independently of p_0, and the government policy is not affected. The change in z is then given by (15).

Model 2: environmental policy under productivity fluctuations

We now consider an extension of model 1 by letting the output price facing each individual firm fluctuate between a high (H) and a low (L) level, p_H and $p_L \in (0,p_H)$. To keep matters simple we now focus on one particular firm, taking the rest of the economy as stationary and exogenous.[10] The transitions between H and L, and L and H respectively, are governed by Poisson processes with constant transition rates equal to a_1 and a_2. In the long run the relative amounts of time spent in the H and L states are then $a_2/(a_1+a_2)$ and $a_1/(a_1+a_2)$, respectively.[11] Using $V_N(i)$ and $V_L(i)$ to denote non-shirking and shirking (with respect to both y and z) respectively, in states $i=H, L$, we may define

$$rV_N(L)=w(L)-x+b[V_u-V_N(L)]+a_2[V_N(H)-V_N(L)] \tag{31}$$

$$rV_S(L)=w(L)+(b+q)[V_u-V_S(L)]+a_2[V_N(H)-V_S(L)] \tag{32}$$

$$\begin{aligned} rV_N(H)=w(H)-x+[b+a_1(1-n)][V_u-V_N(H)] \\ +a_1n[V_N(L)-V_N(H)] \end{aligned} \tag{33}$$

$$\begin{aligned} rV_s(H)=w(H)+[b+q+a_1(1-n)][V_u-V_s(H)] \\ +a_1n[V_N(L)-V_s(H)]. \end{aligned} \tag{34}$$

Here (as in Strand, 1991, 1992a) $n=N(L)/N(H)$ is the fraction of the firm's labour force in H periods that is retained in L periods, where $n \in (0,1]$. Possible L-period layoffs are assumed to be made at random, and laid off workers are not recalled nor given any economic compensation from the firm.[12] For non-shirking to be enforced the conditions $V_N(L) \geq V_S(L)$ and $V_N(H) \geq V_S(H)$ must be imposed. Assuming that these are both fulfilled with equality at equilibrium, and taking V_U as exogenous, the non-shirking constraints are (dropping subscripts)

$$V(i)-V_u=\frac{1}{q}x,\ i=H,L, \tag{35}$$

while the solutions for $w(L)$ and $w(H)$ can be written as

$$w(L)=\frac{1}{q}(r+q+b)x+s+rV_u \tag{36}$$

$$w(H)=\frac{1}{q}[r+q+b+a_1(1-n)]x+s+rV_u. \tag{37}$$

Assume now that the firm maximizes current profits in any given state taking its strategy in the opposite state as given. Current profits in L and H periods are given by

$$\begin{aligned}\Pi(L)&=p_L f(N(L))-w(L)N(L)-t_L h(N(L),z(L))\\&=p_L f(N(L))-\left[\frac{1}{q}(a+b+q)x+s+rV_u\right]N(L)\\&\quad-t_L h(N(L),z(L))\end{aligned} \tag{38}$$

$$\begin{aligned}\Pi(H)&=p_H f(N(H))-w(H)N(H)-t_H h(N(H),z(H))\\&=p_H f(N(H))-\left[\frac{1}{q}(r+q+b+a_1)x+s+rV_u\right]N(H)\\&\quad+\frac{a_1}{q}xN(L)-t_H h(N(H),z(H)),\end{aligned} \tag{39}$$

using that $n=N(L)/N(H)$ in (39). We find the following first order conditions with respect to $N(i)$ and $z(i)$:

$$\frac{\partial\Pi(L)}{\partial N(L)}=p_L y(L)f'(L)-w(L)-h_L h_N(L)=0 \tag{40}$$

$$\frac{\partial\Pi(L)}{\partial z(L)}=-p_L N(L)f'(L)-t_L h_z(L)=0 \tag{41}$$

$$\frac{\partial\Pi(H)}{\partial N(H)}=p_H y(H)f'(H)-w(L)-\frac{a_1}{q}x-t_H h_N(H)\leq 0 \tag{42}$$

$$\frac{\partial\Pi(H)}{\partial z(H)}=-p_H N(H)f'(H)-t_H h_z(H)=0, \tag{43}$$

where L and H now are used as shorthand notation for the levels of y, z and N in the two states. In (42), inequality obtains only if $N(H)=N(L)$. In cases where $N(L)<N(H)$, $w(H)>w(L)$ from (36)–(37). The reason is that L-period layoffs cause an additional exogenous risk for workers in H periods which must be compensated by a higher wage. Moreover, this effect of $N(L)$ on $w(H)$ produces the extra term $a_1(y+z(H))$ in (42), making the firm limit H-period employment and setting $N(H)=N(L)$ on a strictly positive range for p_H-p_L. The solution thus implies that employment is rigid between periods for such a range.[13]

In studying properties of the solution (40)–(43), note that for given $t_L = t$ and $p_L = p$, (40)–(41) for L periods are identical to (8)–(9) for the stationary model. The comparative statics with respect to changes in p_L and t_L are then also identical to those in model 1. As will be seen below, however, the government's optimal solution with respect to t_L (for $p_L = p$) is generally not identical to that in model 1.

Consider next H periods. As already noted, for small relative changes in p, $N(H) = N(L)$, and only z may vary. Then the solution is again equivalent to that of model 1, given $p_H = p$. In this solution we have $t = v$, while $dz(H)/dp_H$ is given by (15). The conclusion then is that with only small relative price changes facing firms between the high- and low-demand periods, employment will stay fixed, the environmental tax in the high-demand period will equal marginal damage cost, and environmental effort will decrease and productive effort increase, in high-demand relative to low-demand periods.

In (42), the term $(a_1/q)x$ creates the difference in firm behaviour from (40). Thus the higher a_1 is, i.e. the shorter H periods are on the average, the greater is the range of $p_H - p_L$ over which N does not vary.

Assume now that $p_H - p_L$ is sufficiently great for $N(H) > N(L)$ at the solution chosen by the firm. (42) is then fulfilled with equality. Since $(a_1/q)x$ in (42) is a constant, this term does not affect any marginal conditions for the firm. This implies that further increases in p_H have the effects given by (40) and (42), for $p_H = p$ and a given t. Thus when t then is a constant, $dN(H)/dp_H > 0$, while $dz(H)/dp_H$ is ambiguous as in (42).

We now study government behaviour when $N(H) > N(L)$. Starting from any given state, the government maximizes present discounted value of the social surplus created by the firm in question, with respect to current parameters t_i and p_i, $i = H, L$. Define these present discounted values by $V(H)$ and $V(L)$. We then have

$$rV(H) = G(H) + a_1[V(L) - V(H)] \tag{44}$$

$$rV(L) = G(L) + a_2[V(H) - V(L)], \tag{45}$$

where $G(H)$ and $G(L)$ are expressions equivalent to (17). These equations yield

$$V(H) = \frac{1}{r(r + a_1 + a_2)}[(r + a_2)G(H) + a_1G(L)] \tag{46}$$

$$V(L) = \frac{1}{r(r + a_1 + a_2)}[a_2G(H) + (r + a_1)G(L)]. \tag{47}$$

First best solutions are also here derived from maximizing $G(H)$ and $G(L)$

individually with respect to $N(i)$ and $z(i)$, $i=H,L$, in a fashion equivalent to (18)–(19) (now assuming equality in (16) for relevant $N(H)$ levels). Properties of such solutions are basically identical to those for model 1 in L periods, i.e. (18)–(19) then still apply, replacing p and p_0 by p_L and p_{0L} (letting p_{0H} and p_{0L} denote the underlying exogenous unsubsidized market prices). For H periods we find the related expressions:

$$\frac{p_H-p_{0H}}{p_{0H}}=\frac{t_H-v}{v}=A_H, \qquad (48)\text{–}(49)$$

where

$$A_H=\frac{r+b+c+a_1}{q}\cdot\frac{x}{s+x}. \qquad (50)$$

Thus in general the relative rates of employment subsidy and pollution tax should both be greater in H periods than in L periods. This is due to the greater amount of distortion from the first best at the market solution in H periods. The result is easily understood by considering a first best solution implemented in L periods, implying a relative output subsidy rate ρ_L^* and pollution tax rate t_L. If the difference between p_H and p_L is small, $N(H)=N(L)$ given the same instrument values ρ_L^* and t_L. This is clearly not first best, and a higher subsidy rate in H periods is required.

We will now study second best solutions for the government assuming either that only pollution taxes are used, or that output may in addition be subsidized, but in a balanced budget way and insufficient for first best implementation.

Pollution taxes only

We may distinguish between two separate cases, namely
(a) $p_{0H}-p_{0L}$ 'small' and $N(H)=N(L)$ at relevant market solutions; and
(b) $p_{0H}-p_{0L}$ 'large' and $N(H)>N(L)$.
We study these in turn.

N(H)=N(L)

In this case we find for H periods:

$$\begin{aligned}\frac{dV(H)}{dt_H}&=\frac{r+a_2}{r(r+a_1+a_2)}\,\frac{dG(H)}{dt_H}+\frac{a_1}{r(r+a_1+a_2)}\,\frac{dG(L)}{dt_H}\\&=\frac{r+a_2}{r(r+a_1+a_2)}\,\frac{\partial G(H)}{\partial z(H)}\,\frac{\partial z(H)}{\partial t_H}=0,\end{aligned} \qquad (51)$$

noting that $dG(L)/dt_H=\partial N(H)/\partial t_H=0$ in this case. This simply implies

$\partial G(H)/\partial z(H)=0$, i.e. $-p_{0H}yf'(H)-vh_z(H)=0$, which in turn implies $t_H=v$, using (43). Intuitively, t_H here only affects $z(H)$, and the government can then set t_H optimally, being concerned about $z(H)$ only.[14]

For L periods we find

$$\begin{aligned}\frac{dV(L)}{dt_L}=&\frac{a_2}{r(r+a_1+a_2)}\left[\frac{\partial G(H)}{\partial N(H)}\frac{\partial N(H)}{\partial t_L}+\frac{\partial G(H)}{\partial z(H)}\cdot\frac{\partial z(H)}{\partial t_L}\right]\\&+\frac{(r+a_1)}{r(r+a_1+a_L)}\left[\frac{\partial G(L)}{\partial N(L)}\frac{\partial N(L)}{\partial t_L}+\frac{\partial G(L)}{\partial z(L)}\cdot\frac{\partial z(L)}{\partial t_L}\right]=0\end{aligned} \tag{52}$$

when $N(H)=N(L)$, $\partial N(H)/\partial t_L=\partial N(L)/\partial t_L$. To determine in which way t_L may deviate from v in this case, assume first the special (but relevant) case of $\partial N(L)/\partial t_L=0$, implying $\partial N(H)/\partial t_L=\partial z(H)/\partial t_L=0$ as well. We then end up with $\partial G(L)/\partial z(L)=0$ and thus $t_L=v$ as the government's optimal solution. Assume instead that $\partial N(L)/\partial t_L>0$. Then $\partial N(H)/\partial t_L>0$, and $dG(H)/dt_L>0$ as well.[15] In the opposite case of $\partial N(L)/\partial t_L<0$, $\partial N(H)/\partial t_L$ and $dG(H)/dt_L$ are also negative, but the absolute value of the last expression is now smaller.[16]

The conclusion is that when $\partial N(L)/\partial t_L>0$, $t_L<v$, while when $\partial N(L)/\partial t_L<0$, $t_L>v$. Qualitatively, this is as in the stationary model 1. The difference between t_L and v is however greater here than in model 1, since the effect on employment now also works via the opposite state H, in which the tax parameter t_L is not directly applied. Note also that the difference t_L-v is larger when it is negative than when it is positive (for given $|\partial N(L)/\partial t_L|$), because of the effects via the term $[\partial G(H)/\partial z(H)]\cdot[\partial z(H)/\partial t_L]$.

$N(H)>N(L)$

In this case (42) holds with equality, and there is no direct interlinkage between the strategies chosen by the firm in the two types of periods. This implies that the discussion from model 1 applies to this case, for H and L periods separately. For H periods there is a slight difference due to the inefficiency of unemployment being greater in this case and $\partial G(H)/\partial N(H)$ generally greater at equilibrium.

Budget balance in each period

In the case of budget balance *as an average* over H and L periods, the following budget condition must hold:

$$a_2B(H)=a_1B(L), \tag{53}$$

where $B(i)$ is the budget surplus in period i given (in a way analogous to (46)) by

$$B(i)=t_i h(i)-(p_i-p_{0i})f(i),\ i=H,L. \tag{54}$$

We will first consider only the case where budget balance is required for any given period, and then subsequently consider whether it is advantageous for the government to make budget transfers between periods. Current budget balance implies, in a fashion similar to (28)–(29),

$$\left(\frac{\partial N(i)}{\partial t_i}\right)_{BB}=\frac{p_i-p_{0i}}{t_i}\cdot\frac{\partial N(i)}{\partial p_i}+\frac{\partial N(i)}{\partial t_i} \tag{55}$$

$$\left(\frac{\partial z(i)}{\partial t_i}\right)_{BB}=\frac{p_i-p_{0i}}{t_i}\frac{\partial z(i)}{\partial p_i}+\frac{\partial z(i)}{\partial t_i}, \tag{56}$$

$i=H,L$. Also now we need to distinguish between the cases where $N(H)=N(L)$, and $N(H)>N(L)$.

N(H)=N(L)

Consider L-period strategies first. Assuming that $\partial z(H)/\partial t_L=0$, we have, much as for $\rho_i=0$ above, that $(\partial N(H)/\partial t_L)_{BB}=(\partial N(L)/\partial t_L)_{BB}$, which are both (almost certainly) positive. When compared both to model 1 and to the case of $\rho_i=0$ above, this makes it more advantageous to raise t_L, since p_L is then also raised, increasing *both* $N(L)$ and $N(H)$. Clearly, this implies that $t_L>v$, and more so than for model 1. Environmental effort is then above its first best level in L periods.

Consider next H-period strategies. Now $\partial N(H)/\partial t_H=0$, and there is nothing to gain in terms of efficiency from increasing p_H in a budget balanced way. Still, provided that balanced budget subsidies are used, t_H is given by an expression analogous to (30).

N(H)>N(L)

We now again have separation of the two types of periods, and t_H and t_L are derived much as in model 1. The only difference is that also here (as with only environmental taxes above) the optimal tax and subsidy rates tend to be somewhat higher in H periods, due to the greater employment distortion in that state.

Intertemporal budget transfers

We may also now distinguish between the two cases considered above.

N(H)=N(L)

In this case it is optimal to tax pollution while paying no employment subsidies in H periods, and to subsidize employment in L periods. Clearly, this implies that it is optimal to transfer funds from H to L periods, making greater employment subsidies possible in the latter periods. Such a strategy will then raise employment in both states, for given implemented environmental efforts.

Moreover, the government's general need for funds to subsidize L-period employment makes it advantageous to set $t_H > v$, and even possibly to *tax* H-period output as long as employment is not then affected negatively.

$N(H) > N(L)$

In such cases the solution deviates little from the one-state model 1, and there is little to gain from intertemporal transfers. However, since the employment distortion generally is greater in H periods, a transfer from L to H periods may now be necessary to implement the government's preferred solution. In both types of periods $t_i > v$, and possibly more so in H periods.

Conclusions and extensions

In this chapter we have studied aspects of efficient government policies to deal jointly with the problems of underemployment and pollution, in an economy where there are inefficiencies in the labour market due to problems of disciplining workers' efforts, and where these efforts are devoted both to the enhancement of firms' output and to the reduction of their pollution levels. A pollution tax will then have a joint effect on output and pollution, and should generally be set below (above) marginal damage cost given that employment is affected positively (negatively) by the same tax. When the government revenue earned through the pollution tax is used to subsidize employment, the preferred pollution tax rate is greater and more likely to exceed marginal damage cost. We also have studied properties of the environmental tax under idiosyncratic fluctuations in demand, and found that the optimal pollution tax levels could be either higher or lower in low demand, and the government budget underbalanced in low demand for small price fluctuations, but underbalanced in high demand for large fluctuations.

The model studied is special since it assumes that the underlying labour market imperfection is of a quite specific nature, namely that of efficiency wages due to the possibility of workers shirking on the job. Still, we argue, the model gives important insights into the relationship between employment and environmental variables, by building a micro-based, coherent model where the nature of the various inefficiencies and policy rules can be studied in some detail. As such, our analysis should be just a start of a much more comprehensive effort to analyse such relationships. It can be argued that the employment and environmental policy areas are the two most important items on the public policy agenda, today and for the foreseeable future, for most rich as well as many poor countries. The interaction of these policies is no less important, but has so far been little studied.

A large number of extensions could be made to improve upon the current analysis and make it more general and topical. We will here briefly outline a few of these.

1. More general environmental problems could be addressed. For one thing, environmental degradation due to the use of particular natural resource inputs could be taken into account, as has already been attempted in a related study by Bovenberg and van der Ploeg (1993). Among other things they address the so-called 'double dividend' hypothesis, whereby higher environmental taxes could at the same time help to increase employment, and answer this in the negative. Our analysis above indicates that the 'double dividend' hypothesis may hold when there are labour market imperfections, but this needs to be studied in a richer model where natural resource inputs are also included. Other extensions would be also to model the effects of firms' investments and of consumer behaviour on pollution and the policies to counteract it. Strand (1993a) attempts to incorporate the latter effect in an otherwise simpler model, and shows that constrained efficient environmental policy rules then generally become more complex. Moreover, general equilibrium effects, with full sets of markets, should be considered.

2. Different and more general types of imperfections may be analysed. More general models of labour market moral hazard and adverse selection, or incorporating imperfections in capital markets affecting the investment in both output- and pollution-related equipment, could be studied within a framework similar to that used here. Alternatively, the robustness and relevance of our conclusions can be checked by constructing different examples of specific informational imperfections or practical limitations on government policies. A relevant line of research, building, for example, on the framework of Gabel and Sinclair-Desgagné, could be to study the firm's incentives to take environmental care in a principal–agent relationship where the incentives of managers would also be of importance, and incorporate this into a more complete markets framework.

3. The nature of the underlying fluctuations causing 'business cycles' in the current model is quite special and could be generalized in future work. One obvious extension is to assume that the firm's output price changes according to a continuous distribution and not in a binary fashion. Such a model is studied in a pure labour market framework by Strand (1993b). I show there that when firms' strategies otherwise are modelled as in the current chapter, many of the same qualitative results are derived: in particular, for a sufficiently dense price distribution employment will not

vary with the price but be depressed to its lower-state level. With a more dispersed distribution of prices employment will vary but less than in a standard competitive economy. Quite likely this model could readily be adapted to incorporate environmental policy as here, with analogous results. Another alternative would be to assume a random walk process for the price, e.g., a continuous-time Brownian motion process. This has been assumed in similar pure labour market contexts, e.g. by Bentolila and Bertola (1990) and by Orphanides (1993). In such cases it is likely that optimal environmental and employment policy will not change much in response to price shocks, but this remains to be demonstrated. Finally, general equilibrium employment fluctuations (and not only partial equilibrium as here) should be studied. Strand (1992a) shows in a pure labour market context that with general equilibrium fluctuations the solution in general becomes more complicated than here. Employment may then fluctuate even less due to workers' alternatives being better in high-demand periods. This probably makes it more advantageous to stimulate employment in high-demand periods, relative to the analysis presented here. This and other extensions must, however, await future research.

Notes

Paper presented at the University of Warwick Macroeconomics Workshop (July 1993). I thank Jon Vislie and conference participants for helpful comments. This research is part of the project 'Environmental policy under asymmetric information', at the SNF Foundation for Research in Economics and Business Administration, Department of Economics, University of Oslo.

1. See Cropper and Oates (1992) for a survey of this literature.
2. Perhaps more satisfactorily, pollution should be a function of the firm's output and total environmental effort. We may then write

$$h = g(f((x-z)N), zN) = h(N, z),$$

implying

$$dh/dN = g_1 y f' + g_2 z \equiv h_N$$
$$dh/dz = (-g_1 f' + g_2)N \equiv h_z.$$

Our assumptions are then made directly on h_N and h_z.
3. See Strand (1992b) for a discussion.
4. A sufficient condition for this to hold is that $f(0) = 0$, i.e. when workers shirk and put up no effort, output is zero. We will adopt this assumption in the following.
5. Note that neither here nor in the following can the government do better by using other environmental policies, e.g. transferable or non-transferable pollution quotas. It is, however, true that in some cases non-linear pollution taxes may be advantageous to use. These would, however, be equivalent to profit

taxes, which we rule out. We may thus limit our attention to proportional pollution taxes.

6. The justification for this assumption may be that working efficiently is something well defined, implying a given total effort level, but that more or less attention can be devoted to environmental care given an efficient total effort.
7. For example, in the case of $h = h(yN)$, $yh_{zz} + Nh_{Nz} = 0$. In such a case we always get a border solution with $z = 0$.
8. This is a basic result derived also for firms behaving monopolistically, and where there are no labour market imperfections; see Buchanan (1969) and Cropper and Oates (1992).
9. It can, however, be argued that vh should be used by the government to clean up damages from pollution and/or compensate damaged parties. Then a net revenue increase of vh would be required for the government to stay in balance, and firms would always be 'subsidized' according to (23).
10. Note that under such independent and firm-specific fluctuations and with a continuum of firms, the entire economy will be stationary. In Strand (1991, 1992a) I have, however, shown that the employment variations of a firm whose productivity fluctuates in such a way, adequately represents the employment fluctuations at the macro level in the case where productivity shocks are common to firms, as long as such fluctuations are 'not too great'. The model can thus also give an approximate indication of how employment would fluctuate at the macro level, given correlated shocks across firms.
11. These assumptions conform to those made in Strand (1991) for the case of no environmental effort.
12. For discussions of these assumptions and implications of alternatives, see Strand (1991).
13. For details of such solutions with $z \equiv 0$, see Strand (1991, 1992a).
14. Here and in the discussions below we assume that the policies by government do not lead to a change in the solution, from one where $N(H) = N(L)$ to one where $N(H) > N(L)$. Such a change is more likely to be implemented by the government, the closer p_h is to the level that yields equality in (42) in the absence of taxation. Given such a switch, the analysis for the case of $N(H) > N(L)$ applies here.
15. In fact the effect through $z(H)$ is also positive, since $\partial G(H)/\partial z(H)$ and $\partial z(H)/\partial t_L$ generally will have opposite signs.
16. This is because $\partial G(H)/\partial z(H)$ and $\partial z(H)/\partial t_L$ now are both positive, and thus the effect on $G(H)$ through $z(H)$ positive.

Part IV

Financial market imperfections

11 The stock market and equilibrium recessions

Jeff Frank

Introduction

A considerable number of recent studies have developed the multiple equilibrium approach to the understanding of Keynesian macroeconomics, as discussed in surveys by Dixon and Rankin (1991) and Silvestre (1993). Models in the literature show how imperfect competition and increasing returns lead to multiple equilibria. Low level equilibria can be viewed as Keynesian in nature, particularly insofar as outcomes can be improved by coordination or aggregate demand policies.

The purpose of this chapter is to explore the extent to which similar results can be found in an economy with price-taking firms and decreasing returns. The novelty in our analysis is the introduction of a stock market. In an overlapping-generations economy with a stock market, there are interesting multiple equilibria. A low output equilibrium, or 'equilibrium recession', has the following characteristics. Real interest rates are high, capital investment is low and share values are low. These features are consistent with stylized facts about economic recessions. This is an equilibrium since the low output and resulting low incomes lead to low savings that match the low investment.

An advantage of the current model is that it has the 'look and feel' associated with traditional macroeconomics as developed in IS-LM analysis. We develop the analysis with a flow market-clearing condition and with an asset market, portfolio-balance condition. The intersections of these curves determine the multiple equilibria.

The model

We examine a simple overlapping-generations stock market economy.[1] d'Aspremont *et al.* (1991a) and Pagano (1990) have explored multiple equilibria in imperfect competition overlapping generations frameworks without stock markets. Frank (1989) considers an overlapping-generations stock market economy with imperfect competition. In that economy, multiple equilibria hold even in the presence of decreasing returns technologies. The current model extends these results to an economy with price-taking firms.

The economy has a fixed number F of identical firms indexed by $f=1,\ldots,F$ that exist indefinitely. All firms produce the same good which can either be consumed or used as capital. Physical capital is the sole factor used in producing output. In each period t, firm f hires capital K_{ft} at the rental R_t and uses the capital as the sole factor in producing its output Q_{ft} to be sold at the price P_t.[2] There are strictly decreasing returns to the firm's scale of production: $Q_{ft}=g(K_{ft})$ where $g(\cdot)$ is strictly concave.

Each period, the firm pays out its profits $P_tg(K_{ft})-R_tK_{ft}$ as dividends to its shareowners. Shares in the firm are traded on a stock market at a price that values the firm in total at the amount V_{ft}. An individual buying γ of firm f pays γV_{ft} for the shares involved. Since there are no persistent effects to the choice of capital, firms maximize profits by choosing K_{ft} at each point in time t such that:

$$P_tg'(K_{ft})=R_t. \tag{1}$$

Households are part of an overlapping-generations structure and live for two periods; households are identical except for their age. H households are born each period and each is indexed by its number $h=1,\ldots,H$ and its date of birth t. When young, households receive an endowment in the amount e. This endowment is best viewed as labour services. The young sell this endowment to the old of the previous generation at the price of labour services W_t. The young do not consume but instead 'invest' their income W_te in physical capital and in buying shares of firms. When old, the household receives rental on capital, dividends on shares, and the proceeds from selling its capital and shares.[3] The household spends the receipts on the produced good and on buying labour services from the young. The household gains utility from its consumption of the two goods (produced goods bought from firms, and labour services bought from the young) following the quasi-concave, increasing utility function $U(\cdot)$.

The young face a portfolio choice between physical capital and shares in the various firms. Depending upon the rates of return, they allocate their endowment income W_te across the assets. In the absence of uncertainty, as in our model, we consider the perfect foresight rates of return. The rate of return on physical capital bought in period t is:

$$(P_{t+1}-P_t+R_{t+1})/P_t.$$

Since each firm is identically placed in a given period, each chooses the same capital and makes the same profits. The return on shares bought in period t is:

$$[V_{t+1}-V_t+P_{t+1}g(K_{t+1})-R_{t+1}K_{t+1}]/V_t.$$

Households will only hold physical capital and shares in firms if each pays the same return. This leads to the portfolio-balance condition for any equilibrium:

$$(P_{t+1}-P_t+R_{t+1})/P_t=[V_{t+1}-V_t+P_{t+1}g(K_{t+1})-R_{t+1}K_{t+1}]/V_t. \quad (2)$$

Further, the total value of assets held by the h households must equal their income from selling labour services. The wealth-holding condition can be written:

$$(V_t+P_tK_{t+1})F=W_teH. \quad (3)$$

The old face a choice between the consumption of the produced good and consumption of labour services. Their income consists of dividends on shares, $P_{t+1}g(K_{t+1})-R_{t+1}K_{t+1}$, rental on capital, $R_{t+1}K_{t+1}$, receipts from the sales of shares at the value V_{t+1} and the sale of capital $P_{t+1}K_{t+1}$. Write the total income at $t+1$ of household h born at time t as $Y_{h,t+1}$. The household spends this on the two goods. The demand for labour services at time $t+1$ from this household can be written:

$$L^d_{h,t+1}=L(W_{t+1}/P_{t+1},Y_{h,t+1}/P_{t+1})$$

using the homogeneity of demand to write this in terms of the real wage and real income. Walras' Law ensures that, if the labour services market and the asset markets clear, then the goods market must clear as well. The labour market clears if the total demand for labour services equals the supply from the next generation:

$$L(W_{t+1}/P_{t+1},Y_{h,t+1}/P_{t+1})=e. \quad (4)$$

Equilibria in the economy must meet conditions (1)–(4) at each period in time t. We will consider only quasi-stationary equilibria. These are defined by a constant inflation rate π and therefore prices $P_t=P_\tau(1+\pi)^{(t-\tau)}$, $V_t=V_\tau(1+\pi)^{(t-\tau)}$, $W_t=W_\tau(1+\pi)^{(t-\tau)}$ and $R_t=R_\tau(1+\pi)^{(t-\tau)}$. Further, capital and output at each firm are constant at levels K^* and $g(K^*)$. Then we have the profit maximization and asset market-clearing conditions corresponding to (1)–(3):

$$g'(K^*)=R_\tau/P_\tau \quad (5)$$

$$[g(K^*)-(R_\tau/P_\tau)K^*]/(V_\tau/P_\tau)=R_\tau/P_\tau \quad (6)$$

$$V_\tau/P_\tau+K^*=(W_\tau/P_\tau)(eH/F). \quad (7)$$

The profit maximization condition (5) is that the marginal product of capital equals the real rental. Portfolio-balance (6) requires that the real rate of return on each asset is the same: the real profit level, divided by the real price of shares, equals the real rental on capital. The wealth-holding condition (7) states that the real value of assets must equal the real income of the young, given that the young under our assumptions save all their income. Finally, we have the condition corresponding to (4):

$$L[W_\tau/P_\tau, g(K^*) + eW_\tau/P_\tau] = e. \tag{8}$$

The interesting feature of this condition is that, in a quasi-stationary equilibrium, the real income of the old is just the real value of the produced good plus the real income of the young. The reason for this is that the old receive dividends and rental income equal to the value of the produced good; they then sell their shares and capital to the young for the total income of the young.

(5)–(8) define the quasi-stationary equilibria in the economy. In this non-monetary economy, it is not surprising that the price level and the rate of inflation are indeterminate in equilibrium. Equilibria are defined by the capital stock K^*, the real rental on capital, the real wage and the real value of shares. We will analyse the set of equilibria in the following section, showing by an example how multiple equilibria can arise.

An example

For further discussion, it is helpful to have a specific example. This requires specifying the production function $g(\cdot)$ and the demand for labour function $L(\cdot)$, as well as numerical values for e, F, H. Consider the quadratic production function:

$$g(K) = aK - (b/2)K^2$$

and the labour demand function (shown for household h at time t):

$$L(W_t/P_t, Y_{h,t}/P_t) = (Y_{h,t}/P_t)/[(W_t/P_t) + (W_t/P_t)^{1/2}]. \tag{9}$$

This is the labour demand function for a household with an underlying utility function in consumption of the two goods (the produced good and labour services) $U(c_1, c_2) = c_1c_2/(c_1 + c_2)$. Also adopt numerical values: $e, H, F = 1$, $a = 10$, $b = 2$.

We proceed by solving (6)–(8), with the specific functional forms, for a relationship between the rental on capital (R_τ/P_τ) and capital utilization K^*. As discussed earlier, the real income of a household in equilibrium meets $Y_{h,t}/P_t = g(K^*) + eW_\tau/P_\tau$. Then from (8) and (9) we can solve for the equilibrium real wage:

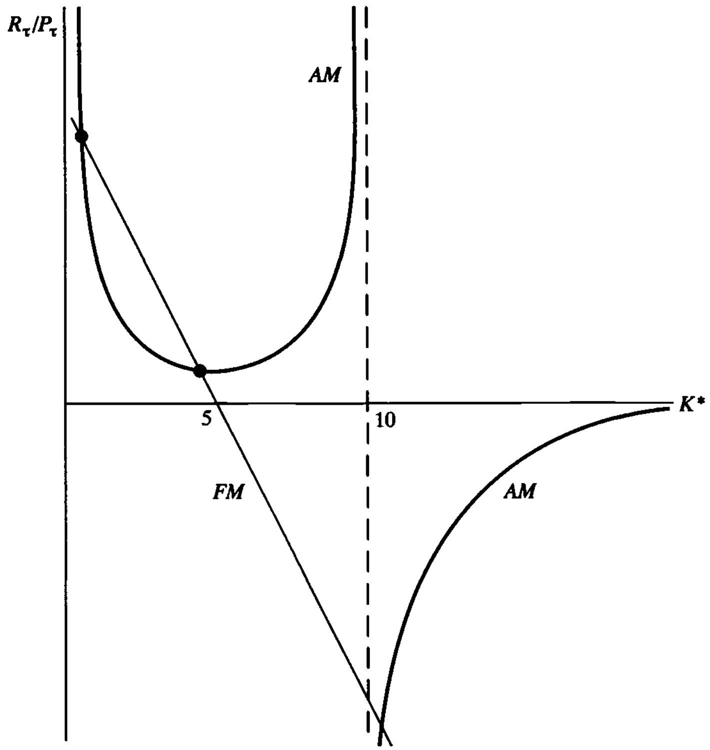


Figure 11.1 Multiple equilibria in the asset market and the flow market

$$W_\tau/P_\tau = [g(K^*)/e]^2. \tag{10}$$

Substituting (10) into (7), and then into (6), with the numerical values $e, H, F = 1$, leads to:

$$R_\tau/P_\tau = 1/[g(K^*)]. \tag{11}$$

This condition represents asset market-clearing in our model. For each level of capital and production, there is an interest rate that leads to a value of shares equalling the labour services income of the young, net of their cost of acquiring the physical capital. Incorporated within this relationship is the determination of the real wage so that the demand for produced goods and labour services equals their supplies, and the portfolio-balance condition that the return on physical capital and shares must be the same.

Under the quadratic production function, (11) takes on the form shown in figure 11.1 for the AM (asset market-clearing) curve. The remaining consideration in determining equilibria in the model is the profit maximiza-

Table 11.1. *Equilibria with positive R*

Equilibrium	K	Q	R_τ/P_τ	V_τ/P_τ
High	4.98	25	0.04	620
Low	0.01	0.1	9.98	0.00002

tion condition (5). We view this as a flow market condition in that the FM curve in figure 11.1 shows how the interest rate induces the firm to choose different levels of capital and production. As seen in figure 11.1, there are three intersections of the AM and FM curves, but only two of these with positive returns to asset-holding.

For numerical calculations in our quadratic production function example, the AM curve takes on the form from (11):

$$R_\tau/P_\tau = 1/(10K^* - K^{*2}) \tag{12}$$

and the FM curve from (5):

$$R_\tau/P_\tau = 10 - 2K^*. \tag{13}$$

Together, these lead to a cubic equation in K^* determining the three roots discussed above. Numerical calculations of the variables of interest for the two equilibria with positive R_τ appear in table 11.1.

These equilibria meet reasonable stylized facts about macroeconomic fluctuations. The equilibrium recession is characterized by low output, high interest rates and low share values.

The intuition behind our results is straightforward. A low interest rate induces the firm to rent more capital and produce more output. But the higher output also leads to greater savings, which allows for higher capital (and share values) to be consistent with a lower interest rate. The particular mechanism by which higher output leads to higher savings is peculiar to the overlapping-generations structure. In that framework, the young save from their wage income to provide pensions for their old age. Hence increased labour income raises the savings ratio, in contrast to the traditional stylized fact that this type of redistribution lowers savings. In our example, higher output of the produced good leads to a sufficiently higher demand for labour services that the savings ratio rises.

Given that we have examined an overlapping-generations economy with a number of stylized features, it may not be clear to what extent our results depend upon the existence of a stock market. One way to investigate this is to consider what happens in the example as the parameter b in the production function approaches zero.[4] When $b = 0$, there are constant returns to scale in production and zero profits; as a result, shares in firms have no value and there is effectively no stock market. To examine the behaviour of the economy as $b \Rightarrow 0$ adopt the general quadratic production function $g(K) = aK - (b/2)K^2$, the labour demand function (9) and parameters $e, H, F = 1$. Then as $b \Rightarrow 0$ the economy approaches the unique equilibrium defined by the equality of (5) and (11) with $K^* = 1/a^2$. This highlights the role of the stock market in creating multiple equilibria.

Conclusions

Our analysis has developed multiple equilibria with price-taking firms and decreasing returns to scale. We feel that this represents a significant strengthening of the case for the coordination failure approach to understanding macroeconomic fluctuations, since it generalizes the technologies and market structures consistent with multiple equilibrium models.

The equilibria in the model meet sensible macroeconomic stylized facts in that the low output equilibrium is also associated with high real interest rates and low share values. Another important feature of the framework is that it displays a macroeconomic 'look and feel'. This is partially in that the model emphasizes the role of financial markets in understanding macroeconomic fluctuations. More directly, the approach leads to asset and flow market-clearing conditions that mimic those of traditional IS-LM analysis.

Natural questions arise about the optimality and stability associated with the multiple equilibria. Scheinkman (1978) and Dechert and Yamamoto (1992) establish that equilibria in overlapping-generations stock market economies are efficient. This is an important difference between the multiple equilibria in our model and those of the imperfect competition literature. A second issue relating to policy implications involves the stability of equilibria. With only two equilibria, one will typically be unstable. These issues are discussed more fully in the related analysis in Frank (1994). In that model, it is seen that a shift to a high output equilibrium benefits all future generations at the expense of the current generation. The initial generation must increase its savings to allow for a higher rate of capital accumulation, even though its income has not risen. The model also leads to an odd number of equilibria, so that there is both a stable low and a stable high output equilibrium.

Notes

I am greatly indebted to Heraklis Polemarchakis and numerous participants at the University of Warwick Macroeconomics Workshop (July 1993). This chapter is produced as part of a Centre for Economic Policy Research (CEPR) programme on *The UK Labour Market: Microeconomic Imperfections and Institutional Features*, supported by the UK Department of Employment (Research Grant No. 4RP-154-90). The views expressed in this chapter are not necessarily those of the CEPR or of the Department of Employment.

1. Scheinkman (1978) and Dechert and Yamamoto (1992) develop the general properties of overlapping-generations stock market economies.
2. The results would be unaffected if the firms owned the capital.
3. Note that the rental on capital, received by a generation born in period t, is R_{t+1} on their capital K_{t+1}. This is an accounting convention, and the alternative assumption, that they received R_t, would make no difference to our analysis.
4. This point was suggested by Neil Rankin.

12 Asymmetric information, investment finance and real business cycles

Brian Hillier and Tim Worrall

It is not money that makes the world go round, but credit.
Stiglitz (1988, p.320)

Introduction

This chapter surveys the literature on the role of financial factors in explaining economic fluctuations. We place special emphasis upon the recent literature on the implications for economic fluctuations of asymmetric information in the market for investment finance.[1] The basic argument of this literature is that, in the presence of informational asymmetries and agency costs, financial factors may affect real variables like investment and output. In dynamic models these real variables may also affect financial factors and may generate persistent effects of shocks even in models which would not display persistence in the absence of the informational asymmetry.

The plan of the chapter is as follows. The second section provides a brief discussion of the views of some earlier writers on the importance of financial factors in the determination of economic activity. The third section reviews the microeconomic arguments concerning informational asymmetries and their implications for investment finance. The fourth section shows one way in which these microfoundations have been used to provide a real business cycle model based on informational asymmetries and agency costs. We show that this model may yield multiple equilibria with the possibility that the economy tends to oscillate around either a high output or a low output equilibrium. The fifth section reviews the literature and the sixth section concludes.

Background

The idea that the role of the financial system is important in explaining the cyclical behaviour of the economy has a long history. Fisher (1933) in his theory of 'debt-deflation' coupled the collapse of the financial system with the collapse of real economic activity in the Great Depression. According to Fisher the high level of borrowers' debt built up during the period of prosperity preceding 1929 made the economy vulnerable to the ensuing downturn which led to a wave of bankruptcies which, in turn, enhanced the downturn. Furthermore, deflation reduced the net worth of borrowers and led them to cut back on their expenditures which, without any offsetting rise in the expenditures of creditors, served to exacerbate the recession. A similar idea is suggested by Keynes who wrote that 'if the fall of wages and prices goes far, the embarrassment of those entrepreneurs who are heavily indebted may soon reach the point of insolvency – with severely adverse effects on investment' (1936, p.264). Keynes used this idea as one of his arguments to explain why the market economy is not returned to full employment equilibrium by falling money wages and prices.[2] In much of the *General Theory*, however, Keynes takes the failure of the price mechanism as given and so works out the equilibrium of the economy for arbitrary values of money wages and prices. This latter exercise, although in many ways less interesting than his arguments about the inefficacy of price and wage flexibility for stabilizing the economy, was much easier to formalize and it provided the foundation stone for much of Keynesian economics and the misguided view that Keynes relied upon the liquidity trap or wage rigidities to explain involuntary unemployment. The subsequent debate between the Keynesians, monetarists and new classical economists also tended to blur or ignore the fact that Keynes felt that the major source of economic fluctuations was to be found in highly volatile demand for investment and attention instead focused upon the monetary sector and developments of the theory of liquidity preference.[3] Two important reasons for the relative neglect of the investment sector were the forceful advocacy by Milton Friedman of the importance of the money supply, and the Modigliani–Miller theorem, which formally showed that in a perfect markets setting real investment decisions and the value of the firm did not depend upon the method of finance.

Despite being out of the limelight, interest in the role of investment and in its mode of finance persisted in attracting the attention of macroeconomists. Notable among these were Gurley and Shaw (see, for example, 1955), who noted the role of financial intermediation in the credit supply process and called attention to the importance of 'financial capacity'. Financial capacity was an aggregate indicator of borrowers' ability to support debt without having to cut back current or future spending in order to avoid default or rescheduling. The role of financial and balance sheet variables on investment and output was thus emphasized in a manner similar to Fisher and Keynes.

Another important writer was Rosa who put forward the so-called Availability Doctrine that

> in essence, it is not necessarily interest rates as a cost to the borrower, nor as an inducement to the saver, but rather interest rates as a reflection of underlying changes in credit availability, that have an important (though certainly not always a decisive) impact upon the generation of business cycles. (1951, p.276)

Rosa's arguments focused attention on the issue of credit rationing and stimulated attempts to provide a sound theoretical explanation for the failure of interest rates to rise to equate supply and demand in the market for loanable funds. These attempts in time led to the development of a substantial literature on the role of asymmetric information in the credit market. This literature provides the microeconomic foundations for the models of the business cycle which we review in this chapter. The following section looks at these microfoundations.

Asymmetric information and investment finance

Developments in the economics of information and incentives have been applied to both the equity and debt markets for investment finance and have been used to explain the forms of financial contracts and intermediation.[4]

Consider the market for debt. There are three types of informational asymmetry dealt with in the literature, either singly or in combination:

(a) Borrowers may be indistinguishable *ex ante*. This may give rise to adverse selection and Akerlof's (1970) 'lemons' problem.

(b) Banks may be unable to observe the use to which borrowers put their funds. This may give rise to the problem of moral hazard with hidden actions as in Arrow (1963, 1968).

(c) Banks may be unable to observe the returns to a project without incurring a cost as in Townsend's (1979) model of costly state verification. This may give rise to the problem of moral hazard with hidden information since the borrower has an incentive to declare a project return so low as to make him unable to repay his debt to the bank even if the return is in fact much greater than would be needed to do so. In response to this problem, banks commit themselves to monitoring, either for sure or with some given probability, borrowers who default.[5]

Any of the above asymmetries of information may yield the result that an increase, beyond a certain level, of the interest rate on loans may adversely affect the rate of return to banks. In the first case this happens by driving borrowers with safer projects out of the market which, given the asymmetry in payoffs induced by the standard debt contract, is undesired by even risk-neutral banks. In the second case it happens by driving borrowers to choose riskier projects, and in the third it happens by causing a rise in bankruptcies and an increase in monitoring costs. In each case the non-monotonic relationship between the rate *charged* by banks and the return *received* by banks may be used to explain the phenomenon of credit rationing: banks may wish to hold the interest rate below the market clearing level since raising the rate would lower bank returns.

The early literature on asymmetric information and credit rationing placed the emphasis on adverse selection and assumed that borrowers issue standard debt contracts that pay lenders a fixed yield if the project return is sufficiently high, or pay the entire project return if this is below the required fixed yield; see, for example, Jaffee and Russell (1976), Keeton (1979) and Stiglitz and Weiss (1981). However, a problem with this literature is that the results are sensitive to the nature of the distributions from which project returns are drawn; see, for example, de Meza and Webb (1987) who replace the assumption used by Stiglitz and Weiss (1981) that the distributions differ across projects in variances but not means by the assumption that project distributions differ in expected returns. Another problem is that the results are sensitive to the nature of the financial contracts; see, for example, Bester (1985) who showed how the introduction of collateral requirements might be used by banks to induce borrowers to self-select themselves into different categories and eliminate rationing, or de Meza and Webb (1987) who showed that the optimal mode of finance given the Stiglitz and Weiss (1981) framework was equity and not debt.

Even if equity is the optimal mode of finance, however, it is possible to show that the agency problems which beset the debt market have their counterparts in the market for equity; see, for example, Myers and Majluf (1984), Stiglitz and Weiss (1981), Greenwald, Stiglitz and Weiss (1984) and Stiglitz (1988). Problems of a moral hazard or incentive type occur because when a firm is equity-financed managers receive only a small fraction of any extra profit so their incentive to expend effort on making profits is attenuated. Alternatively, since the owners or managers of firms have private information about their firms' expected returns, it may be those with the lowest expected returns who are most willing to sell their shares, thus leading to adverse selection problems.

These informational asymmetries in either the credit or equity market show that investment may be constrained. The next section examines the dynamic macroeconomic implications of these constraints and attempts to find some microeconomic foundations for the ideas of the earlier writers which were presented on pp.245–7. Given the variety of informational asymmetries there are modelling choices to be made and we follow the route of Bernanke and Gertler (1989). They introduce a costly state verification problem where debt and retained earnings or net worth play important roles. There are good reasons for adopting this approach.

First it is possible to show that in models with costly state verification the optimal form of financial contract is a standard debt contract and that this contract is best intermediated by banks; see, for example, Diamond (1984), Gale and Hellwig (1985), Boyd and Prescott (1986) and Williamson (1986). The intuition for this result is simple. Given the informational asymmetry the non-default payoff is a constant because no borrower would ever choose to pay to the lender more than the minimum amount necessary to prevent monitoring. The default payoff will equal the return to the project, since if it was less then it would be possible to raise it whilst lowering the non-default payoff so that the borrower's expected repayment remains the same; this would leave the borrower no worse off but yield a gain to the lender by reducing expected monitoring costs. Lenders monitor whenever entrepreneurs claim to be unable to repay their loan since if they did not do so entrepreneurs would have an incentive to default and keep returns to themselves even when projects were successful. Intermediation dominates direct lending since banks economize on monitoring costs; a bank monitors a defaulting loan only once compared with each lender needing to monitor individually under direct lending.

Secondly, debt and retained earnings are empirically by far the major sources of investment funds, especially for small and medium sized firms (see Fazzari *et al.*, 1988, Stiglitz, 1992 for some convincing evidence). According to Stiglitz this is partially a result of agency problems in the equity market. Thus he claims that

> the cost of issuing equity is sufficiently great that most firms act as if they were equity rationed. When they are denied credit, they do not raise capital by issuing new equity, but rather constrain their capital expenditures to retained earnings. (Stiglitz, 1988, p.313)

A real business cycle (RBC) model

Overlapping-generations model

In this section we present a slightly modified version of Bernanke and Gertler (1989) who show how monitoring costs can produce low investment equilibria and real business cycles. Their model introduces intragenerational heterogeneity and an asymmetry of information into the overlapping-generations model of Samuelson (1958). There are infinitely many periods but in any one period there are two equally sized cohorts, one in its youth and the other in its old age. Each cohort, distinguished by its date of birth, consists of a continuum of risk-neutral agents called entrepreneurs who live two periods, their youth and old age, but consume only in their old age.

There are two produced goods, an output or consumption good and a capital good. The capital good is produced using the output good as an input and the output good is produced by a constant returns to scale technology using the capital good and labour as inputs. Capital goods depreciate completely after one period in use. There is also a constant returns to scale storage technology to which everyone has access and which transforms one unit of the output good at the start of the period into $r \geq 1$ units of the output good at the end of the period. We call r the gross rate of interest.

In their youth entrepreneurs are endowed with a single unit of labour which they supply inelastically to a competitive labour market which pays a wage w (i.e. they receive w units of the output good) equal to the marginal product of labour at the end of the period. They save this w units of output which then becomes their initial wealth or savings at the start of their old age. Entrepreneurs must then decide whether to become capital goods producers. It is at this stage that heterogeneity is introduced: entrepreneurs have access to different capital good production technologies. In particular each entrepreneur has access to an investment project which yields Z units of capital goods, where Z is a random variable, but requires a fixed input of x units of the output good. The random variable Z is the same for each entrepreneur independent of x, but x varies across entrepreneurs. Entrepreneurs with low values of x are thus more likely to undertake investment projects.

There is also a simple asymmetry of information: when an entrepreneur undertakes an investment project only he can observe costlessly the actual number of units z of capital goods produced. Any other agent must pay a monitoring cost to observe the number of units of the capital good produced. It is assumed that monitoring of any project uses up m units of the capital good independent of z or x. We thus have a costly state verification model of the Townsend (1979) type. As we have seen on p.249 financial intermediaries or banks arise naturally in such an environment (see Diamond, 1984, Williamson, 1986) in order to economize on monitoring costs. These banks take in funds from entrepreneurs who decide not to invest or have excess savings and they lend to others who wish to invest but have insufficient funds. It is assumed that the banking sector is perfectly competitive so that each bank will make zero profits in equilibrium and that each bank has a well diversified portfolio of loans. In addition it is assumed that there are always sufficient funds in the economy to finance any level of investment which the banks wish to fund, so the gross rate of interest paid on deposits at banks is r.

Capital goods produced but not used up in monitoring will be supplied to a perfectly competitive capital market at a relative price q in terms of the output good. Let k be the aggregate[6] quantity of capital supplied. In aggregate one unit of labour is supplied. Hence aggregate output is given by $f(k)$, where $f(\;)$ is the aggregate production function which is increasing and concave in k. As there are constant returns to scale in output good production, the capital price equals the marginal revenue product of capital, $f'(k)$ and wages are equal to the marginal revenue product of labour, $f(k)-f'(k)k$. There are no profits in the output good sector.

The economy then proceeds as follows: at any date t the current old have savings of s_t (the same for every member of the cohort). Old entrepreneurs invest or save and this determines the amount of capital goods produced, some of which may be used up in monitoring defaulting loans.[7] The net amount of capital goods and the fixed amount of labour supplied by the young are then used as inputs to produce the output good. Because the marginal product of labour increases with capital, a larger net capital stock will produce higher wages and hence higher savings for next period. The key role in Bernanke and Gertler (1989) is played by monitoring costs which provide a link between the entrepreneur's wealth and capital goods production. No such link is present under perfect information.

First best case

As a benchmark we shall consider the first best case where there is perfect information, i.e. no private information, so that the outcome of the investment process is observable to any agent at no cost, $m=0$. Assume that the random variable Z has a continuous, differentiable probability distribution function $G(z)=\text{prob}\{Z\leq z\}$ with support $[z_{\min}, z_{\max}]$ and density function $g(z)$.[8] Let $z^e=\int z dG$ denote the expected value of Z. Then an individual entrepreneur, taking q as given and borrowing if necessary at the rate r, will undertake his investment project if $z^e \geq rx/q$, i.e. if the expected return exceeds the opportunity cost of investment in terms of capital goods. Write $x(q)=qz^e/r$, since z^e and r are given parameters of the model but q is an endogenous variable. We shall assume that agents have perfect foresight so the capital goods price q they expect is the equilibrium price. An entrepreneur with $x \leq x(q)$ will undertake his investment project but an entrepreneur with $x > x(q)$ will put any savings in the storage technology or on deposit at a bank. Let $H(x(q))$ denote the proportion of entrepreneurs with $x \leq x(q)$ where we assume $x \in [0, x_{\max}]$. We can treat $H(x)$ as a distribution function which we assume has a continuous density function $h(x)$ with $h(x)>0$ on $(0, x_{\max})$. Hence aggregate investment in the economy is

$$i(q)=\int_{0}^{x(q)} x\,dH. \tag{1}$$

We refer to $i(q)$ as the perfect information investment schedule. Although there is uncertainty about the production of capital goods at an individual level, since there is a continuum of entrepreneurs the law of large numbers can be invoked so that at an aggregate level capital goods production is non-stochastic and given by

$$k(q)=z^{e}H(x(q)) \tag{2}$$

where $H(x(q))$ is the proportion of entrepreneurs who undertake their investment projects and z^e is the average capital goods production of every project. We refer to $k(q)$ as the perfect information supply of capital goods schedule. The demand for capital goods schedule $k^d(q)=f'^{-1}(q)$ is given by equating the capital goods price to the marginal product of capital. As $f''(k)<0$ and $H(x)$ is increasing there is a unique equilibrium capital goods price satisfying $z^eH(x(q))=f'^{-1}(q)$ which we shall denote q^*, with $x^*=x(q^*)$, $i^*=i(q^*)$ and $k^*=k(q^*)$ denoting the first best levels. In order to maintain the assumption that r is fixed it is necessary to assume that there is always an excess supply of savings, i.e. $s_t \geq i^*$ for each time period t. The dynamics in the perfect information case are trivial. In period 0 with savings s_0, i^* is invested producing a capital stock k^* which commands a price q^*. The wages of the young are $w^*=f(k^*)-f'(k^*)k^*$, the marginal product of labour. This becomes the next period's wealth or savings, $s_1=w^*$ so that i^* is again invested, the capital stock is k^* and output is $f(k^*)$, i.e. the equilibrium level of savings $s_t=w^*$ is attained after one period.

Optimum financial contract

Now consider the situation with the asymmetry of information where the outcome z is private information of the entrepreneur but can be observed by other agents at a monitoring cost of $m>0$. First consider an entrepreneur entering his old age with savings s, who requires x units of the output good to undertake his investment project and who expects a capital goods price of q. Note that since the entrepreneur is small relative to the market, q will be independent of his realized production level z. As in the perfect information case, we maintain the perfect foresight assumption that the q which each entrepreneur expects turns out to be the equilibrium q.[9]

To determine whether an entrepreneur will undertake his investment project it is necessary to know on what terms he can borrow to finance his project. Suppose that the entrepreneur undertakes his project and let $e \leq \min(x, s)$ be his equity stake, i.e. how much he himself invests in the project. If $e < x$ then he must borrow the difference $b = x - e$ from a bank.[10] We assume that the loan contract with the bank can specify the equity participation e and that the creditor can costlessly access the entrepreneur's savings $r(s-e)$, whether invested at another bank or invested in the storage technology, in the event of default on the loan. We can then appeal to the results of Gale and Hellwig (1985) that any loan contract with equity participation e is weakly dominated by a loan contract with maximum equity participation $e = s$ and that for the reasons described on p.249 the optimum loan contract is a standard debt contract. A standard debt contract has the following features: (1) there is a fixed repayment R, (2) the debtor repays R if he is able, i.e. if $z \geq R$ and (3) if he is unable to repay R then the debtor is in default and the creditor monitors at a cost m but recovers the maximum amount of capital goods, i.e. z. Thus the optimum loan contract is entirely described by the repayment factor R.

Given the standard debt contract and with R specified, the borrower's expected repayment in terms of capital goods is

$$p(R) = \int_{z_{\min}}^{R} z \, dG + \int_{R}^{z_{\max}} R \, dG.$$

Integrating both terms (the first term by parts) and rearranging gives

$$p(R) = R - \int_{z_{\min}}^{R} G(z) \, dz. \tag{3}$$

Likewise the creditor's expected return in terms of capital goods is

$$\rho(R) = \int_{z_{\min}}^{R} (z - m) \, dG + \int_{R}^{z_{\max}} R \, dG = p(R) - mG(R). \tag{4}$$

There are two points to notice here. First that $\rho(R)$ is independent of x since the returns to investment are independent of x and second that the creditor's expected return is equal to the borrower's expected repayment less the expected monitoring cost, the latter being m times the probability of default, $G(R)$. The function $\rho(R)$ is drawn in figure 12.1a. We shall assume

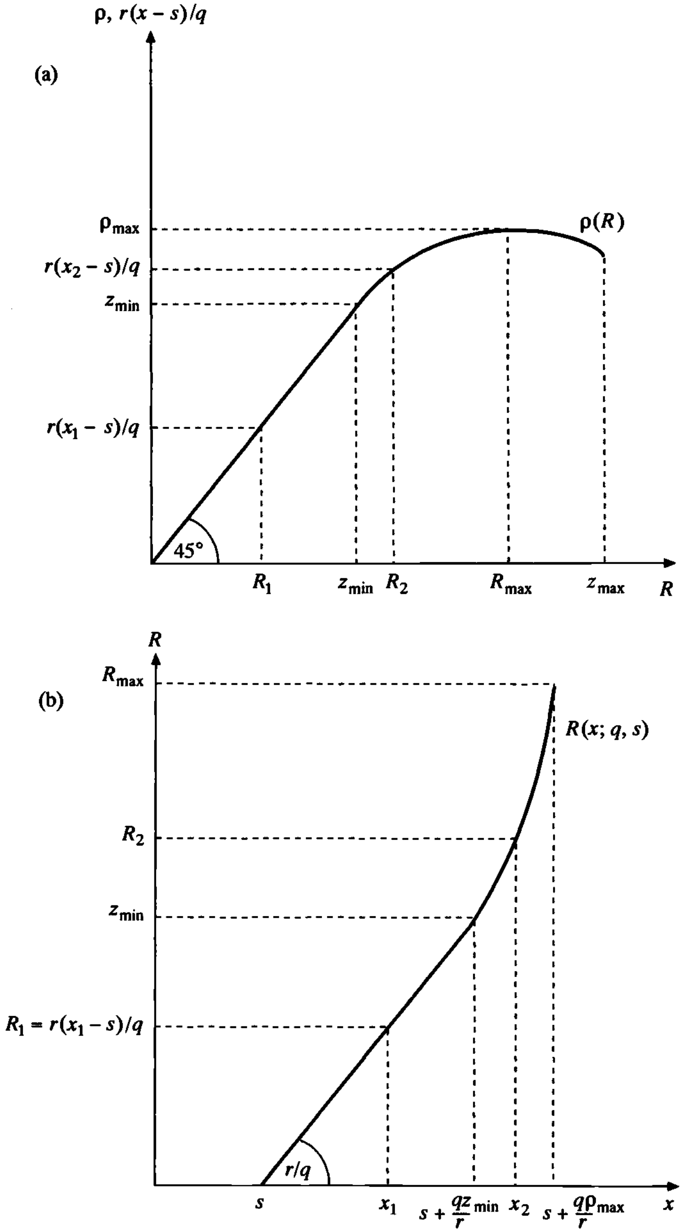


Figure 12.1 Returns and repayments
(a) Lender's return function (b) Repayment function

that it is a concave function, i.e. that $mg'(R)+g(R)\geq 0$, and let $R_{max}\in[0,z_{max}]$ denote the repayment factor which maximizes $\rho(R)$ and let $\rho_{max}=\rho(R_{max})$ denote the maximum value of the function. Intuitively, as R increases the creditor earns larger returns if the borrower does not default, but expected monitoring costs increase because there is a higher probability of default. Beyond R_{max} the latter effect begins to dominate and the creditor's return actually falls as R increases. Competition between banks means that R is never set above R_{max}, otherwise another bank could undercut and earn a higher return.

In a competitive equilibrium with free entry of banks and, hence, zero profits in the banking sector, the creditor's expected return on a loan of size b must equal the rate of return on a loan of size b available from the storage technology in terms of capital goods, i.e. $r(x-s)/q$. Hence

$$\rho(R)=p(R)-mG(R)=r(x-s)/q. \qquad (5)$$

(5) determines R as a function of x for a given s and q, i.e. the terms on which different entrepreneurs can borrow, which we denote $R(x;q,s)$ (entrepreneurs differ only in x; each has the same s and expects the same q). Figure 12.1b illustrates $R(x;q,s)$ for a given q and s. For $x\leq s$ there is no need for the entrepreneur to borrow as he has sufficient funds himself, so $R=0$. For $s<x\leq s+qz_{min}/r$ then $R=r(x-s)/q\leq z_{min}$ so there is no probability of default, because the borrower can repay in full even in the worst possible outcome. The creditor therefore faces no risk at the individual level. For $s+qz_{min}/r<x\leq s+q\rho_{max}/r$ the creditor does bear some default risk, $G(R)>0$, and $R>z_{min}$ satisfies (5). For $x>s+q\rho_{max}/r$ there is no feasible repayment such that any lender does not prefer to place funds in the storage technology.

Within-period equilibrium

Having worked out the terms on which an entrepreneur can borrow it is now possible to determine whether he would wish to borrow and therefore the within-period equilibrium investment, capital goods production and prices. An individual entrepreneur will wish to invest if $z^e-p(R)\geq rs/q$, i.e. if the expected return less the expected payment is no less than the opportunity cost of his funds. But from (5), $p(R)=(r(x-s)/q)+mG(R)$, where $R=R(x;q,s)$. That is, the borrower's expected repayment must cover the opportunity cost of the loan in terms of capital goods available from the storage technology, plus the expected monitoring cost. Hence the entrepreneur will wish to invest if

$$z^e-mG(R(x;q,s))\geq rx/q. \qquad (6)$$

We graph both sides of (6) against x in figure 12.2 for given values of q and s and let $x(q,s)$ denote the critical value of x beyond which investment is not undertaken. For $x \leq s + qz_{min}/r$, $G(R(x;q,s)) = 0$ so the left-hand side is constant at z^e, but for $x > s + qz_{min}/r$ it declines until $z^e - mG(R_{max})$ at $x = s + q\rho_{max}/r$, where we have drawn figure 12.2 assuming $z^e - mG(R_{max}) > 0$. Figure 12.2a shows the case where savings, s are large enough or q small enough, i.e. $s + qz_{min}/r \geq x(q) = qz^e/r$ or $s \geq q(z^e - z_{min})/r$, so that there is no need to monitor.[11] In this case there is no risk of default for any borrower and the asymmetry of information makes no difference. Projects are undertaken if $z^e \geq rx/q$, just as in the perfect information case.

Figures 12.2b and 12.2c illustrate cases which involve monitoring. Figure 12.2b shows the case where $x(q,s) < s + q\rho_{max}/r$. The marginal entrepreneur, for whom $x = x(q,s)$, can obtain funds from a bank by offering the bank an expected return of $r(x-s)/q$ but will make an expected profit of zero from his project. The extra-marginal entrepreneur could also obtain funds, but he would make negative expected profits and so would not undertake his investment opportunity at the terms available in the market. In Figure 12.2c, on the other hand, the marginal entrepreneur, for whom $x = x(q,s) = s + q\rho_{max}/r$, will expect strictly positive profits but the extra-marginal entrepreneur is unable to obtain funds in the market because no lender will wish to lend to him. Since the extra-marginal entrepreneur would accept the loan contract of the marginal entrepreneur if offered to him, there is credit rationing.[12] The important point for what follows, though, is not that credit may be rationed but rather whether or not creditors expect to monitor, for if they do this will raise the cost of finance and reduce investment.

In the private information case the marginal investor has an investment cost of $x(q,s)$, which depends both on the expected price of capital goods and, because of its effect on the probability of default, the savings level of the entrepreneur. In the perfect information case the marginal investor has an investment cost $x(q) = qz^e/r$ independent of entrepreneurial savings. It follows from what was said above that $x(q,s) \leq x(q)$, with equality if $s \geq q(z^e - z_{min})/r$ (figure 12.2a) but with strict inequality for $s < q(z^e - z_{min})/r$ (Figures 12.2b and 12.2c). For entrepreneurs with an investment cost of x such that $s + qz_{min}/r < x \leq x(q,s)$, the existence of monitoring costs does not change their investment decision but means that they must pay a higher cost to obtain finance. For entrepreneurs with $x(q,s) < x \leq x(q)$, monitoring costs mean that they do not invest where in their absence they would have done. Aggregate investment in the economy is then simply given by

$$i(q,s) = \int_0^{x(q,s)} x \, dH \leq i(q) \tag{7}$$

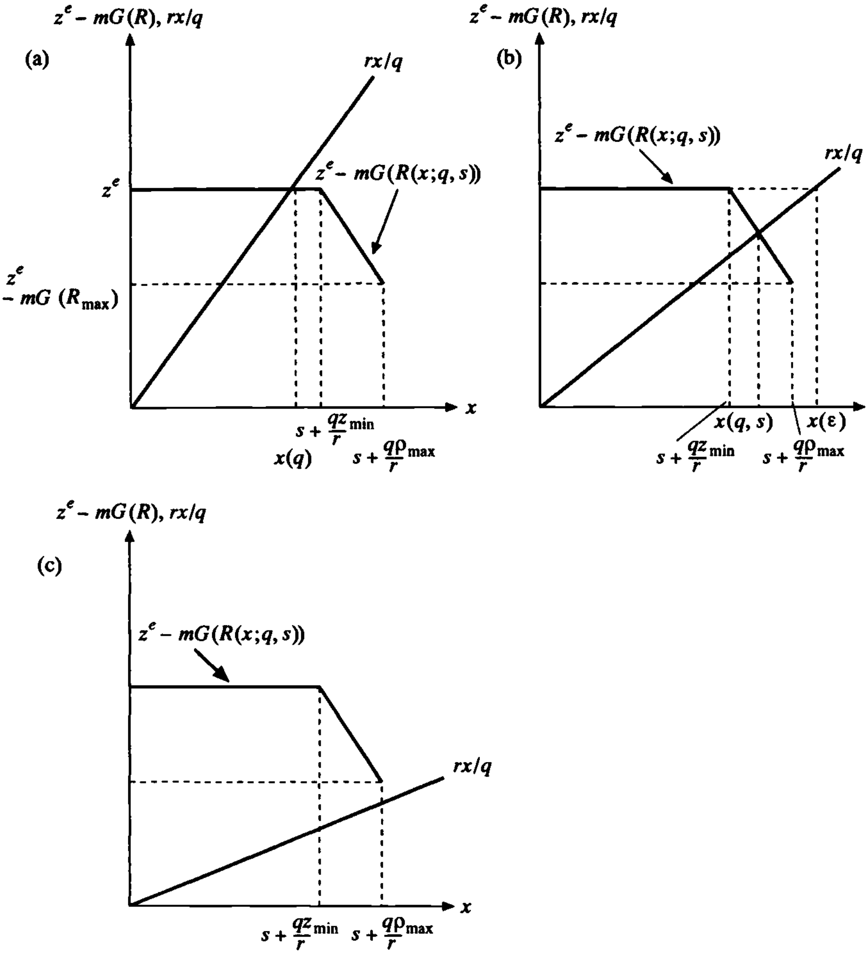


Figure 12.2 Within-period equilibrium
(a) No monitoring (b) Monitoring (c) Monitoring and credit rationing

with strict inequality for $s<q(z^e-z_{\text{min}})/r$. This means that for a given q investment is either at or below the first best level.[13]

Figure 12.3 shows the response of $x(q, s)$ to changes in q and s. It can be seen from figure 12.3, or implicitly differentiating (6), that x is increasing in q and is non-decreasing (increasing for $s<q(z^e-z_{\text{min}})/r$) in s. An increase in q has two effects (1) a direct effect on the value of the entrepreneur's return and (2) an indirect effect through the reduction of monitoring costs: a rise in q reduces the repayment factor R for a given level of savings $s<q(z^e$

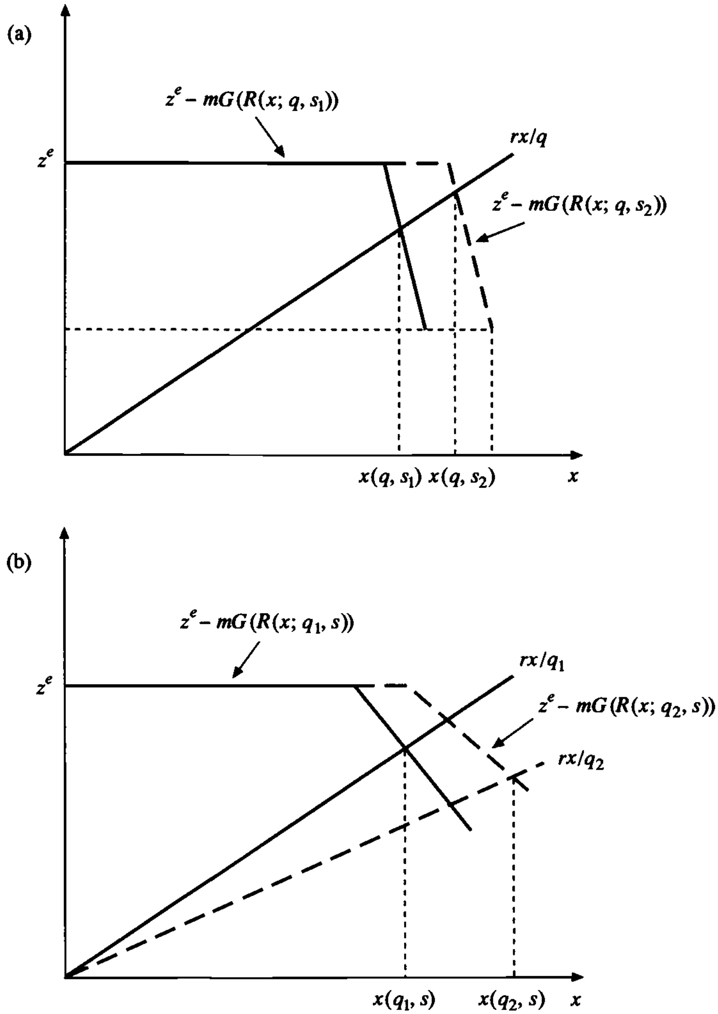


Figure 12.3 The response of x to changes in q and s
(a) The effect of an increase in s: $s_2 > s_1$
(b) The effect of an increase in q: $q_2 > q_1$

$- z_{\min})/r$ with the consequent reduction in the probability of default and hence monitoring costs. That is, an increase in q both increases the entrepreneur's gross expected return, qz^e/r, and reduces his expected payment, p (see (3)), since R is decreasing in q (see (5)). Hence marginal projects become strictly profitable as q increases. An increase in s has only the indirect effect of reducing R and thus the probability of default and expected monitoring costs; when s is larger, the loan size and repayment required for any given x is smaller.

As in the perfect information case, because there is a continuum of entrepreneurs, the law of large numbers can be invoked to show that even though capital goods production is random at the individual level, in the aggregate it is non-stochastic. Since all entrepreneurs with $x \leq x(q,s)$ undertake their investment projects, total capital goods production is $z^e H(x(q,s))$. Some of this capital goods production may, however, be dissipated in monitoring defaulting loans. Since default is a random event at the individual level the cost of monitoring an individual project will be a random variable,[14] but again by the law of large numbers aggregate monitoring costs will be non-stochastic and the quantity of capital goods supplied as an input to the output market will be

$$k(q,s) = z^e H(x(q,s)) - m \int_0^{x(q,s)} G(R(x;q,s))\, dH \leq k(q) = z^e H(x(q)) \tag{8}$$

with equality for $s \geq q(z^e - z_{\min})/r$ (where there is no risk of default) and strict inequality for $s < q(z^e - z_{\min})/r$. We refer to $k(q,s)$ as the private information supply of capital goods schedule. For $s < q(z^e - z_{\min})/r$ the capital goods supply to the market is less than in the perfect information case, $k(q)$, because (1) fewer projects will be undertaken, $x(q,s) \leq x(q)$, and (2) some proportion of the loans made to finance projects will be in default and so must be monitored, which dissipates some of the capital stock. Differentiation of (8) using (5) shows that $k(q,s)$ is increasing in q and is non-decreasing (increasing for $s < q(z^e - z_{\min})/r$) in s. An increase in q has the above mentioned direct effect of increasing the supply of capital goods but the indirect effect works in two ways; it reduces the borrower's repayment R, which encourages more entrepreneurs to undertake projects and also decreases the probability of default and so the capital lost through monitoring. Increases in s have only these later indirect effects of decreasing capital dissipated in monitoring costs and increasing the proportion of projects undertaken.

The capital market equilibrium is illustrated in figure 12.4. The demand schedule for capital, $k^d(q) = f'^{-1}(q)$, is drawn showing the equation of price and marginal product of capital. Also drawn is the perfect information supply schedule for capital, $k(q) = z^e H(x(q))$. The first best occurs at the equilibrium (q^*, k^*). Also drawn are the private information supply schedules $k(q,s)$ for three particular values of s. Higher values of s shift the schedule to the south-east. Let $s^* = q^*(z^e - z_{\min})/r$ be the smallest value of s such that the equilibrium outcome is (q^*, k^*). It can be seen from figure 12.4

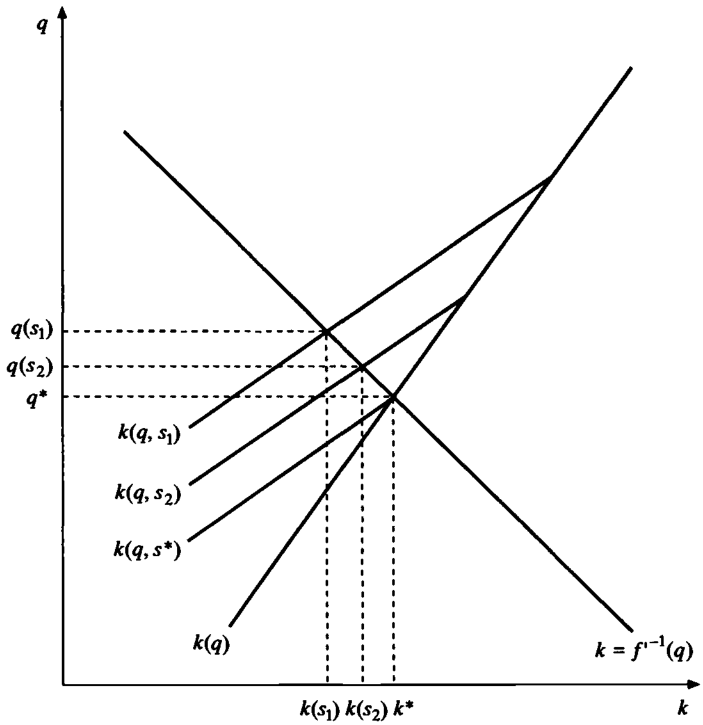


Figure 12.4 Capital market equilibrium: $s^* > s_2 > s_1$

how the equilibrium quantity of capital goods supplied to the market and the capital goods price depend upon the initial level of savings for $s < s^*$. We can write these functions $k = k(s)$ and $q = q(s)$.

From figure 12.4 it is easy to see that in equilibrium q is non-increasing in s and k is non-decreasing in s. Intuitively higher savings reduce monitoring costs so that there is a larger supply of capital for a given price q, i.e. an outward shift of the supply schedule which depresses price and increases quantity. Thus for $s \geq s^*$ the first best outcome is sustainable even with private information, but for $s < s^*$ there will be fewer capital goods used in output goods production and hence lower output. Once the capital goods price is determined for a given s, investment is determined by savings, $i(s) = i(q(s), s)$. Similarly since the wage w equals the marginal product of labour and this increases with the amount of capital used in production, wages are a non-decreasing function of savings since $k(s)$ is non-decreasing, i.e. $w(s) = f(k(s)) - f'(k(s))k(s)$ with $w'(s) \geq 0$. In principle it is necessary to check that there is an excess of savings over investment, i.e. $s_t \geq i(q(s_t), s_t)$ at every time period in order to maintain the assumption that r is fixed, but following Bernanke and Gertler we shall simply assume that parameters are such that it is always satisfied.

Deterministic dynamics

As we have said, even though capital goods production is random at the individual level it is deterministic at an aggregate level. We therefore say the economy is deterministic. The only dynamics in the deterministic economy are provided by the savings variable; what entrepreneurs earn when young becomes their savings at the start of their old age (remember that entrepreneurs do not consume until the end of their old age). But as we have seen wages in the current period depend upon savings in the current period, so savings evolve according to a simple first order non-linear difference equation given by $s_{t+1}=w(s_t)$.

Figure 12.5 illustrates the possible steady-state equilibrium savings, s_e, where the function $w(s)$ cuts the 45° line.[15] Once equilibrium savings are determined, steady-state values of investment, capital goods production, output and prices are determined from the analysis of pp.255–61. There are a number of cases to consider depending on the relative magnitudes of w^* and s^* and whether or not there is a unique equilibrium. First imagine that savings at time t are no less than s^*. Then the equilibrium capital supply is k^* with equilibrium wages $w^*=f(k^*)-f'(k^*)k^*$, so that next period's savings will be $s_{t+1}=w^*$. Hence $w(s)=w^*$ for $s\geq s^*$. Since $s^*=q^*(z^e-z_{\min})/r$ it is possible that either $w^*>s^*$ or $w^*<s^*$. In figure 12.5a $w^*>s^*$ so that savings are sufficiently high in equilibrium to eliminate any possibility of default. Hence there is a steady-state equilibrium with investment at the first best level and equilibrium savings $s_e=w^*$. In figure 12.5b $w^*<s^*$ and there is a unique low investment equilibrium. Figures 12.5c and 12.5d illustrate the possibility of multiple equilibria.[16] Both illustrate a stable high investment equilibrium and a low investment equilibrium with an unstable equilibrium in the middle. The economy may tend to either stable equilibrium depending upon initial conditions. In figure 12.5c the high investment equilibrium is below the first best level but in figure 12.5d the high investment equilibrium is equal to the first best level.

Stochastic dynamics

The dynamics in the deterministic case are monotonic: savings, prices, output and capital stock simply increase or decrease toward their equilibrium values. If, however, there is a stochastic i.i.d. technological shock to output production, then it is possible to show that monitoring costs can

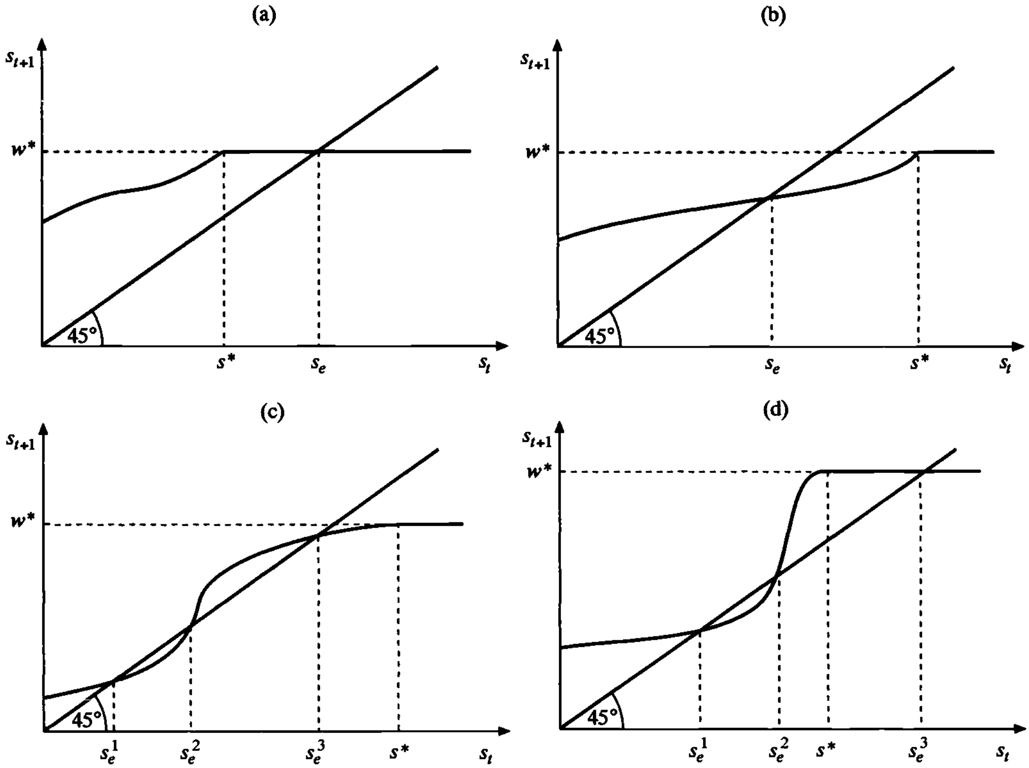


Figure 12.5 Dynamics
(a) Unique full investment equilibrium
(b) Unique underinvestment equilibrium
(c) Multiple underinvestment equilibria
(d) Multiple mixed equilibria

produce cycles or persistence and asymmetric responses. Suppose that there is a random variable θ, with expected value of unity, which acts as a multiplicative shock to the output technology, so that output $y = \theta f(k)$. We follow Bernanke and Gertler and assume that debt contracts are made before θ is known, but that labour and capital are hired after θ is known. The capital goods price will then be a random variable and entrepreneurs must base their investment decisions upon their expectations of the price. Since they are risk-neutral their investment decisions are taken as if the capital goods price were sure to be the expected price. The capital goods supply curve will therefore be exactly as in the deterministic case, i.e. $k(q^e, s)$ where q^e is the expected price. Given that the shock is unity on average, the expected demand curve is just that in the deterministic case and it can be seen from figure 12.4 that for a given level of savings, s, the equilibrium price corresponding to the expected shock is $q(s)$. Hence the rational expectations equilibrium supply of capital is $k(s)$. Once θ is known, prices will adjust to clear the market, i.e. $q(s,\theta)=\theta f'(k(s))$ since the supply of capital goods is fixed in the short run, i.e. unresponsive to θ, and wages are $w(s,\theta)=\theta(f(k(s))-f'(k(s))k(s))=\theta w(s)$.

Consider first the perfect information case in this stochastic environment. In this case $w(s,\theta)=\theta w^*$, so a good shock produces higher output and higher wages. Despite the fact that wages and thus savings are stochastic, investment next period is unaffected, since in the perfect information case investment depends only on the expected capital goods price, q^*, and hence is independent of savings. Thus i.i.d. shocks to productivity will cause i.i.d. shocks to prices and output but have no effects on investment or capital goods production.

Now consider the private information case. Since the first best equilibrium capital goods supply is k^* at an expected price of q^*, it can be seen that s^*, the smallest savings such that the first best outcome is sustainable in the private information case, is independent of θ. A positive shock, $\theta > 1$, then shifts the $w(s)$ curve upward with no effect on s^* and a negative shock, $\theta < 1$, shifts it downward. In order to examine the possible implication of this stochastic environment consider the following simple thought experiment. Suppose that there is a sequence of $\theta = 1$ shocks, so that the economy settles down at the steady-state equilibrium as if it were deterministic, but then there is one positive shock, $\theta > 1$, before again the economy experiences a sequence of $\theta = 1$ shocks. Figure 12.6a illustrates the case where there is a unique low investment equilibrium. The economy begins in the low investment equilibrium E when it is hit by a positive shock $\theta > 1$, which shifts the curve upward for one period. Wages and output are higher than anticipated which feeds through to higher savings next period. By reducing total monitoring costs this allows more investment and capital goods production next period, resulting in higher output and wages than normal even though there is no new positive shock. There is persistence. In the words of Bernanke and Gertler 'Strengthened borrower balance sheets resulting from good times expand investment demand, which in turn tends to amplify the upturn' (1989, p.27). There is also an obvious asymmetric response to shocks between an equilibrium with low investment and one with the first best level of investment. If there were a unique equilibrium which sustained the first best level of investment with $w^* > s^*$ then small shocks would have no effect upon investment next period, whereas if there is a unique low investment equilibrium even small positive or negative shocks will cause persistent changes in future investment and capital goods production. Furthermore, if there were a unique equilibrium which sus-

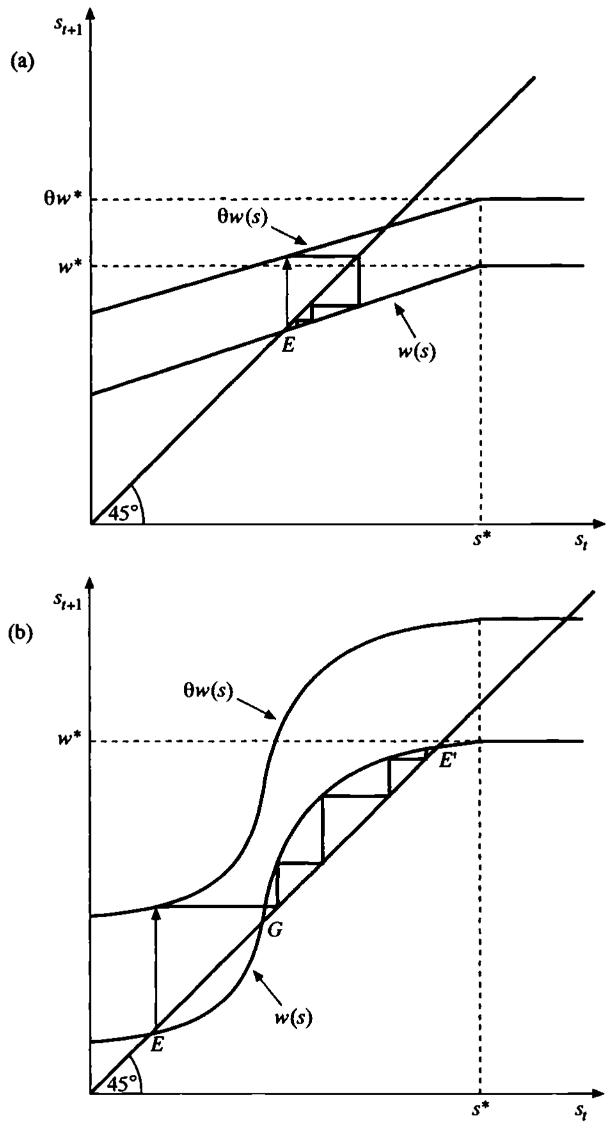


Figure 12.6 Effects of a temporary technological shock
(a) Unique equilibrium with positive shock θ
(b) Multiple equilibria with positive shock θ

tained the first best level of investment with $w^* > s^*$, then whilst large negative shocks would have an effect on investment, large positive shocks would have no such effects.

An even more interesting possibility is shown in figure 12.6b which shows multiple equilibria; a high level locally stable equilibrium at E', an unstable middle equilibrium at G, and a locally stable low equilibrium at E. It is now possible that the economy displays temporal agglomeration – once at either the high or low level equilibrium the economy tends to oscillate around it for quite a while until a big shock pushes it past the middle equilibrium at G and it then moves towards the other stable equilibrium. Suppose that the economy is at equilibrium E following a sequence of $\theta = 1$ shocks when there is a large positive shock which shifts the $w(s)$ curve upward. This generates higher output and higher savings for next period and this may push the economy beyond G and on toward the better equilibrium E' even though thereafter the economy experiences only shocks $\theta = 1$. It is also possible that there may be asymmetries in the durations around high and low level equilibria; if G is nearer to the low level stable equilibrium than the high level stable equilibrium, the economy would tend to be shocked away from the low level equilibrium more easily than it would be shocked away from the high level equilibrium and so would tend to be more often around the high level equilibrium. Evidence for such asymmetric cyclical behaviour may be found in Neftci (1984), Hamilton (1989) and Diebold and Rudebusch (1990).

Conclusions

Bernanke and Gertler (1989) have developed an equilibrium business cycle model in which financial intermediation plays a crucial role in explaining cyclical behaviour. The intermediation which takes place in the model fits well with some of the features of real world financial markets. Intermediaries carry out the task of monitoring defaulting projects, they borrow from large numbers of depositors and lend to large numbers of borrowers, writing debt contracts with borrowers and offering depositors financial assets which pay an expected return equal to the safe rate. The main implications of their model are:

(a) Absent informational asymmetries, investment is unaffected by i.i.d. shocks to productivity and output responses are non-cyclical.

(b) With informational asymmetries i.i.d. shocks to productivity produce cyclical responses in investment and output. The reason for the persistence is that a good shock increases borrower net worth, reducing monitoring costs and increasing investment and future net worth, and vice versa for a bad shock.

(c) The model may yield multiple equilibria with the possibility of temporal agglomeration or, in other words, the possibility that the economy tends to oscillate around either a high output or a low output equilibrium with a big shock needed to push it away from one equilibrium to the other.

The next section reviews a number of similar models which can be found in the literature.

Financial factors and business cycles

Apart from Bernanke and Gertler (1989) there are a number of other models of business fluctuations in which financial factors play a role. Bernanke and Gertler (1990) pursues the theme of the importance of borrower net worth, but in a model where the asymmetric information problem is based not upon costly state verification but upon moral hazard with hidden actions; entrepreneurs borrow in order to evaluate projects but lenders cannot observe whether they actually do evaluate (moral hazard), nor can they observe the information revealed by evaluation (which reveals the success probability associated with a project to the evaluating entrepreneur). Despite the differences, the critical role of net worth in easing agency problems remains essentially the same as in their 1989 study, although the latter does not emphasize its role in propagating business cycles. In an important development, Gertler (1992) returns to the problem of costly state verification but extends the analysis to allow for multiperiod financial arrangements. This makes agency costs depend upon the present discounted value of future project earnings as well as upon the borrower's current net worth. Thus a small change in macroeconomic conditions may have substantial effects upon expected future earnings and produce large short-run shifts in financial constraints.

Farmer (1984) and Williamson (1987) also embed a costly state verification problem in an overlapping-generations framework similar to Bernanke and Gertler (1989). In Williamson projects differ in terms of the monitoring costs in the case of default, rather than the cost of funding, and there are some agents who have no investment projects of their own. As in Bernanke and Gertler (1989), and for similar reasons, credit may be rationed in equilibrium; although in Williamson the allocation of credit is based upon the (observable) cost of monitoring if a project fails, with credit going to borrowers with the lower monitoring costs. Net worth, however, has no role to play since Williamson assumes that individuals with projects have zero net worth. The type of shock considered by Williamson also differs from the shocks considered on p.261. In Williamson there are shocks to the riskiness of investment projects; he assumes there are two possible states of the world, where one differs from the other by applying a mean preserving spread to the distribution of project returns. With risk-neutral agents such a shock would have no macroeconomic consequences in the absence of information problems, but with costly state verification the riskier state increases agency problems and leads to less investment in projects. Since there is a one-period lag between making an investment and the resulting production of output, this has a negative effect on investment the following year too, producing cyclical dynamics or persistence as in the model on pp.249–66. The difference is that Williamson's model formalizes the idea that business confidence is important in explaining output fluctuations.

In Williamson (1987) an important role is played by the demand for money. He assumes that the consumption good is perishable between periods so that his lender class, who supply labour when young but not when old, must either invest in projects (via intermediaries) or hold money if they wish to consume when old, thus yielding a role for money (i.e. valued, unbacked government securities). His model yields a positive correlation between the price level and real output, and negative correlations between real output and business failures and real output and risk premia. It predicts that intermediary loans and a nominal monetary aggregate Granger cause output. The results concerning the price level and nominal monetary aggregates cannot be derived from Bernanke and Gertler (1989) which neglects monetary factors, but clearly the two models are complements rather than substitutes.

Greenwood and Williamson (1989) develop a two-country version of Williamson's model which examines the role played by financial factors and exchange rate systems in the international transmission of business cycles. The model displays monetary non-neutrality and positive comovements among national outputs, inflation and interest rates. The correlation between output and prices depends upon the source of disturbances, monetary shocks yielding a positive correlation and technological shocks a negative one. Exchange rate regimes matter for the variance of output, but which yields the lower variance depends on the source of shocks, for example a flexible exchange rate regime yields a lower output variance in response to foreign monetary shocks than either of the two fixed exchange rate regimes considered.

There are also a number of other studies in the literature which include important roles for financial factors in the propagation of business cycles. One approach adopted by Kiyotaki and Moore (1993) is to provide a dynamic model of the enforcement problems in the credit market considered by Hart and Moore (1989).[17] Although the model is rather different from the models of asymmetric information discussed so far, the results are similar. As borrowers may repudiate their debt, creditors protect themselves by securitizing loans and never allowing the size of the debt to exceed the value of the collateral. Investment, and therefore output, is determined by collateral values. Investment and output in turn determine future collateral values and yield multiplier and cyclical effects.

Another approach is presented in an interesting series of studies by Greenwald and Stiglitz (1987, 1988, 1993) and Stiglitz (1992). They develop a model which produces a role for net worth or retained earnings in propagating cycles very much like that in Bernanke and Gertler (1989) but without any credit rationing or explicit modelling of informational asymmetries. Greenwald and Stiglitz rule out equity sales by using the type of arguments discussed on p.248, and they assume a perfect credit market; agency costs and asymmetric information, therefore, are used to justify their model but have no formal role to play within it. The role for net worth or retained earnings is introduced by assuming that the managers of firms are risk-averse. Risk-aversion is introduced either by placing a cost on bankruptcy in the utility function of managers or by making their utility a concave function of profits. In either case this makes managers wary of debt financing, since this increases the probability of bankruptcy. Hence managers increase investment and output if net worth or retained earnings increase, as this enables a substitution of retained earnings for debt. Increased investment and output in turn maintain relatively high retained earnings so shocks have persistent effects in much the same way as in Bernanke and Gertler (1989). Investment is also sensitive to the riskiness of the environment in a similar way to that discussed with respect to Williamson (1987) above, but now directly due to risk-aversion rather than agency costs.[18]

Yet another approach is taken by Blinder (1987) who, like Greenwald and Stiglitz, assumes that there is no equity market but allows for credit rationing. Firms unable to obtain credit are unable to hire factors of production and there is a failure of effective supply. Whilst the models Blinder uses are rather more ad hoc than those we have examined, he is able to take advantage of their relative simplicity to integrate issues of aggregate supply and aggregate demand. It would be interesting to do this in the models with richer microfoundations which have so far tended to focus more on the supply side; this is clearly an important task for future research.

Conclusions

We have reviewed the recent literature which provides microeconomic foundations for the long-established idea that financial factors are important in explaining business cycles. Agency costs, enforcement problems or risk-aversion have been shown to offer possible explanations of cycles. In this concluding section it is useful to compare these explanations with others, briefly review the available evidence and offer a general evaluation.

It is possible to discern four major theories of business cycles (see Stiglitz, 1992): the theory we have reviewed here based on asymmetries of information, the standard real business cycle (RBC) model (without informational asymmetries), the new classical model of price forecast errors, and models of imperfect competition. The standard RBC model has been much criticized on the grounds that it relies too much on large technological shocks as the primary source of economic fluctuations (it is difficult to find negative shocks to productivity of the size required to explain the Great Depression), and too much on intertemporal substitution effects to explain fluctuations in employment (see Mankiw, 1989, p.79). Furthermore it might be expected that changes in investment are dampened rather than exacerbated as entrepreneurs take advantage of the reduced costs of investment in recessions. The new classical model introduces the extra element that agents may mistake general price level changes for relative price changes and respond accordingly, but this still fails to explain the cyclical volatility of investment. Similarly whilst the degree of competition might decrease in a downturn causing further reductions in output, such effects also have unrealistic implications, e.g. that profit margins are countercyclical.

The view that financial factors are important in propagating business cycles might be viewed as complementary to the above three approaches. But it also has some advantages. First, as we have seen, it is not necessary that average productivity be affected since a small change in the perceived riskiness of the economic environment can have significant effects via agency costs (Williamson, 1987) or risk-averse behaviour (Greenwald and Stiglitz, 1993). Second it has been suggested by Gertler (1992) that small disturbances can induce large output fluctuations when borrowers and lenders enter into long-term debt contracts. In that case net worth includes the present discounted value of anticipated future project returns and, since this may be quite volatile, it is possible that small shocks produce large changes in the cost of external finance.

We have so far concentrated upon theoretical issues. There is some empirical evidence which offers some support to the importance of financial factors in explaining business fluctuations. Fazzari *et al.* (1988) examine evidence from the USA and conclude that financial constraints are important for many firms and that their investment is positively related to retained earnings or cash flow. An important aspect of their analysis is that they study investment behaviour in groups of firms with different financial characteristics and so offer a potential reconciliation of the mixed results of earlier studies on financial factors in investment: financial factors may well matter more for some firms (new and small ones) than for others (established and large ones). Fazzari and Petersen (1993) present further evidence of financial constraints on investment and consider the role of working capital. On a more macroeconomic level, Mishkin (1978) and Bernanke (1983) marshal evidence in support of the hypothesis that financial factors contributed to the depth and persistence of the Great Depression in the USA. More recently, Gertler and Gilchrist (1994) have found that, in the USA, small firms account for a disproportionate share of the manufacturing decline following a tightening of monetary policy, and that balance sheets significantly influence small firm inventory demand.

We close with a few words on policy implications and directions for future research. The literature is, in fact, rather light on detailed derivations of policy implications. Williamson, for example, concluded that 'there appears . . . to be no obvious role for "stabilization policy"' (1987, p.1215). Stiglitz, on the other hand, concluded that 'an effective stabilization policy of the government should be directed at overcoming the limitations of [this] rationing' (1988, p.320) after discussing agency problems but without offering a detailed analysis of such policy. Hillier and Worrall (1994) have examined a static version of Williamson's (1986) model and found the result that if rationing does occur then investment is likely to be excessive. The reason is that the market outcome produces too much monitoring, so that a cut in loan quantity increases welfare by cutting total monitoring costs. This result, whilst special to the model considered, illustrates that the policy implications of such models may be quite different from the aggregate demand management policies one might have expected from models with Keynesian features like interest rate rigidities. This is perhaps not so surprising once one remembers that the models do not so much explain rigidities, but explain why prices which could move are sometimes held constant by rational optimizing agents. Nevertheless, there is a clear need for further examination of policy, preferably within versions of the models extended to contain a government sector.

One way in which macroeconomic policy can have an effect in these models has been examined by Farmer (1984, 1985), who showed that it is possible for government debt and spending to generate real effects by changing the real interest rate and affecting the level of investment. The real interest rate is often fixed exogenously in the literature (it is given by the return on storage in the model on p.250, and by the imposed rate of time preference in the Greenwald and Stiglitz models discussed above), and it might be interesting to endogenize it to examine Farmer's arguments in other models. Another obvious implication of the models where retained earnings matter for investment is that the average as well as marginal rate of profits tax matters for investment, a lower average tax rate raising retained earnings and investment. This implication would appear to be testable.

We conclude that the literature on financial factors in macroeconomic models is interesting and promising. Further theoretical work is needed to examine the policy implications of this sort of model and to introduce roles for other potentially important factors such as imperfect competition and monetary surprises. Empirical work in this area is still relatively rare, and more would be useful.

Notes

1. For a general survey of the role of asymmetric information in the market for credit see Hillier and Ibrahimo (1993). An excellent earlier review with an emphasis upon the macroeconomic issues is Gertler (1988).
2. Recently Bohn and Gorton (1993) have presented a model of coordination failure which explains the use of non-indexed debt.
3. See the *General Theory* (1936, pp.149–55, and 319–25) for support for the claim that Keynes saw the volatility of investment to be the prime cause of output fluctuations. Chapters 2, 3 and 4 of Hillier (1991) offer further discussion of these and related issues.
4. For surveys of the literature examining the implications of agency costs for optimal contracts and the financial structure of the firm see Dowd (1993) and Harris and Raviv (1990).
5. The asymmetry of information identified in (a) is clearly *ex ante* because it exists before the debt contract is signed. That in (b) is *ex post* in the sense that the hidden action occurs after the contract is signed but *ex ante* in the sense that it arises before the project return is observed. That in (c) is clearly *ex post* since it arises after the contract and after the return is observed by the borrower.
6. All aggregate quantities are measured in per cohort terms.
7. Initially it may seem puzzling that it is the old generation who may invest rather than the young, but it is notationally convenient to keep investment and the returns to investment in the same period. Thus the young enter the economy with no wealth and must work in the labour market to acquire wealth. As they enter old age they make their investment choices and consume only at the end of their old age.
8. As stated above this section presents a slightly modified version of Bernanke and Gertler (1989). The main differences are: (1) they assume that Z is a discrete random variable; (2) they allow for stochastic monitoring of returns in the case of default whereas here we consider only deterministic monitoring: although the macroeconomic properties of the model are unaffected by this change it allows us to interpret the optimal financial contract as a standard debt contract; (3) they introduce a class of lenders in their model, who do not have access to a project of their own: this allows them to consider redistributions between

borrowers and lenders and show that such redistributions 'that may affect borrowers' balance sheets (as may occur in a debt-deflation) will have aggregate real effects' (Bernanke and Gertler, 1989, p.28).

9. On p.261 an aggregate shock is introduced into the model so the perfect foresight assumption will then be replaced by a rational expectations assumption.
10. We do not allow borrowers to borrow funds from several banks.
11. If s is large or q is small the demand for funds will be low, either because entrepreneurs can finance their own projects or because the expected return to investment in terms of output goods is low.
12. Credit is rationed, though, according to observable borrower characteristics, i.e. x, and is therefore quite different from the rationing of typical adverse selection models (see, for example, Stiglitz and Weiss, 1981).
13. We make no statements about whether it is at, below or above the second best level.
14. If there is default then monitoring occurs for sure because we do not consider stochastic monitoring, but the event of default is a random variable.
15. The existence of an equilibrium is guaranteed since given our assumptions $w(s)$ is continuous, $w(0) \geq 0$ and the maximum of $w(s)$ is w^*.
16. Multiple equilibria appear to be a common feature of models of coordination failure in the New Macroeconomics literature. See, for example, Cooper and John (1988) and Bohn and Gorton (1993) as well as Frank (Chapter 11 in this volume).
17. Scheinkman and Weiss (1986) referred to such enforcement problems in their study on borrowing constraints and aggregate activity, but did not present detailed microfoundations.
18. Stiglitz (1992) extends the analysis to argue that banks are a specialized kind of firm and that the principles which have been applied to other firms should also be applied to them. Thus, 'a reduction in the net worth of banks and an increase in the riskiness of their environment will lead them to contract their output, i.e. to make fewer loans' (Stiglitz, 1992, p.290).

Part V

Nominal rigidities and bounded rationality

13 Hedging, multiple equilibria and nominal contracts

Daron Acemoglu

Introduction

Not all transactions take place in spot markets. Contracts are often written to determine the terms of future trades. Understanding the nature of contracts is therefore an important step in the analysis of the allocation of resources. A robust observation is that most contracts are not indexed to the relevant price level despite the fact that prices at the time of transaction are uncertain and that the parties involved are risk-averse. This constitutes an important puzzle. A risk-averse agent would by definition benefit from a reduction in risk. It would thus seem to be the case that two risk-averse agents would always prefer to turn a nominal contract they have signed into a real one in order to reduce the risk they are bearing. The fact that apparently risk-averse agents write nominal contracts when the aggregate price level is unpredictable therefore constitutes an important puzzle.

A number of explanations have been offered to account for the presence of nominal contracts. Gray (1976), Fischer (1977), Cooper (1988a) show how in the presence of both supply and demand shocks the optimal degree of indexation is less than full, and therefore how some degree of nominal rigidity can arise as an equilibrium phenomenon. However, this approach does not in itself explain why the majority of observed contracts are purely nominal (for instance Card, 1983, reports that about 50 per cent of US wage contracts are not indexed at all, and the same applies with greater force to European contracts). Here we can either appeal to bounded rationality or to costs of writing complicated contracts. In particular, an argument similar to that of Akerlof and Yellen (1985a) and Mankiw (1985) can be suggested in this context (for example, Ball, 1988); costs of not choosing the right degree of indexation to prices may be of second order to individuals, while being first order to society. However, this argument is not very compelling. Many contracts involve transactions for very substantial amounts between risk-averse parties and costs of non-optimal contracts may be very large. Also in labour contracts signed between firms and unions, the cost of devising an optimal contract (especially if this contract requires full indexation) that will cover a large number of workers is likely to be small relative to the benefit. A different line of attack is to ask whether an equilibrium with nominal contracts can be a 'self-fulfilling prophecy'. This intuition first suggested by Fisher has been developed by Cooper (1990). In Cooper's model, risk-neutral firms write commodity and labour contracts with risk-averse workers/consumers. If all other firms are offering real contracts, it would be cheaper for each firm to offer a real contract but in contrast, if all other firms are expected to offer nominal contracts, the best thing that the firm in question can do is to offer a nominal contract.

This can be thought of as part of a more general phenomenon we refer to as 'hedging via contracts'. When outside insurance opportunities are limited, contracts written for transaction purposes will in general play two roles; allocation of resources and allocation of risks. Suppose next period's prices are random; a risk-averse agent would prefer to pay \$50 and receive \$50 in current prices simultaneously rather than receive \$50 in current prices and pay \$50 in fixed prices. This is because in the second situation he is subject to the risk created by price uncertainty. By making both his receipts and payments in current prices (or fixed prices) he avoids this risk. Thus the agent may well prefer a nominal contract and the statement we started with, that two risk-averse agents would *always* benefit from turning a nominal contract into a real one, is not true. This intuition suggests that the degree of indexation in contracts can in general be used to provide insurance against risks associated with variations in the price level (i.e. 'hedging').

We will first construct a simple model that exhibits these features and, as Cooper's economy (1990), has multiple equilibria. As our economy is simpler than Cooper's, the basic mechanism will be clearer and we will be able to locate all the equilibria. Also as in Cooper's example, all these equilibria will be efficient because they offer 'perfect hedging' to traders. However, more importantly we will show that when 'perfect hedging' through nominal contracts is not possible, all these equilibria disappear and we end up with a unique equilibrium without any nominal rigidity. It thus appears that nominal contracting equilibrium is not a generic phenomenon. Nevertheless, we also show that when other imperfections are present, this need not be so. This result is in a similar spirit to Ball and Romer (1991), who show that real rigidities increase the impact of small nominal rigidities. However, the imperfections we consider are not real wage rigidities, but limited insurance possibilities for agents and restrictions on the complexity of contracts that can be written. If there exist non-diversified risks associated with changes in prices, the equilibrium contracts will no longer be fully indexed. In particular, we show that in a dynamic overlapping-generations model where agents have limited access to financial instruments so that they cannot insure against relative price changes, a nominal contracting equilibrium may emerge, while a real contracting one fails to exist. Alternatively, when we restrict the complexity of contracts that can be written, we see that as well as the Pareto efficient real contracting equilibrium, an inefficient nominal contracting equilibrium becomes possible. We also obtain a number of interesting results; as in Ball (1988), writing a nominal contract creates a negative externality but through a completely different mechanism, and as in Mankiw (1985), small menu costs may have large impacts but again via a different channel. Underlying all these effects is the desire of risk-averse agents to hedge. The more risk-averse are these agents, the stronger is this mechanism. This can be contrasted to the Ball–Mankiw mechanism which gets weaker as the agents become more risk-averse, since increased risk-aversion raises the costs of non-indexation.

The plan of the chapter is as follows. The next section shows how a multiplicity of equilibria arises when agents trade with more than one party, but it is also shown that this multiplicity is not a generic feature. The following section analyses a dynamic economy with financial imperfections and demonstrates how a nominal contracting equilibrium may naturally arise. The fourth section shows how a restriction on the complexity of contracts can lead to the presence of Pareto ranked equilibria. The chapter concludes with a brief discussion in the fifth section.

A model with perfect hedging and multiple contracting equilibria

Consider an economy consisting of N agents denoted by $i = 1, \ldots, N$. At $t = 1$, if he buys one unit of good $i-1$ as input from agent $i-1$, agent i can produce one unit of consumption good and one unit of good i. Agent 1 buys his input from agent N. Good i has no other alternative use and the consumption good is sold in a competitive market. This situation can be represented as the circle shown in figure 13.1. The single arrows show the flow of inputs and the double arrows show the supply of the consumption good.

We assume that agents can write contracts at time $t = 0$ to determine how they will trade at $t = 1$. However, the general price level of this economy, p, at which they will also sell the consumption good to the market, is uncertain with a known distribution at $t = 0$, say because the nominal money supply is stochastic. Therefore, at $t = 0$ expected utility calculations are possible and we ask what the equilibrium level of indexation will be in the contracts that the agents write at $t = 0$. All agents again only care about their real income (i.e. deflated by the price of the consumption good, p), and we assume that they all start with no outside real income and have the same strictly concave (i.e. risk-averse) von Neumann–Morgenstern utility function $u(\cdot)$. We also

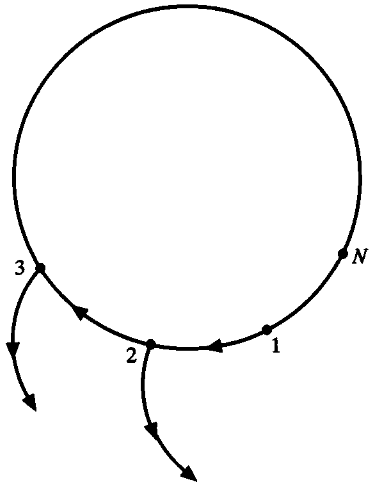


Figure 13.1 Trading relationships in the basic model

assume that goods are freely disposable and that breach of contract is prohibitively costly.

To determine the contracting equilibria of this economy we need to make certain assumptions about the structure of bargaining prior to contracting. We assume that the seller makes the first offer. If this is rejected, no contract is signed at $t=0$. At $t=1$, the buyer makes a final offer, and if this too is rejected, no trade takes place. There is no discounting between $t=0$ and $t=1$, but both parties incur a real cost $\beta<1$ if they cannot agree after the first offer. All agents make their offers simultaneously at time $t=0$ and $t=1$. The existence of the bargaining cost implies that agents would like to write a contract at $t=0$ rather than carry out transactions spot at $t=1$. As the game the agents play in this economy is dynamic, the equilibrium concept we use is subgame perfect. In what follows we will solve the game by backward induction, which will give us the subgame perfect equilibria. In all cases, there also exist other Nash equilibria, however they are supported by implausible out-of-equilibrium beliefs, and thus are not subgame perfect.

The time arrow for the events is given in figure 13.2. As can be seen, the price level, p, is not known at the time of contracting. However, since it has a well-defined distribution, agents can calculate their expected utilities. The situation is obviously very symmetric and this will lead to what we call 'perfect hedging'. We also concentrate on symmetric equilibria. We denote the price level by p and let c_i be the contract offered by i. In this case $c_i(p)$ is the nominal payment that $i+1$ makes to i in return for the input that he buys from i. We trivially have:

Figure 13.2 Time arrow for the basic model

Lemma 1: There exists a no trade equilibrium in which there is no production.

Lemma 2: There exists a real contracting equilibrium.

Proof: The method of proof is to consider a situation in which all parties other than i and $i-1$ offer and accept real contracts and to study the optimal contract offer of $i-1$. Suppose i expects to receive q_i for his input from $i+1$ and is bargaining with $i-1$. If he rejects $i-1$'s offer, at $t=1$ he can offer a spot contract at the price q'_{i-1} for this input. The minimum that will be accepted by $i-1$ is $q'_{i-1}=0$ because of free disposal (note that since $i-1$ has already agreed with $i-2$, he is producing for sure). In this case i would obtain $u(1+q_i-\beta)$ as he pays nothing for the input, receives q_i from $i+1$ but incurs the bargaining cost β. Agent $i-1$ can demand q_{i-1} at the first stage of bargaining that just gives this utility to i. Therefore,

$$u(1+q_i-q_{i-1})=u(1+q_i-\beta) \tag{1}$$

which gives $q_{i-1}=\beta$. However, we also need to show that $i-1$ would not prefer to offer any other contract. Suppose that all the other agents offer and accept real contracts at $q=\beta$ and that $i-1$ offers a contract $c_{i-1}(p)$ that is not real. In this case, from the above argument, i's utility is given by

$$Eu\left(1+\beta-\frac{c_{i-1}(p)}{p}\right)=u(1) \tag{2}$$

where i can obtain at least $u(1)$ because if he disagrees in the first period, he will get the input for free in the second period, which will give him $u(1)$. As i is risk-averse, i.e. $u(\cdot)$ is concave, $E(c_{i-1}(p)/p)<\beta$. Therefore, when he offers this contract to i, $i-1$'s utility is

$$Eu\left(1+\frac{c_{i-1}(p)}{p}-\beta\right). \tag{3}$$

As $u(\cdot)$ is concave and $E(c_{i-1}(p)/p)<\beta$, (3) is less than $u(1)$ which $i-1$ would have obtained with the real contract $q_i=\beta$. Therefore, $i-1$ will choose to offer a real contract. □

The existence of a real contracting equilibrium is naturally not surprising. However more interestingly:

Lemma 3: Any situation in which all agents offer and accept the same contract $c(p)$ is an equilibrium if $c(p)$ satisfies

$$Eu\left(1-\beta+\frac{c(p)}{p}\right)=u(1). \tag{4}$$

Proof: Take an arbitrary contract $c(p)$ and suppose all agents other than i and $i-1$ have signed this contract. At $t=1$, the second stage of bargaining, i can offer a price of zero for the input. Therefore, at $t=0$, $i-1$ can offer a contract $c^*(p)$ which satisfies

$$Eu\left(1+\frac{c(p)}{p}-\beta\right)=Eu\left(1+\frac{c(p)}{p}-\frac{c^*(p)}{p}\right) \tag{5}$$

and he will maximize

$$Eu\left(1+\frac{c^*(p)}{p}-\frac{c(p)}{p}\right) \tag{6}$$

If (4) is satisfied $c^*(p)=c(p)$ trivially maximizes (6) subject to (5). □

This result states that an agent who has signed a contract that exposes him to the risk of aggregate price variability can increase his utility by obtaining hedging from another contract which has the same form but now determines his income inflow rather than outflow. Condition (4) is necessary because the seller can always improve upon a contract that does not satisfy (4). It can thus be seen that a contract that does not satisfy (4) cannot be a symmetric equilibrium of this economy as it will not be best response to itself.

Lemma 3 also implies that a pure nominal contracting equilibrium (in which all contracts are purely nominal) exists. To see this, note that $Eu(1-\beta)<u(1)$ and

$$\lim_{Q\to\infty} Eu\left(1-\beta+\frac{Q}{p}\right)>u(1) \tag{7}$$

thus by continuity of $u(\cdot)$, there exists a value of Q such that $Eu(1-\beta+Q/p)=u(1)$ and also as $Eu(\cdot)$ is everywhere increasing in Q, this value of Q is unique.

Intuitively, if i has signed a contract that makes him pay a nominal amount, he would prefer to sign a contract that pays him the same amount irrespective of the price level so as to obtain perfect hedging rather than receive the expected value of this amount in real terms. To see the intuition more clearly, consider the case in which i and $i-1$ have signed a nominal contract and so have $i+1$ and $i+2$. If i offers a real contract (or a partially indexed contract) to $i+1$, then he needs to compensate $i+1$ for the extra risk he would be asking him to bear. However, i himself will be bearing more risk and also compensating $i+1$. Therefore, he cannot prefer to offer a contract that is not purely nominal. The crucial point is that all agents pay the same amount to each other, and this implies that a contract that has a different degree of indexation cannot be preferred to the contract that the others have signed. This also explains why this section is called 'a model with perfect hedging': because of the symmetric nature of the problem, all agents can obtain perfect hedging by writing nominal or partially indexed contracts. In the equilibrium in which all contracts are nominal, agents do not bear the risk of price variability, although their payments and receipts vary with the price level. This is also why in this model nominal contracts yield the same level of utility to all agents as the real contracting equilibrium. However, note that the different equilibria are not merely distinguished by the numéraire, because the prices of inputs relative to that of the consumption good vary across the equilibria.

We summarize the findings of this discussion in Proposition 1:

Proposition 1: In the above economy, there always exist a no trade equilibrium, a unique real contracting equilibrium, a unique nominal contracting equilibrium and also equilibria with partially indexed contracts. In all symmetric equilibria with trade, all agents get the same level of utility, $u(1)$.

Further, we can see that if a very small group of agents (for instance one of these N agents) suffers from money illusion and thus does not want to write an indexed contract, this will be sufficient to make the only possible outcome of this economy a nominal contracting one.

The model considered above is obviously very specific. First, we have a bargaining game in which each agent simultaneously bargains with more than one party, we also assume a specific ordering of moves; secondly, all agents are identical and by writing nominal contracts they can obtain perfect hedging; and thirdly there exist no fundamental risks associated with the price level in the sense that the outside wealth of the agents does not vary with changes in the price level.

The first assumption is made for tractability and simplification. As long as we maintain the symmetry assumption we can obtain exactly the same result by considering different bargaining structures before the contracting stage. Also the simultaneous bargaining assumption is not crucial as nominal contracts offer perfect hedging; if we change the model such that 1 first bargains with 2 and then 2 and 3 until at the end N bargains with 1, we will get the same result as in Proposition 1, since knowing that he can obtain perfect hedging, 1 can start by offering a nominal contract. In fact, looking at this set-up from an efficiency point of view we can see that all agents trading through nominal contracts is an efficient solution to the allocation problem. This again is due to the specific set-up of the model which gives perfect hedging. Also in this respect the model is similar to Cooper (1990) where the workers (the risk-averse agents in his model), sign both commodity and labour contracts and obtain perfect hedging.

If we introduce certain features in this model that will prevent perfect hedging, we lose the multiplicity of equilibria of Proposition 1. To illustrate this suppose that agent i has an obligation to pay a real amount x to an outside party (say, buy an input in spot market) and that all agents except i and $i-1$ have signed a partially indexed contract $c(p)$. Will i and $i-1$ do so too? The answer is no, because in this case i would not obtain perfect hedging and all the risks would be borne by him. As both utility functions are concave, they can share this risk and be better off. Thus, their best response to all other agents signing the contract $c(p)$ will be to sign a contract that is more closely indexed to the price level than $c(p)$, say $c^*(p)$. Thus both i and $i-1$ would be bearing some risk. However, if all other agents sign $c(p)$ and i and $i-1$ sign $c^*(p)$, the best response of $i-1$ and $i-2$ would be to share the risk that is now being borne by $i-1$. Reasoning similarly, no contract $c(p)$ that does not offer perfect hedging can be an equilibrium. There is a unique contract that does so, the real contract, $c^{**}(p)=qp$. Therefore,

Proposition 2: When perfect hedging is impossible and there are no price-related risks, there exists a unique equilibrium which is a real contracting one.

This proposition shows that our result in Proposition 1 was rather special. However, the caveat in Proposition 2 has to be borne in mind: 'when ... there are no price-related risks'. In practice, insurance markets often appear to fail in eliminating inflation-related variability in the real incomes of many households. When perfect hedging is not possible but there exist price-related risks, there will no longer exist a multiplicity of equilibria, yet the unique equilibrium of this economy will not be a real contracting one. We can model the presence of price-related variability in real incomes by writing the real income of agent i as $y_i(p)$. The existence of a unique equilibrium in this case follows a similar argument to that of Proposition 2. To see that it is not a real contracting one suppose everyone else signed a real contract except for i and $i-1$. Also i's real income is $y_i(p)$ and that of $i-1$ is $y_{i-1}(p)$. Generically these two real incomes will not be perfectly correlated, and thus by writing a contract $c(p)$ which makes the real payment a function of the price level, i and $i-1$ will be able to hedge some of these risks.

Proposition 3: When perfect hedging is impossible and there are price-related risks, there exists a unique equilibrium which is a generically not a real contracting one.

Proposition 3 establishes that a real contracting equilibrium may be a special case too. However, this does not really take us in the desired direction since this equilibrium will not be a nominal contracting one and thus does not help us in answering why the majority of real world contracts are not indexed at all. However, we will see in the next section that if the limited insurance opportunities take the specific form in which agents do not receive real returns on part of their savings, a nominal contracting equilibrium exists while a real contracting one fails to exist.

Nominal contracts in a dynamic economy with financial imperfections

In this section we consider an overlapping-generations version of the static model of the previous section. The main differences are as follows:

1. There are N agents of each generation in an infinite horizon economy and agent i of generation t is referred to as agent (i, t).
2. Agent $(1, t)$ does not buy from (N, t) but from $(N, t-1)$.

A diagrammatic representation of this set-up is given in figure 13.3. Agent (i, t) only cares about real income at time t, but he conducts his negotiations before the price p_t is revealed. This is also the time when agent (N, t) negotiates with $(1, t+1)$ to supply the latter with the necessary input.[1] All agents have the same utility function $u(\cdot)$. We also make the crucial assumption that agent $(1, t)$ can borrow money at zero nominal interest rate and pay for the purchase of a unit of input from $(N, t-1)$. Therefore, if he pays $c(p_{t-1})$, this will cost him $c(p_{t-1})/p_t$ in real terms. Basically, this assumption is equivalent to assuming that the current interest rate is fixed in nominal rather than real terms, or that agents cannot obtain real returns on some part of their wealth.[2] This assumption ensures that agents care about the returns and the price level in their period and it therefore makes a nominal contracting equilibrium possible. If agents receive interest at a

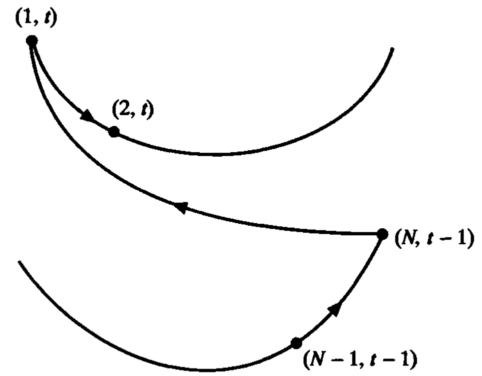


Figure 13.3 Trading relationships in the overlapping-generations model

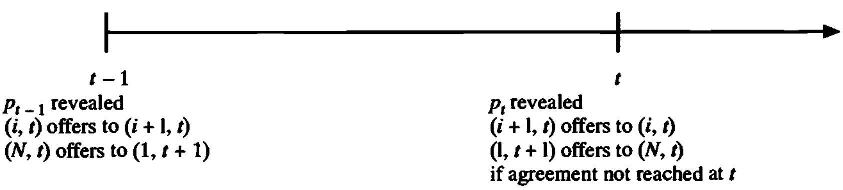


Figure 13.4 Time arrow for the overlapping-generations model

given real rate, then nominal contracting equilibrium disappears – see the discussion below and in the following section.

The other assumptions are the same as on p.000, and we concentrate on non-decreasing contracts. Contracting for trade at t is made at time t after p_{t-1} is revealed, and the trade between $(N, t-1)$ and $(1, t)$ takes place at $t-2$ before p_{t-1} is known. The time arrow is given in figure 13.4. The main result of this section is:

Proposition 4: In the overlapping-generation model outlined in this section

1. A pure nominal contracting equilibrium exists.
2. A pure real contracting equilibrium does not exist.

Proof: If $(1, t)$ and $(N, t-1)$ sign nominal contracts and so do (N, t) and $(1, t+1)$, we know from Lemma 3 that all agents at t will sign nominal contracts. Suppose at $t-1$, all contracts are at the nominal price Q_{t-1} and at t, at the nominal price Q_t. At the second stage of bargaining, $(1, t)$ can get the input at zero price and his expected utility at time $t-2$ (i.e. when he is offered the contract, see figure 13.4), will be

$$E_{t-2}u\left(1+\frac{Q_t}{p_t}-\beta\right)=u(1). \tag{8}$$

This expression is equal to $u(1+y)$ by the argument in the proof of Lemma 3 because Q_t is an equilibrium price for a nominal contract. Now $(N, t-1)$ can offer $c^*(p_{t-1})$ such that

$$E_{t-2}u\left(1+\frac{Q_t}{p_t}-\frac{c^*(p_{t-1})}{p_t}\right)=E_{t-2}u\left(1+\frac{Q_t}{p_t}-\beta\right) \tag{9}$$

and $c^*(p_{t-1})$ will be chosen to maximize $(N, t-1)$'s utility

$$E_{t-2}u\left(1-\frac{Q_{t-1}}{p_{t-1}}+\frac{c^*(p_{t-1})}{p_{t-1}}\right). \tag{10}$$

The contract can only depend on p_{t-1} (i.e. not on p_t) because agent $(N, t-1)$ consumes at $t-1$ before p_t is revealed. Now we show that there exists a nominal contracting equilibrium with $Q_{t-1}=Q_t=Q$. The argument goes in three steps:

(a) Suppose Q_t is expected to be equal to Q_{t-1}. Maximization is then achieved by setting $c(p_{t-1})=Q_{t-1}$. Also as Q_{t-1} is an equilibrium contract offer it must satisfy condition (4).

(b) When $(1, t)$ signs a contract $c(p_t)=Q$ which satisfies condition (4), by Lemma 3, all agents at t will sign the same nominal contract.

(c) The first two steps show that the situation is an equilibrium; when an agent believes that everyone else will offer a nominal contract, he too signs a nominal contract. Finally we need to show that the equilibrium is subgame perfect. In other words, when (i, t) deviates and takes account of the fact that the offer that $(i+j, t)$ will make to $(i+j+1, t)$ may no longer be a pure nominal contract, he still prefers to offer a nominal contract. The key observation is that $(1, t)$ can get $u(1)$ by rejecting $(N, t-1)$'s offer and by waiting until $t-1$ to get the input for free. Thus $(N, t-1)$'s offer $c^*(p_{t-1})$ must satisfy:

$$E_{t-2}u\left(1+\frac{c(p_t)}{p_t}-\frac{c^*(p_{t-1})}{p_t}\right)=u(1) \tag{11}$$

where $c(p_t)$ is the best offer that $(1, t)$ can make to $(2, t)$ in the case where he accepts $c^*(p_{t-1})$. $c^*(p_{t-1})$ will be chosen to maximize

$$E_{t-2}u\left(1+\frac{c^*(p_{t-1})}{p_{t-1}}-\frac{Q_{t-1}}{p_{t-1}}\right). \tag{12}$$

Now $c(p_t)$ must satisfy (4) since it is the best offer of $(1, t)$ and $(1, t)$ only contracts with agents from his generation. Next choose Q_t such that it also satisfies (4) and note that $c(p_t)$ varies with price level while Q_t does not, therefore[3] $E_{t-2}(c(p_t)/p_t) \leq E_{t-2}(Q_t/p_t)$. But also $E_{t-2}(c^*(p_{t-1})/p_t) \leq E_{t-2}(c(p_t)/p_t)$ can be deduced from (11) by the same argument since p_{t-1} varies less than perfectly with p_t, thus the left-hand side of (11) is more variable than the right-hand side, and thus must have a higher expected value. This implies that $E_{t-2}(c^*(p_{t-1})/p_{t-1}) < E_{t-2}(Q_t/p_{t-1})$ and so

$$E_{t-2}u\left(1+\frac{Q_t}{p_{t-1}}-\frac{Q_{t-1}}{p_{t-1}}\right) \geq E_{t-2}u\left(1+\frac{c^*(p_{t-1})}{p_{t-1}}-\frac{Q_{t-1}}{p_{t-1}}\right) \qquad (13)$$

which proves that a nominal contract is preferred to any other contract, and therefore we have a dynamic nominal contracting equilibrium.

To prove 2 above take the case in which $(1, t)$ signs a real contract at the price q^* with $(N, t-1)$. He would try to maximize

$$E_{t-2}u\left(1-\frac{q^*p_{t-1}}{p_t}+\frac{c(p_t)}{p_t}\right) \qquad (14)$$

by choosing $c(p_t)$ at time $t-1$ after p_{t-1} has been revealed. So at the time he is offering the contract, (i, t) is bearing the price risk through q^*p_{t-1}; he could therefore reduce this risk by offering $c(p_t)$, that is increasing in the price level p_t. If $c(p_t)$ satisfies (4), it will be accepted and (i, t) will be better off. Therefore we cannot have a real contracting equilibrium. □

The intuition of the proposition is that nominal contracts are offering sufficient hedging, in particular to agents who buy in $t-1$ and sell in t. However, the crucial point is that these agents care about their returns in the period they live in (i.e. in which they spend their money) and in the absence of real returns on the money they save (by paying a lower price at time $t-1$), the best they can do is to sign a purely nominal contract. The essence of this result is that because of the no real rate assumption, agent $(1, t)$ is forced to bear some risks related to relative price variability (i.e. changes in p_t relative to p_{t-1} because $(1, t)$ cares about p_t while $(N, t-1)$ cares about p_{t-1}). The best thing to do is to write a nominal contract that offers some hedging. This leads to the existence of a nominal contracting equilibrium.

To illustrate, take agent $(1, t)$; if we let the amount he pays be Q, then $(1, t)$ cares about Q/p_t. Suppose he signed a real contract, then $Q = qp_{t-1}$. However, as p_{t-1} and p_t are not perfectly correlated, he would be subject to more variability in his return. The same applies to any contract that is not purely nominal. Given that $(N, t-1)$ has already signed a nominal contract, he prefers a nominal contract too, and there is no reason to deviate. On the other hand a real contracting equilibrium does not work because in the case where $(N, t-1)$ has already signed a real contract, $(1, t)$ and $(N, t-1)$ would still prefer to write a non-real contract and share the risks that a real contract would have imposed solely on $(1, t)$.

Another way of viewing this result is that the 'imperfection' in the 'financial sector' of this economy, which is in essence the absence of indexed bonds, makes the optimal contract nominal. However, in this case we can ask why is the interest rate not fixed at a real rate independent of the change in the price level between $t-1$ and t. First, as the model stands there is no reason to have such a market. Secondly, agents may need to carry some money for transactions (matching imperfections) or as precaution (uncertainty) on which they will not get any interest, and this will be sufficient to get us a result as in Proposition 4. Finally, when all contracts are nominal, the banking sector may prefer to pay at a nominal rate.

We can also note that the no interest rate assumption is not necessary for this result. We can have a constant interest rate or alternatively an interest rate paid at t that depends only on p_t. What would reverse the result would be to have a rate that depends on the ratio p_{t-1}/p_t. If this were the case we would have a real contracting equilibrium and no nominal contracting equilibrium.

Finally we can discuss the efficiency properties of this equilibrium. The nominal contracting equilibrium is efficient given the financial imperfection. All agents except $(1, t)$ receive full insurance and $(1, t)$ does as well as he can given the fact that there exist non-diversifiable risks associated with changes in prices. Thus although we have demonstrated that a nominal contracting equilibrium is possible when there exist other imperfections in the economy, we have not offered an explanation of how nominal contracts can arise even when they create inefficiencies. To analyse this case, we now turn to a situation in which there are restrictions on the set of possible contracts.

A model with imperfect hedging

In this section we will analyse a contracting situation between three risk-averse parties, a worker, an intermediate product firm and a final product firm. However, in contrast to the previous sections, we will restrict the strategy space of these agents and assume that only purely real and nominal contracts can be written. This is certainly a radical assumption, and the arguments suggested in the Introduction can be used to say that writing partially indexed contracts should not be too expensive. We can, however, justify a restricted strategy space by bounded rationality; it is difficult for agents to understand a general function $c(p)$ which maps every price level into a payment level. It can also be argued that if one side disobeys the contract, the parties need to go to a court. It is conceivable that partially indexed contracts are prohibitively costly to adjudge and one of the parties can always disobey the partially indexed contract. This may force the parties to write only purely nominal and real contracts. A more subtle and plausible justification will be to argue that, as we will see, partially indexed contracts will never be used in equilibrium, thus if it is costly to learn to do these calculations, the agents in this economy may decide that it is not worthwhile to acquire the skills necessary to understand these contracts despite the fact that it may not be too costly to do so. However, if they do not know how to use these complicated contracts, an inefficient nominal contracting equilibrium becomes possible (see below).

Let us denote the utility function of the three players by U_W, U_I and U_F. All these utility functions are defined over real incomes and are concave. As in the last section the price level which determines how much a given nominal amount is worth in real terms is unknown at $t=0$ when contracts are written. As I is faced with two take-it-or-leave-it offers, it will be forced down to its outside option which is zero. If we let the contract offered by the worker be $W(p)$ and the contract offered by firm F be $Q(p)$ (mappings that give the nominal payment for each realization of the price level), then we need

$$EU_I\left(\frac{Q(p)}{p}-\frac{W(p)}{p}\right)=0 \tag{15}$$

where the expectation is taken over the realizations of p, and in this case the worker will get utility $EU_W(W(p)/p)$ and F will get $EU_F(1-Q(p)/p)$ as it receives 1 in real terms for the unit of output it sells in the spot market at time $t=1$, and pays $Q(p)$ to I in nominal terms. Take a real contract offer, $Q(p)=qp$, by firm F, then W's best response is to offer the real contract $W(p)=qp$. Further $Q(p)=qp$ is the best response for F against $W(p)=qp$. Of course, this does not tell us how q is determined, but shows that a real contracting equilibrium exists. However, the argument used for Proposition 2 shows that $W(p)=Q$ is not a best response to $Q(p)=Q$ and so a nominal contracting equilibrium does not exist unless we restrict the strategy space. Therefore,

Proposition 5: When general contracts can be written, the unique equilibrium is the efficient outcome in which all contracts are fully indexed.

Now suppose that we restrict the strategy space of the agents such that they can only write nominal and real contracts and that W offers $W(p)=Q$

to I; what is F's best response? F can offer either $Q(p)=Q$ which will give $EU_I=0$ or $Q(p)=qp$ such that

$$EU_I\left(q-\frac{Q}{p}\right)=0. \tag{16}$$

Therefore the question is whether $EU_F(1-Q/p)$ is larger than $U_F(1-q)$ where q satisfies (16). If U_I is more concave than U_F in the sense that $U_I(\cdot)=g(U_F(\cdot))$ where $g(\cdot)$ is everywhere concave, then $EU_F(1-Q/p)$ will be larger than $U_F(1-q)$. Similarly, $W(p)=Q$ will be the best response within this restricted strategy set against $Q(p)=Qp$ if U_I is more concave than U_W. To see that concavity is sufficient for this argument, take q_F such that $EU_F(1-q_F)=EU_F(1-Q/p)$ and q_I such that $EU_I(q_I-Q/p)=0$. If $q_F>q_I$ then F will prefer to offer a real contract. Thus for a nominal contracting equilibrium to exist we need $q_F\leq q_I$, and this is equivalent to U_I being more concave than U_F. An analogous argument establishes that U_I needs to be more concave than U_W for W to offer a nominal contract when F is expected to do so. Therefore,

Proposition 6: In the above model, with only real and nominal contracts allowed, we can have nominal and real contracting equilibria if U_I is more concave than both U_F and U_W.

This result is again cast in a very specific and simple model, and so does not have general applicability. However, it again demonstrates that if we have other imperfections nominal contracting equilibria can again exist. Thus although this model does not seem very robust, the intuition that follows from it is.

Another interesting feature of this model is that in contrast to the model considered in the last section, there is a clear welfare ranking between the two equilibria. The real contracting equilibrium *ex ante* Pareto dominates the nominal contracting one. This of course begs the question of why the nominal contracting equilibrium will ever arise. The natural answer is obviously a coordination failure among the agents. In fact, when F and I write a nominal contract they create a negative externality on W, and vice versa when W and I write a nominal contract. These externalities imply that the nominal contracting equilibrium is inferior to the real contracting equilibrium. However, it is optimal for W to offer a nominal contract if he believes that F is doing so.

We can also compare the results of this section to related literature on indexation. First, the multiplicity obtained here is different from the one on p.281 and in Cooper (1990), because perfect hedging no longer holds and the nominal contracting equilibrium is now inefficient (Pareto dominated). Secondly, as in Ball (1988), an agent who chooses less than first-best indexation (i.e. signs a nominal contract) is creating a negative externality on other agents. In particular, when firm *F* offers a nominal contract to *I*, the worker *W* is suffering as a result. In Ball (1988), these effects arise through the aggregate demand externality when we close the economy with a demand side whereas in this section, the mechanism is the risk-aversion of firm *I*; for firm *I*, contracts do not only allocate resources but also risk, so when it signs a nominal contract with *F*, it would also like to sign a nominal contract with *W*, ignoring the costs that these nominal contracts impose on other agents. Finally our results can be interpreted in an alternative way in order to enable easier comparison with the 'near-rationality' or 'small menu costs' literature. Note that the optimal contract in our economy is a real one. Thus, as suggested earlier, if there exists a small cost of investing in a technology that will enable agents to write complicated contracts, they may find this unnecessary, because in equilibrium these contracts will not be written. However, once these contracts are unavailable, as Proposition 6 shows, we can no longer rule out the inferior nominal contracting equilibrium.[4] Thus, innocuous behaviour at the individual level may have important aggregate consequences, as in the Akerlof–Yellen–Mankiw studies. Nevertheless, as our discussion illustrates, the intuition is quite different and relies on the risk-aversion of the agents – in particular, on the risk-aversion of firm *I*. Also, investing in the technology necessary to write complicated contracts would create an externality which is again not internalized. If *F* invests in this technology the nominal contracting equilibrium will disappear and *W* will also benefit, but this of course does not feature in *F*'s calculations.

Discussion and conclusion

This chapter builds upon the intuition that for an agent who is already subject to the risk of price variability, a contract that links the level of his real payments (or receipts) to the price level can provide hedging and increase his welfare. As Cooper (1990) shows this may contribute to our understanding of why most observed contracts are not indexed on the price level. However, we also demonstrate that the existence of an equilibrium with nominal contracts requires that they are as good in allocating risks as fully indexed contracts are. Thus we need perfect hedging. As in most situations certain transactions have to take place in the spot market, perfect hedging will not be possible and thus a nominal contracting equilibrium will not exist.

However, this is not the whole story. Trading in spot markets may imply that nominal contracts do not offer perfect hedging, but in the presence of fundamental risks related to the price level, agents can do better than write real contracts. We show that for certain forms of financial imperfections, a nominal contracting equilibrium may exist in a dynamic context, especially when agents cannot insure against risks of relative price variability, while in the same situation a real contracting equilibrium fails to exist. However, in this equilibrium nominal contracts efficiently allocate risks, and thus we are unable to answer the question of whether nominal contracts may introduce inefficiencies. In order to tackle this issue we restrict the complexity of the contracts that can be written. The issue of hedging enriches the analysis at this juncture, too. If bounded rationality or transaction costs make partially indexed contracts too costly to write, a nominal contracting equilibrium may exist under less restrictive circumstances. A Pareto inferior nominal contracting equilibrium may arise in this case, demonstrating that even though nominal contracts may have some positive role in efficiently allocating risk, they may also cause important distortions.

Therefore, this chapter identifies hedging as a mechanism that can potentially explain the existence of nominal contracts. As emphasized in the chapter, this is unlikely to be the whole story, but it can contribute to our understanding of nominal contracts. In particular, we can ask whether certain nominal contracts we observe in the real world provide hedging. First, since most firms buy and sell through predetermined contracts, our theory predicts that these firms could have both sets of contracts nominal or real but should not have nominal contracts determining their payments and real contracts determining their receipts. Secondly, the intuition of this chapter can be applied to mortgage contracts. These contracts determine the payments of agents who often receive wages and salaries through nominal contracts. We should therefore expect mortgage payments to fall in real terms with increases in the price level. Finally, since the tax brackets in most countries are infrequently adjusted for inflation, this can also act as an imperfection against which nominal contracts can be used to obtain hedging. On the other hand, an important class of nominal contracts is the wage contracts, but hedging appears to be less of an issue here. Some workers sign contracts that make the level of their real payments a function of the price level – such as nominal mortgage contracts – and wage contracts may be offering hedging against these risks. However there is much more at stake when nominal wage contracts are signed because they also make the level of employment a function of the price level and they are unlikely to be explained by hedging only. Thus there are still many unanswered questions before we can reach an understanding of why most contracts, and in particular most wage contracts, are in nominal terms.

Notes

This chapter is based on ch. 4 of my Ph.D. thesis at the LSE. I am grateful to Charlie Bean, John Moore, Neil Rankin, Kevin Roberts, Ulf Schiller, Andrew Scott and seminar participants at the LSE and the University of Warwick Macroeconomics Workshop (July 1993) for helpful comments.

1. This assumption implies that there are at least $2N+1$ agents alive at any point in time, but crucially all agents of generation t only consume at time t (I am grateful to Neil Rankin for pointing out a confusion about the number of agents alive in the earlier version of the chapter).
2. Putting it differently, the existence of financial imperfections creates risks against which agents cannot insure. An alternative form for this assumption is to assume that each agent living at t receives money income M_t at time $t-2$ but can get only nominal interest rate on this amount. However this assumption complicates the analysis because the real income received by agents also depends on the price level, and they will prefer contracts that offer them hedging against this risk too.
3. To see this, note that if $E(c(p)/p)=E(Q/p)$, then $Eu(1-\beta+c(p)/p) > Eu(1-\beta+Q/p)$ by the concavity of $u(\cdot)$.
4. If we model this situation in a game theoretic context, we will end up with a mixed strategy equilibrium in which the cost to write complicated contracts will be incurred by F and W some of the time.

14 Information acquisition and nominal price adjustment

Torben M. Andersen and Morten Hviid

Introduction

The failure of prices to adjust plays a crucial role in Keynesian macroeconomics. In particular the failure of nominal prices to adjust instantaneously to nominal shocks is important for the role of demand shocks as a source of business cycle fluctuations. Empirical evidence shows that nominal shocks contribute to business cycle fluctuations (see, for example, Andersen, 1994). Insufficient price adjustment may at a general level be caused by either adjustment being costly in terms of explicit or implicit costs or by price-setters lacking sufficient information to make the proper adjustment.

The most widespread model of price inflexibility is the so-called menu cost model, assuming price adjustment to be costly. This model has recently been extensively analysed (see Ball, Mankiw and Romer, 1988 and Andersen, 1994 for introductions and references) and has provided a number of important insights on price adjustment. Still, the empirical relevance of price adjustment costs remains an open question, and it is not obvious whether price adjustment costs are more important than costs of adjusting quantities.

An alternative approach to the explanation of price rigidities is to consider informational problems arising in decentralized economies. Small departures from the benchmark case of full information are sufficient to cause adjustment failures, especially if firms are differently informed or if there is confusion between permanent and transitory changes (see Andersen, 1994). In this chapter we extend the analysis of informational problems by making information acquisition by actors in the economy explicit.

This has several purposes. First, it allows us to check whether menu costs can be interpreted as information costs, and thus act as a simple modelling device. Secondly, it allows us to check whether information acquisition has other implications for price adjustment than the obvious one that if the value of information falls short of the costs firms do not acquire information and hence there will be price adjustment failures. Finally, the analysis is relevant for the more general question of the ability of decentralized market economies to aggregate and disseminate information. This issue plays a central role in economics but has hitherto been analysed rigorously only for competitive financial markets (see Grossman and Stiglitz, 1976, 1980; Radner, 1981). Clearly, the question is highly relevant in relation to product markets as analysed in this chapter.[1]

The framework adapted here is a model of monopolistic competition which has proved a useful vehicle for macroeconomic analysis. Firms set prices simultaneously given their private information, but price decisions interact since firms compete over market shares. Considering first an exogenous information structure with informed and uninformed firms, we find that the uninformed firms have a disproportionately large effect on the price level due to strategic complementarities in price-setting.

Endogenizing the information structure by assuming that information can be acquired at a cost, we find as expected that for sufficiently high information costs information acquisition is never worthwhile and nobody acquires information, implying completely rigid nominal prices. However, there are more important results going beyond the obvious implications of information acquisition being costly. First, there may be multiple equilibria in information acquisition, due to a self-fulfilling property in information acquisition. The incentive for any single firm to acquire information depends on how many other firms are expected to acquire information. We show that at some cost levels, there exist equilibria in which all firms acquire information and others in which no firms acquire information. This implies that even for moderate information costs no acquisition may be an equilibrium outcome, and moreover information acquisition and thus price adjustment may be path-dependent. Secondly, the interplay between firms may also preclude the existence of equilibrium in the information acquisition game. This arises due to a Grossman–Stiglitz (1976, 1980) type of information paradox. Information is valuable to firms, and if no other firms are acquiring information there is an incentive for each firm to do so. However, if all firms are acquiring information it is not worthwhile for any single firm to do so. Our setting differs from Grossman and Stiglitz as firms cannot infer the information held by informed firms directly, but their decisions are interrelated through the implications of information for the aggregate price level, which in turn affects the optimal price for each firm and the incentive to acquire information.

The second section sets up a model with monopolistic competition. The adjustment of the aggregate price level to nominal shocks is considered in the third section, presuming an exogenous asymmetry in information between firms. The fourth section endogenizes information acquisition and considers equilibria to the information acquisition game. The fifth section summarizes the chapter's arguments.

Price determination

Consider a monopolistically competitive economy, where demand for goods produced by firm $j \in J$ is given as[2]

$$d_j = \left(\frac{P}{P_j}\right)^{\alpha} \left(\frac{M}{P}\right) \quad \alpha > 1, \tag{1}$$

where P_j is the price charged by firm j, P the aggregate price index, and M the level of nominal demand (or money stock). The aggregate price level is defined as

$$P = \prod_j P_j^{1/J}$$

where J is the number of firms.

To focus on nominal adjustment, the money stock M is the only state variable and it is assumed to be a random variable. For tractability, it is assumed to be lognormally distributed, i.e.

$$\ln M \sim N(0, \sigma^2) \tag{2}$$

To the extent that firms can condition their prices on the true value of M, the general price index will also be a random variable. The specific form of the demand function is chosen to capture two essential variables affecting demand, namely relative prices and the aggregate demand level.

Firms produce output subject to the following decreasing returns to scale production function

$$y_j = \frac{1}{\gamma} l_j^{\gamma} \quad 0 < \gamma < 1 \tag{3}$$

where l_j is labour input.

As is usual for monopolistic competition models, each firm ignores its effect on the other firms' prices which occurs through the price index. Firms set prices simultaneously to maximize expected real profits $E\Pi_j$ conditioned on their information set I_i which, given (1) and (3), can be written as

$$E\Pi_j = E\left[\frac{P_j}{P}\left(\frac{P}{P_j}\right)^{\alpha}\left(\frac{M}{P}\right) - \frac{W}{P}\left[\gamma\left(\frac{P}{P_j}\right)^{\alpha}\left(\frac{M}{P}\right)\right]^{\frac{1}{\gamma}} \middle| I_j\right] \tag{4}$$

where W is nominal wage which is exogenously determined. To rule out nominal wage rigidities, it is simply assumed that the nominal wage level is proportional to the nominal shock variable, i.e.

$$W = M.$$

It is well known that nominal wage rigidity causes nominal price rigidity. Hence, the above mentioned assumption serves the purpose of showing that in the process of price-setting there may be reasons for nominal rigidities beyond those arising in wage-setting.

The first order condition for profit maximization can be written as

$$\frac{\partial E\Pi_j}{\partial P_j} = E\left[(1-\alpha)P_j^{-\alpha}P^{\alpha-2}M - \gamma^{\frac{1}{\gamma}}M^{1+\frac{1}{\gamma}}P^{\frac{\alpha-1-\gamma}{\gamma}}\left(-\frac{\alpha}{\gamma}\right)P_j^{-\frac{\alpha}{\gamma}-1} \middle| I_j \right] = 0. \tag{5}$$

We have assumed above that M is lognormally distributed. Below we also show that P is lognormally distributed. This allows us to simplify (5). If x is random variable lognormally distributed, we have that (see, for example, Aitchison and Brown, 1957)

$$\ln E(x) = E(\ln x) + 1/2\mathrm{VAR}(\ln x). \tag{6}$$

Using (6) on (5), the resulting price decision rule can be written as

$$\ln P_j = \lambda_{0j} + \lambda_1 E(\ln P|I_j) + \lambda_2 E(\ln M|I_j) \tag{7}$$

where expressions for λ_{0j}, λ_1, and λ_2 are given in appendix 1, and where

$$\lambda_1 + \lambda_2 = 1$$

reflecting that the nominal price quoted by firm j is homogenous of degree 1 in the two exogenous nominal variables P and M.

Asymmetric information

We want to investigate the incentives of an individual firm to acquire information, but before we proceed with this task, it is useful to analyse the case of an exogenous information structure. We therefore first solve the model assuming two groups of firms – informed who know the realization of M, and uninformed who only know the distribution of M as given in (2). Index these firms by I and U, respectively. Let h be the fraction of informed firms and assume that h is known by all. Although this assumption in itself raises informational problems, it is made here to focus on the implications of lack of information concerning exogenous state variables.[3] Note that for $h > 0$, a fraction of firms can condition their price on the realization of M, making the price index a random variable from the point of view of the uninformed firm.

Because of the symmetry, all informed firms will quote the same price and similarly for the uninformed firms. Denote the price quoted by an informed firm by P_I and by an uninformed firm P_U. The aggregate price level becomes in this case

$$P = P_I^h P_U^{1-h} \tag{8}$$

where h is the fraction of firms being informed.

We shall prove the existence of an equilibrium to the model under this asymmetric information structure, and provide a characterization of equilibrium prices by use of the so-called undetermined coefficients method. Conjecture that the equilibrium price is determined as

$$\ln P = \rho_0 + \rho_1 \ln M. \tag{9}$$

Then

$$E(\ln P|I_I) = \rho_0 + \rho_1 \ln M$$
$$E(\ln P|I_U) = \rho_0$$

and we find

$$\ln P_I = \lambda_{0I} + \lambda_1(\rho_0 + \rho_1 \ln M) + \lambda_2 \ln M \tag{10}$$

$$\ln P_U = \lambda_{0U} + \lambda_1 \rho_0. \tag{11}$$

Using the definition of the aggregate price level given in (8), we find by inserting (10) and (11) that the resulting aggregate price level can be written as

$$\ln P = h(\lambda_{0I} + \lambda_1\rho_0) + (1-h)(\lambda_{0U} + \lambda_1\rho_0) + h(\lambda_2 + \lambda_1\rho_1)\ln M. \tag{12}$$

For (12) to be consistent with (9), we require

$$\rho_0 = h\lambda_{0I} + (1-h)\lambda_{0U} + \rho_0(h\lambda_1 + (1-h)\lambda_1)$$
$$\rho_1 = h(\lambda_2 + \lambda_1\rho_1)$$

or

$$\rho_0 = \frac{h\lambda_{0I} + (1-h)\lambda_{0U}}{1-\lambda_1} \tag{13}$$

$$\rho_1 = \frac{h\lambda_2}{1-h\lambda_1} = \frac{(1-\lambda_1)h}{1-h\lambda_1}. \tag{14}$$

Consequently, the aggregate price level is determined in (9) with the coefficient given above in (13) and (14), i.e.

$$\log P = \frac{h\lambda_{0I} + (1-h)\lambda_{0U}}{1-\lambda_1} + \frac{(1-\lambda_1)h}{1-h\lambda_1} \ln M. \tag{15}$$

Hence, we have found a unique rational expectations equilibrium to the model within the class of loglinear solutions. As assumed above, P is lognormally distributed with mean ρ_0 and variance $(\rho_1\sigma)^2$. Further, the covariance between $\ln P$ and $\ln M$ is $\rho_1\sigma^2$.

We note that the coefficient to the money stock is less than one ($\rho_1 < 1$), reflecting that there is a nominal rigidity. This is no surprise given that a fraction $1-h$ of the firms is uninformed about the nominal shocks.

It is more interesting to note that the adjustment coefficient is actually less than the fraction of informed firms, i.e.

$$\rho_1 < h \quad \text{for} \quad h < 1.$$

Hence, even though a fraction h of firms knows the true nominal shock the aggregate price level is adjusted by less than this fraction. The reason is the strategic complementarity in price-setting ($\lambda_1 > 0$), implying that informed firms take into account the prices set by uninformed firms, and since these prices by definition cannot be adjusted to the nominal shock variable, it follows that the informed adjust their prices by less. This is a variant of the result proven by Haltiwanger and Waldman (1989) that with strategic complementarities the naive agents – here, the uninformed firms – have a disproportionately large effect on the equilibrium outcome compared to the sophisticated agents – here, the informed firms.

The importance of the interaction between differently informed price-setters is seen clearly by comparing the equilibrium price level (15) to the hypothetical price level given as a weighted average of the price level if all firms are either informed or uninformed, i.e.

$$\ln\tilde{P} = h\ln\tilde{P}_I + (1-h)\ln\tilde{P}_U \tag{16}$$

where $\ln\tilde{P}_I(\ln\tilde{P}_U)$ is the aggregate price level if all firms are informed (uninformed). The prices are weighted by the fraction of informed and uninformed agents respectively. The price level in (16) does not, therefore, take into account any interaction between differently informed price-setters.

We find that

$$\ln P - \ln\tilde{P} = (\rho_1 - h)\ln M.$$

Hence, the interaction between differently informed firms does not affect the average level of prices but only the adjustment to the state variable. The interaction implies that the price level becomes less sensitive to the state variable, as is seen by noting that

$$\chi(h) \equiv \rho_1 - h = \frac{(h-1)h\lambda_1}{1-h\lambda_1} < 0 \quad \text{for} \quad 0 < h < 1.$$

It turns out that the reduced sensitivity of the price level to the state is dependent on the fraction of informed firms

$$\frac{\partial \chi}{\partial h}=\frac{\lambda_1+2h\lambda_1-h^2\lambda_1^2}{(1-h\lambda_1)^2}\gtreqless 0 \quad \text{for } h\lesseqgtr 1-\sqrt{1-\lambda_1}$$

$$\frac{\partial^2 \chi}{\partial h^2}=\frac{2\lambda_1(1+\lambda_1)}{(1-h\lambda_1)^3}>0.$$

Hence, χ is convex in h, zero at the upper and lower bound on h, negative elsewhere and achieves a minimum for an interior value of h. It follows that the interaction between differently informed price-setters has important implications for the behaviour of aggregate prices, and we next turn to an analysis of how this affects the incentives to acquire information.

Endogenous information acquisition

In the preceding section, the information structure was exogenously imposed. In this section, the information structure is endogenized. Assuming information acquisition to be costly, the issue is whether firms find it worthwhile to acquire information, thus removing all uncertainty that they face, and how this in turn affects the formation of prices.

The decision whether to acquire information is an *ex ante* decision where the firm must decide whether it will incur a fixed real cost c of acquiring information on the state of the market[4] or whether it will stay uninformed. Relevant for the information acquisition decision is thus the expected profit when informed and uninformed. Further we assume that price-setting occurs after the fraction of informed firms h has become common knowledge.

Using the first order condition in (5) – which must hold for all values of h – in the expected profit expression (4), the equilibrium expected profits of a firm who has acquired information can be written as

$$E\Pi_I^*(h)=\left(1-\frac{\gamma}{\alpha}(\alpha-1)\right)E\left[\left(\frac{P}{P_I}\right)^{\alpha-1}\left(\frac{M}{P}\right)\right]. \tag{17}$$

If the firms decides not to acquire information, expected profits are

$$E\Pi_U^*(h)=\left(1-\frac{\gamma}{\alpha}(\alpha-1)\right)E\left[\left(\frac{P}{P_U}\right)^{\alpha-1}\left(\frac{M}{P}\right)\right]. \tag{18}$$

where the dependence of expected profits on the share of informed firms has been made explicit to stress the interrelationship between firms.

Clearly, information is acquired if, for a given fraction h, the net gain from becoming informed outweighs costs, i.e.

$$E\Pi_I(h) - c \geq E\Pi_U(h)$$

or if

$$E\Pi_U(h)[\Gamma(h) - 1] \geq c$$

where

$$\Gamma(h) = \frac{E\Pi_I(h)}{E\Pi_U(h)}$$

measures the ratio of expected profit of informed firms to that of uninformed firms. In appendix 2 it is shown that

$$\ln\Gamma(h) = (\alpha - 1)(\lambda_{0U} - \lambda_{0I}) \\ + \tfrac{1}{2}(1-\alpha)\left[(1-\alpha + h(\alpha-2)(\rho_1 h^{-1})^2 + \rho_1 h^{-1}\right]\sigma^2.$$

We further show in appendix 2 that

$$\Gamma(0) > 1$$

and

$$\Gamma(1) > 1$$

implying that information is valuable to the firm no matter whether all other firms are informed or uninformed. The question is whether information is sufficiently valuable to justify costly acquisition of information.

Define the critical cost level which just makes information acquisition worthwhile as

$$\tilde{c}(h) \equiv E\Pi_U(h)[\Gamma(h) - 1]. \tag{19}$$

Clearly, if $c \leq \tilde{c}(h)$, there is an incentive for uninformed firms to acquire information while if $c > \tilde{c}(h)$ there is no such incentive.

Our interest here is to consider Nash equilibria to the information game. Since we consider only symmetric equilibria in pure strategies, we consider the conditions under which no information acquisition ($h^* = 0$) is a Nash equilibrium as well as whether information acquisition by all firms ($h^* = 1$) is a Nash equilibrium.

No information acquisition $h^* = 0$ is a Nash equilibrium if $c > \tilde{c}(0)$, while information acquisition is a Nash equilibrium if $c \leq \tilde{c}(1)$.

The existence of equilibrium is made non-trivial by the fact that

$$\tilde{c}(0) \neq \tilde{c}(1).$$

Due to interdependencies in information acquisition, the sign of $\tilde{c}(0) - \tilde{c}(1)$ is in general ambiguous. Thus we have to consider both $\tilde{c}(0) < \tilde{c}(1)$ and $\tilde{c}(0) > \tilde{c}(1)$. It is easily verified that $\tilde{c}(1) < \tilde{c}(0)$ for α close to unity, while $\tilde{c}(1) > \tilde{c}(0)$ is possible for γ close to unity, as can be seen from

$$\tilde{c}(0)-\tilde{c}(1)=(\Gamma(0)-1)E\Pi_U(0)-(\Gamma(1)-1)E\Pi_U(1)$$

because for $\gamma \to 1$, $\ln\Gamma(1) \to 0$, and hence the second term drops out.

Case I: $\tilde{c}(0)<\tilde{c}(1)$

Consider first the case where $\tilde{c}(0)<\tilde{c}(1)$. For sufficiently small information costs, $c \leq \tilde{c}(0)$ we find that there exists an equilibrium where information is acquired by all firms, i.e. $h^*=1$. For sufficiently large information costs $c>\tilde{c}(0)$, no information is acquired by any firm, i.e. $h^*=0$. For the intermediary case where $\tilde{c}(0)<c\leq\tilde{c}(1)$, we have that both information acquisition by all firms, i.e. $h^*=1$ is an equilibrium, as well as the case where none of the firms acquires information, i.e. $h^*=0$. That is, we have two equilibria to the information acquisition game.

Case II: $\tilde{c}(0)>\tilde{c}(1)$

Consider next the case where $\tilde{c}(0)>\tilde{c}(1)$. For sufficiently small information costs, $c\leq\tilde{c}(1)$, we find that there exists an equilibrium where information is acquired by all firms, i.e. $h^*=1$. For sufficiently large information costs $c>\tilde{c}(1)$, no information is acquired by any firm, i.e. $h^*=0$. For the intermediary case where $\tilde{c}(1)<c\leq\tilde{c}(0)$, we have that neither having all firms acquiring information nor having no one acquiring information is an equilibrium. That is, we have no symmetric equilibrium to the information acquisition game.[5]

To understand the intuition underlying these results, it is useful to note that the following externalities are present in information acquisition.[6] The larger the fraction of informed firms h, the more the aggregate price level adjusts to changes in the money stock and hence an increase in the fraction of informed firms has a positive externality to uninformed firms by stabilizing real balances (M/P). More variability in the aggregate price level following from a larger fraction of firms being informed also means that the relative price for uninformed firms becomes more variable (P_j/P), i.e. a negative externality. In case I the negative externality is dominating – if all others are acquiring information, each single firm is also more inclined to do so ($\tilde{c}(1)>\tilde{c}(0)$), while in case II the positive externality is dominating – if all others are acquiring information, the individual incentive to do so is less ($\tilde{c}(1)<\tilde{c}(0)$).

The non-existence result in case II is related to the so-called information paradox of Grossman and Stiglitz (1976, 1980). If no firm is acquiring information, it is worthwhile for each single firm to incur the information costs. However, if all firms are acquiring information, it is not optimal for any single firm to acquire information. The fact that all others have acquired information therefore has a negative externality on the incentive of a firm to acquire information.

Important differences between Grossman and Stiglitz (1976, 1980) and the present analysis should be noted. In the former agents can infer information instantaneously from the prices called by the auctioneer and modify their plans accordingly, while this is not possible here since prices are preset by firms. Prices therefore serve no signalling role, but information is useful in predicting both the state of nature and the behaviour (prices) of other firms. In the Grossman and Stiglitz framework, observation of market prices is a substitute for information acquisition, and this creates a free rider problem in information acquisition which may lead to non-existence of equilibrium. A free rider problem is also present here but arises via the stabilizing role price adjustment has for aggregate demand – the more firms acquiring information the more stable is aggregate demand, and this has a positive externality for other firms. The positive externality need not be dominating, as revealed by case I, driven by the negative externality.

The preceding argument has considered the implications of variations in information costs. Precisely the same argument could be followed by assuming an invariable cost c, and then considering how the incentives to acquire information depend on the variability (σ^2) of the nominal state variable. It is easily seen that this case will yield the same qualitative implications.

Concluding remarks

The importance of information for price-setting and thus for the incentive to acquire information has been considered in the case where firms make the information acquisition decision before knowing the price set by other firms. This precludes that firms may infer information from the prices set by competing firms, and the public value of information is therefore not at the centre of the present discussion. Crucial here is the interrelationship between price decisions of firms. There is a strategic complementarity in price-setting since the price decision of a single firm is increasing in the prices set by other firms as captured by the aggregate price level. Information is thus of relevance not only for predicting the state variables but also for inferring the decisions taken by competing firms.

This interrelationship turns out to have important implications for the incentive to acquire information. It is especially interesting to note that there may be multiple equilibria or non-existence in the information acquisition game, depending on whether negative or positive externalities in information acquisition are dominating.

The multiplicity of equilibria found here in the presence of information costs resembles the multiplicity arising in menu cost models (Ball and Romer, 1991). There are two important differences. First the condition determining whether prices are adjusted in menu cost models is state-dependent; that is, whether it is optimal to incur the menu costs depends on the actual change in market conditions. In the case of information acquisition, the change in market conditions is obviously not known. The value of information is determined by the variability of market conditions. Secondly, we found that despite strategic complementarity in price-setting, there could be a strategic substitutability in information acquisition precluding the existence of a symmetric equilibrium in pure strategies.

We have not commented on the welfare consequences of nominal rigidities. Using the reasoning of Mankiw (1985) and Bénassy (1987), it can be concluded that expansionary shocks in combination with nominal rigidities are potentially welfare improving by expanding activity, the reason being that nominal rigidities in combination with positive nominal shocks mitigate the consequences of imperfections in the product market. Conversely for negative shocks. Hence, 'small' information costs may have 'large' consequences for welfare.

Adopting a sequential decision structure implies that firms may infer information from the prices set by competing firms. This raises new issues in relation to the use of and incentive to acquire information, as firms may try to affect the information competitors extract from prices (see Andersen and Hviid, 1993, 1994).

Appendix 1

To derive (7), use (6) on (5) to get

$$\lambda_{0j}=\left(\frac{\alpha}{\gamma}+1-\alpha\right)^{-1}\left[\ln\left(\frac{\alpha\gamma^{\frac{1}{\gamma}}}{\gamma(\alpha-1)}\right)+\frac{1}{2}\left[\left(1+\frac{1}{\gamma}\right)^{2}-1\right]\operatorname{var}(\ln M|I_j)\right.$$

$$+\frac{1}{2}\left(\left(\frac{\alpha-1-\gamma}{\gamma}\right)^{2}-(\alpha-2)^{2}\right)\operatorname{var}(\ln P|I_j)$$

$$\left.+\left(\left(1+\frac{1}{\gamma}\right)\left(\frac{\alpha-1-\gamma}{\gamma}\right)-(\alpha-2)\right)\operatorname{cov}(\ln M,\ln P|I_j)\right]$$

$$\lambda_1=\left(\frac{\alpha}{\gamma}+1-\alpha\right)^{-1}(\alpha-1)\left(\frac{1}{\gamma}-1\right)$$

$$\lambda_2=\left(\frac{\alpha}{\gamma}+1-\alpha\right)^{-1}\frac{1}{\gamma}.$$

Note that only λ_{0j} depends on the parameter h. For completeness we write down the equilibrium values of λ_{0I} and λ_{0U}. Using from (10) and (11) that given the information of the informed, M is not a random variable, we find

$$\lambda_{0I}=\left(\frac{\alpha}{\gamma}+1-\alpha\right)^{-1}\ln\left(\frac{\alpha\gamma^{\frac{1}{\gamma}}}{\gamma(\alpha-1)}\right)$$

$$\begin{aligned}\lambda_{0U}=\lambda_{0I}+\left(\frac{\alpha}{\gamma}+1-\alpha\right)^{-1}\Bigg[&\frac{1}{2}\left(\left(1+\frac{1}{\gamma}\right)^2-1\right)\\&+\frac{1}{2}\left(\left(\frac{\alpha-1-\gamma}{\gamma}\right)^2-(\alpha-2)^2\right)\rho_1^2\\&+\left(\left(1+\frac{1}{\gamma}\right)\left(\frac{\alpha-1-\gamma}{\gamma}\right)-(\alpha-2)\right)\rho_I\Bigg]\sigma^2.\end{aligned}$$

Appendix 2

Expected profits can be written

$$E\Pi_j=\left(1-\frac{\gamma}{\alpha}(\alpha-1)\right)E(MP^{\alpha-2}P_j^{1-\alpha})$$

implying that

$$\begin{aligned}\ln E\Pi_j=\ln\left(1-\frac{\gamma}{\alpha}(\alpha-1)\right)&+(\alpha-2)\rho_0+(1-\alpha)E(\ln P_j)\\&+\frac{1}{2}(1+\rho_1(\alpha-2))^2\sigma^2+\frac{1}{2}(1-\alpha)\mathrm{var}(\ln P_j)\\&+(1-\alpha)(1+\rho_1(\alpha-2))\mathrm{cov}(\ln M,\ln P_j)\end{aligned}$$

where it has been used that $\ln P=\rho_0+\rho_1\ln M$.

The value of information can now be expressed as

$$\begin{aligned}\ln\Gamma(h)&=\ln E\Pi_I-\ln E\Pi_U\\&=(1-\alpha)(E(\ln P_I)-E(\ln P_U))+\frac{1}{2}(1-\alpha)^2\mathrm{var}(\ln P_I)\\&\quad+(1-\alpha)(1+\rho_1(\alpha-2))\mathrm{cov}(\ln M,\ln P_I).\end{aligned}$$

Using that

$$\ln P_U=\lambda_{0U}+\lambda_1\rho_0$$

$$\ln P_I=\lambda_{0I}+\lambda_1\rho_0+(\lambda_1\rho_1+\lambda_2)\ln M$$

and

$$\lambda_2 + \lambda_1\rho_1 = \rho_1 h^{-1} \equiv \rho_3$$

we get

$$\begin{aligned}\ln\Gamma(h) &= (\alpha - 1)(\lambda_{0U} - \lambda_{0I}) \\ &\quad + \frac{1}{2}(1-\alpha)[(1-\alpha+h(\alpha-2))\rho_3^2 + 2\rho_3]\sigma^2.\end{aligned} \tag{A1}$$

Next we consider the signs of $\Gamma(0)$ and $\Gamma(1)$. Using that $\rho_3 = 1 - \lambda_1 = \lambda_2$ for $h = 0$ in (A1), we have

$$\ln\Gamma(0) = \frac{1}{2}\frac{\alpha-1}{\frac{\alpha}{\gamma}+1-\alpha}\left(\left(1+\frac{1}{\gamma}\right)^2 - 1\right)\sigma^2 + \frac{1}{2}(1-\alpha)((1-\alpha)\lambda_2^2 + 2\lambda_2)\sigma^2.$$

Hence,

$$\ln\Gamma(0) = \frac{1}{2}(\alpha-1)\frac{\frac{\alpha}{\gamma^3}\sigma^2}{\left(\frac{\alpha}{\gamma}+1-\alpha\right)^2} > 0. \tag{A2}$$

Using that $\rho_1 = 1$ and $\rho_3 = 1$ for $h = 1$, it follows from (A1) that

$$\begin{aligned}\ln\Gamma(1) &= (\alpha-1)\frac{1}{2}\left(\frac{\alpha}{\gamma}+\alpha-1\right)^{-1}\left[\left(1+\frac{1}{\gamma}\right)^2 - 1 + \left(\frac{\alpha-1-\gamma}{\gamma}\right)^2\right. \\ &\qquad \left. -(\alpha-2)^2 + 2\left(1+\frac{1}{\gamma}\right)\left(\frac{\alpha-1-\gamma}{\gamma}\right) - 2(\alpha-2) + \frac{1}{2}(1-\alpha)\right]\sigma^2 \\ &= \frac{1}{2}(\alpha-1)\left(\frac{\alpha}{\gamma}-\alpha\right)\sigma^2 > 0.\end{aligned} \tag{A3}$$

Comparing (A2) and (A3), we note that for γ close to unity, $\ln\Gamma(0) > \ln\Gamma(1)$. Further the limit of $\ln\Gamma(h)$ as γ approaches unity is zero for $h = 1$, and positive for $h = 0$.

Notes

This chapter was initiated during the University of Warwick Macroeconomics Workshop (July 1993). An earlier version was presented at CentER, Tilburg and the conference 'Alternative Approaches to Macroeconomics' Madrid. Comments by Neil Rankin, Franck Portier, Christian Schultz and Jean-Pascal Bénassy are gratefully acknowledged.

1. See Andersen and Hviid (1993, 1994) for an analysis of these issues in the context of duopoly markets with sequential price-setting.
2. See Dixit and Stiglitz (1977) and Blanchard and Kiyotaki (1989) for a derivation from a CES utility function.
3. This assumption may be justified by thinking of this as a repeated event where firms from past observations can infer the fraction of informed firms, or that the agency selling information uses h in its marketing strategy.
4. The cost could be interpreted as the subscription fee for obtaining a forecast from an agency selling business cycle information.
5. It is easily verified that there exists an asymmetric equilibrium with a fraction of firms acquiring information. However, this equilibrium is unstable.
6. We have that $\partial^2\pi/\partial(P_j/P)^2 < 0$ from the second order condition assumed to be fulfilled, and it is easily verified that $\partial^2\pi/\partial(M/P)^2 < 0$. Hence, the firm dislikes variation in both its relative price and real demands.

15 Expectation calculation, hyperinflation and currency collapse

George W. Evans and Garey Ramey

Introduction

Economists widely agree that inflation expectations play a crucial role in hyperinflationary episodes. There is little agreement, however, on how these expectations are formed, and two competing approaches to the question of expectation formation have emerged. On one hand, the *rational expectations* approach posits that agents base their expectations on full knowledge of the structure and equilibrium of the economy. On the other, the *passive adaptive learning* approach endows agents with simple learning rules that map past inflation observations to future forecasts, with the form of the mapping incorporating little direct information about the structure of the economy.

Each of these approaches faces difficulties from both conceptual and empirical standpoints. The rational expectations approach (for example, Sargent and Wallace, 1987) gives no account of how agents come to acquire rational expectations, and in particular there is no specification of what agents would believe were they to observe a history that violated rational expectations. For models with multiple rational expectations solutions, the approach faces the problem of selecting among the (possibly infinite) set of equilibrium paths. Empirically, it is difficult to reconcile rational expectations paths in the paradigm hyperinflation model with the historical record, which shows hyperinflations to have involved increasing and accelerating inflation rates, terminating in major fiscal restructurings or 'currency collapse'.[1] Finally, there are problems in reconciling the rational expectations approach with experimental results. For example, rational expectations predicts that an announced policy change should initiate a reaction at the time of announcement, but laboratory tests of hyperinflation models, reported in Marimon and Sunder (1988, 1991), have found that significant reactions occurred only in the periods immediately preceding changes.

Passive adaptive learning rules, in contrast, posit extremely limited analytic capabilities on the part of agents, far below what one might reasonably expect of real world agents. The least squares learning algorithm studied by Marcet and Sargent (1989), for example, requires agents to form forecasts based on economic models that are misspecified (except asymptotically), and agents are not allowed to respecify their models even in response to known structural changes, such as announced fiscal policy adjustments. Experimental subjects in the Marimon–Sunder laboratory tests reacted more rapidly to structural changes than would be consistent with least squares learning.[2] Moreover, least squares learning algorithms and other simple adaptive schemes such as in Bruno (1989) must converge to stationary rational expectations equilibria (REE), but some of Marimon and Sunder's experimental economies were found to settle at outcomes that were not near any such equilibria.

In this chapter we develop a new approach to modelling expectation formation in hyperinflationary settings that provides a resolution to these difficulties. Our approach, introduced in Evans and Ramey (1992a), posits that a complete specification of the economic environment must include an *expectation formation technology* together with *preferences* over forecast errors. Like the rational expectations approach, we endow our agents with correct structural models of the economy, and in particular the models are adjusted to incorporate structural changes as they occur. In contrast to rational expectations, however, agents do not know in advance the equilibrium path of inflation, but rather must use their models to calculate inflation forecasts.

We assume that agents calculate expectations by means of an iterative algorithm that derives from their knowledge of the economic structure. The algorithm is specified as part of the expectation formation technology. Agents also have bounded abilities to calculate, in that using the calculation algorithm imposes time and resource costs. We propose an *optimality criterion* that governs agents' expectation revision decisions: agents balance the costs of calculation against the benefits of improved forecasts, subject to the calculation decisions made by other agents and the informational restrictions of the algorithm. Agents make optimal calculation decisions in each period, and this gives rise to the equilibrium paths of our model. In this way, inflation dynamics emerge from the agents' capabilities and incentives to calculate expectations.

Our approach yields an empirically appealing theory of hyperinflationary episodes. In the case of low deficit levels, convergence toward steady-state inflation rates can be quite rapid compared, for example, to least squares learning. Equilibrium paths characterized by accelerating inflation and terminating in currency collapse arise naturally in our framework, when initial inflation expectations or the deficit level are sufficiently high. Such paths give equilibria even in the high deficit case where no stationary REE exist; here the rapid acceleration of inflation gives agents a very strong incentive to calculate, and this provides impetus for continued acceleration. These equilibrium paths do not unravel from the end, as they would under rational expectations, because agents do not respond to the possibility of currency collapse until the hyperinflationary episode has advanced far enough to allow them to calculate it.

A change in economic structure leads to rapid adjustment in equilibrium expectations beginning in the period immediately prior to the change, since at this point agents incorporate the change into their calculation models. The anticipation of a change occurring more than one period in the future will not affect agents' behaviour, however, because of calculation costs together with the limited forecast horizon of the calculation algorithm. This provides an explanation for the experimental findings discussed above.

When resource costs of calculation are non-negligible, our equilibrium paths can settle at 'near-rational' steady-states at which the gains from expectation revision, in the form of reduced forecast errors, are small relative to the costs. These near-rational steady-states need not lie near any stationary REE, although they do lie at points where realized inflation is closest to forecast levels. Interestingly, this is precisely where Marimon and Sunder's experimental economies settled in the cases where no stationary REE existed; such outcomes could not occur under either rational expectations or least squares learning, but they are entirely consistent with our theory.

This chapter significantly extends the applicability of the framework developed in our earlier study. We previously analysed a 'natural rate' model of monetary policy possessing a number of simplifying features: (1) the model had a linear structure; (2) agents formed expectations of only current variables; and (3) the model had unique REE. Further, we imposed the restriction that in each period agents should choose from among only two possible calculation decisions. Here we consider a deficit finance model in which money demand is a non-linear function of expected inflation, inflation expectations incorporate current and future price levels, and there are multiple REE. We also allow for an arbitrarily large finite number of possible calculation decisions.

The deficit finance model is presented in the second section. The third section lays out the structure of the calculation model. An optimality criterion for calculation decisions is developed in the fourth section. The fifth section gives examples and discusses implications for policy, while the sixth section concludes.

Hyperinflation model

We adopt a simple version of the hyperinflation model studied by Cagan (1956), Fischer (1984, 1987), Sargent and Wallace (1987), Bruno (1989), Marcet and Sargent (1989), Bental and Eckstein (1990), and others. In this model, the government's ability to finance deficits through seigniorage depends on the willingness of private citizens to hold money balances.

Let H_t denote the stock of high powered money in period t, and P_t the price of consumption goods in terms of money. Let D_t be the real government budget deficit in period t, to be financed through seigniorage. The government budget constraint is given by:

$$\frac{H_t - H_{t-1}}{P_t} = D_t. \tag{1}$$

We assume that money is demanded by a continuum of asset-holding agents, having unit mass. Agent ω's demand for real balances in period t is:

$$\frac{H_t(\omega)}{P_t} = K\hat{\pi}_{t+1}(\omega)^{-a}$$

where $H_t(\omega)$ is demand for nominal balances, $\hat{\pi}_{t+1}(\omega)$ gives the expected rate of inflation between period t and period $t+1$, i.e. the agent's forecast of P_{t+1}/P_t, and K and a are positive parameters. For simplicity we will assume that agents have homogeneous expectations, so that total demand for real balances is:

$$\frac{H_t}{P_t} = K\hat{\pi}_{t+1}^{-a}. \tag{2}$$

Thus agents will be more willing to hold money if they expect a lower one-period ahead inflation rate.

Let the realized inflation rate be denoted by $\pi_t \equiv P_t/P_{t-1}$.[3] For low levels of D_t, (1) and (2) may be solved for the inflation rate that equates supply and demand for money:

$$\pi_t = \frac{K\hat{\pi}_t^{-a}}{K\hat{\pi}_{t+1}^{-a} - D_t} \equiv T(\hat{\pi}_t, \hat{\pi}_{t+1}). \tag{3}$$

The function T, which gives realized inflation conditional on inflation expectations and the deficit level, is called the *T-mapping*; note that we have suppressed D_t in the definition of T.

When D_t or $\hat{\pi}_{t+1}$ is too large, the supply of real balances in period t exceeds demand at every inflation rate, and the government defaults on its current obligations; this situation is called *currency collapse*. Currency collapse occurs in period t if and only if:

$$\hat{\pi}_{t+1} \geq \left(\frac{D_t}{K}\right)^{-1/a} \equiv \bar{\pi}_{t+1}.$$

We assume that P_t becomes infinite in this case, and any balances that are held in this period have zero value in all future periods; thus our model ends following the occurrence of collapse.[4] When $\hat{\pi}_{t+1} \geq \bar{\pi}_{t+1}$, the T-mapping sets realized period-t inflation at infinity. Thus (3) gives realized inflation when $\hat{\pi}_{t+1} < \bar{\pi}_{t+1}$, while realized inflation is infinite for $\hat{\pi}_{t+1} \geq \bar{\pi}_{t+1}$.

A *rational expectations equilibrium* (REE) is given by a path $\pi_1, \pi_2, \ldots$ of realized inflation rates such that agents base their expectations on full knowledge of the equilibrium path, i.e. $\hat{\pi}_t = \pi_t$ for all t. Thus REE paths are generated by the difference equation:

$$\pi_t = T(\pi_t, \pi_{t+1}). \tag{4}$$

Combining this with an arbitrary initial rate π_1 gives a continuum of REE. We can also define a 'collapse' REE by specifying that money has no value in any period.

REE dynamics are illustrated in figure 15.1, where we put $D_t = D$ and $\bar{\pi}_{t+1} = \bar{\pi}$ for all t. The curves in the diagrams depict solutions to (4). Figure 15.1a represents the *low deficit case* in which D satisfies:

$$\frac{D}{K} < \frac{a^a}{(a+1)^{a+1}} \equiv \frac{\bar{D}}{K}.$$

Inflation trajectories are shown along the π_t axis. Here the inflation rates π_L and π_U give steady-state REE, and only the high steady-state π_U is stable under the REE dynamics. Note that currency collapse at a future date can never arise as the outcome of an REE path. This follows from a *backward unravelling* argument: if collapse occurs in period t, then the T-mapping gives $\pi_t = \infty$, and so collapse occurs in $t-1$, etc. Figure 15.1b depicts the *high deficit case* of $D > \bar{D}$, in which no inflationary steady-state exists. In this case all REE paths (other than currency collapse in the initial period) involve hyperdeflation, with real balances becoming infinitely large along the equilibrium path. In general, all of the REE (other than immediate collapse) involve either decreasing inflation rates or increasing inflation rates that decelerate to a limiting steady-state.[5]

Expectation calculation model

Calculation technology

The first ingredient in our expectation calculation framework is a *technology* that is used by the agents to calculate inflation expectations at any given point in time. We will assume that, in each period, agents are endowed with: (1) an *information set* formed from the past history of the economy, which includes an agent's own past calculations along with other variables;

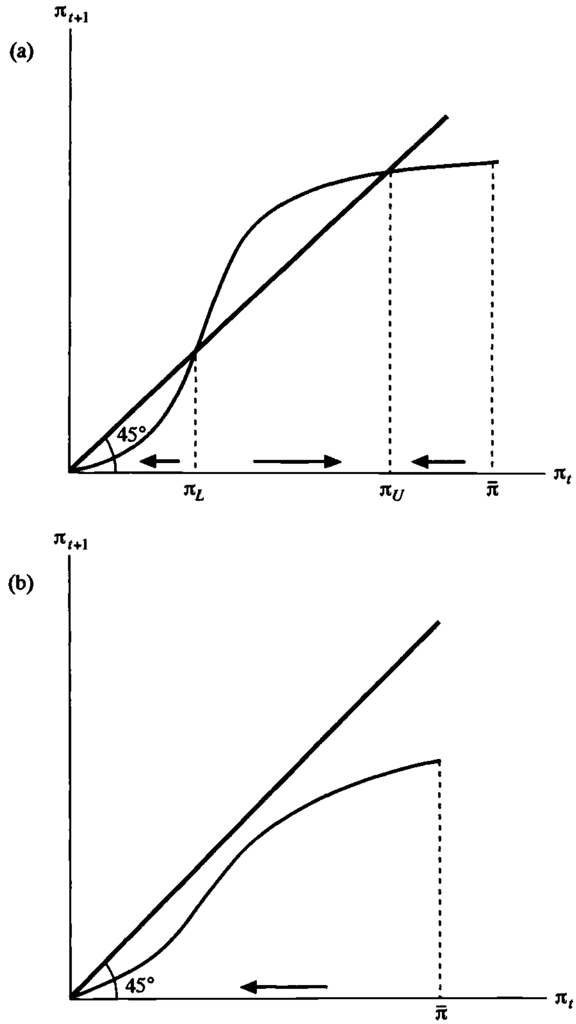


Figure 15.1 REE dynamics
(a) Low-D case (b) High-D case

(2) a *true model* of the economy, which expresses the realized inflation path conditional on the actual path of deficit levels together with any given path of expected inflation, i.e. the T-mapping; and (3) a *calculation algorithm* that combines history with the T-mapping to produce a calculation of the inflation rate for the following period, on which desired money holdings can be based. These three components together make up the calculation technology.

We will consider a class of calculation algorithms that are built on iteration of the T-mapping. The basic idea is that an agent forms a hypothesis as to the path of future inflation rates, and he checks this hypothesis against the prediction of his economic model to see if it is consistent. If the hypothesis is inconsistent with the model, then he revises the hypothesis in the direction indicated by the model. In the next iteration the revised hypothesis is checked, etc. Thus the agent calculates expectations by revising his inflation hypothesis according to the predictions of a true model of the economy.

In this chapter we will focus on algorithms in which the true model is used to calculate one-period ahead inflation expectations, i.e. use of the T-mapping is restricted to a one-period forecast horizon. Revisions of expectations beyond the one-period horizon, which do not directly enter the money demand function, will be tied to the one-period ahead calculations in a manner to be specified below. Let the inputs and outputs of the algorithm be parameterized by β and γ, where:

$$\beta=\hat{\pi}_{t+1}, \qquad \gamma\beta=\hat{\pi}_{t+2}.$$

Thus β, called the *base rate*, parameterizes an agent's expectation of the period $t+1$ inflation rate, and γ, called the *growth coefficient*, gives the proportional change in the inflation rate that the agent expects between periods $t+1$ and $t+2$. We require β and γ to be strictly positive.

Calculation in a given period proceeds as follows. From the observed past history, an agent forms initial inputs β^0,γ^0. The agent then calculates the inflation rate that would actually be realized if expectations were formed according to β^0 and γ^0:

$$\beta^1=T(\beta^0,\gamma^0\beta^0).$$

If $\beta^1>\beta^0$, the agent takes his initial estimate of the base rate to be too low, and he revises it upward, while $\beta^1<\beta^0$ leads to downward revision. In particular, we suppose that the base rate is revised to β^1. If $\beta^1=\beta^0$, then the agent's inputs are consistent with the economic model, and no revisions are needed.

The growth coefficient γ is revised as well, according to the following general specification:

$$\gamma^1=\psi(\beta^1-\beta^0|\beta^0,\gamma^0)$$

where we assume $\psi(0|\beta^0,\gamma^0)=0$. Observe that the calculated value of one-period ahead expectations enters into the revision of two-period ahead expectations; in particular, if $\beta^1=\beta^0$, so that β^0 is consistent with the T-mapping, then γ^0 is not revised.

A second round of calculation takes inputs β^1,γ^1 to produce outputs β^2,γ^2 in an analogous manner, etc. In general, k iterations of the algorithm produce outputs β^k, γ^k that are determined recursively from β^0, γ^0 according to:

$$\beta^k = \beta^{k-1} + x^k \tag{5a}$$

$$\gamma^k = \gamma^{k-1} + \psi(x^k | \beta^{k-1}, \gamma^{k-1}) \tag{5b}$$

where $x^k = T(\beta^{k-1}, \gamma^{k-1}\beta^{k-1}) - \beta^{k-1}$. (5a) and (5b) are valid as long as $\beta^{k-1} < \bar{\pi}_{t+1}$, so that collapse would not occur in the current period. If $\beta^{k-1} \geq \bar{\pi}_{t+1}$, then on the kth iteration the agent calculates that collapse occurs in the current period, in which case he ceases calculation and his desired money holdings are zero. Note that $\gamma^{k-1}\beta^{k-1} \geq \bar{\pi}_{t+2}$ implies that $\beta^k = \infty$ is calculated in the next iteration, and again desired money holdings are zero.

The following specialization of ψ illustrates the behaviour of our calculation algorithm:

$$\psi(x | \beta, \gamma) = \frac{\lambda \gamma x}{\beta} \tag{6}$$

where $\lambda \in [0,1]$.[6] Note that in this example, ψ has the same sign as x, which implies the algorithm revises γ in the same direction as β; we refer to this as the *acceleration property*, since the path of expectations is accelerated in the direction of the base rate revision.

Figure 15.2 illustrates trajectories of the calculation algorithm for this example.[7] Figure 15.2a depicts a low deficit case, while Figure 15.2b considers a high deficit case. The curves γ^S, defined by $\beta \equiv T(\beta, \gamma^S(\beta)\beta)$, give the fixed points of the calculation algorithm. These in turn correspond to the first two values of inflation (π_{t+1}, π_{t+2}) along some REE path; thus the fixed points of the algorithm give expectations that are fully rational relative to the forecast horizon of the algorithm.[8] It is easily verified that $T(\beta, \gamma\beta)$ is strictly increasing in both β and γ. Thus for points (β, γ) above γ^S we have $\beta < T(\beta, \gamma\beta)$, and the algorithm adjusts β and γ upward, as indicated. Points below the curve have $\beta > T(\beta, \gamma\beta)$, and downward adjustments are made. Further, (β^k, γ^k) must lie on the same side of the curve as does $(\beta^{k-1}, \gamma^{k-1})$, i.e. outputs cannot cross over the curve. It follows that all trajectories of the algorithm in this example are monotonic. The paths *AB* and *CD* give trajectories that converge to points on the γ^S curve.

Figure 15.2 also shows regions labelled 'collapse.' If on the kth iteration we have $\beta^k \geq \bar{\pi}_{t+1}$, then the implied inflation expectation $\hat{\pi}_{t+1} = \beta^k$ implies currency collapse in period t. If the agent iterates one more time from this point, then the algorithm reveals that collapse obtains. Similarly, if $\beta^k \gamma^k > \bar{\pi}_{t+2}$, then another iteration indicates that collapse will occur in the next period. Thus the collapse region gives the set of inputs such that one

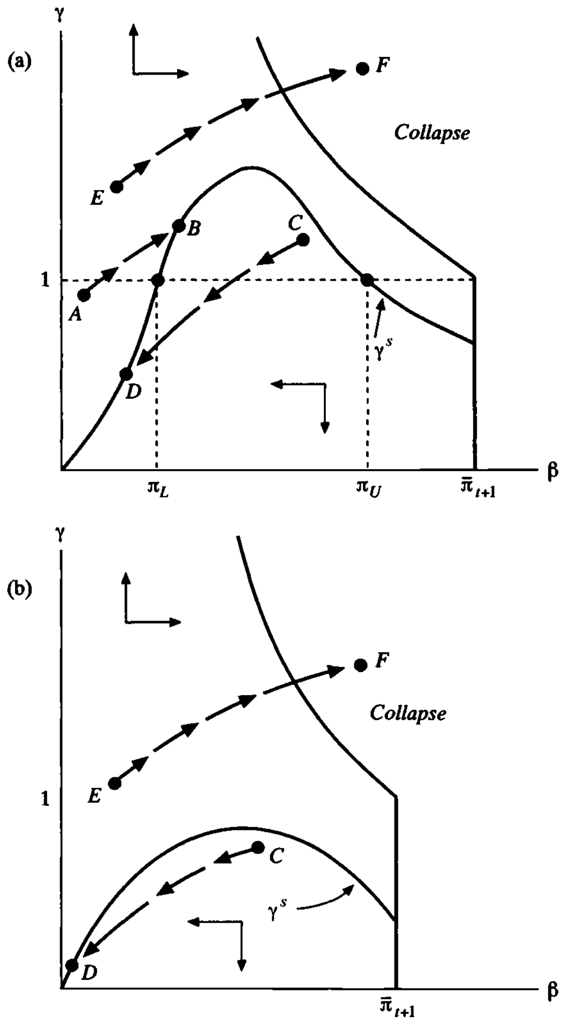


Figure 15.2 Examples of calculation trajectories
(a) $D_t < \bar{D}$ (b) $D_t > \bar{D}$

more calculation indicates collapse in the current or subsequent period. The paths EF depict trajectories that terminate in collapse.

Our algorithm has the important limitation that it effectively looks ahead only one period, taking as inputs one- and two-period ahead inflation forecasts. The algorithm could be modified by extending the forecast horizon to $H > 1$ periods: the T-mapping for periods $t+1, \ldots, t+H$ would be used to calculate revised values of $\hat{\pi}_{t+1}, \ldots, \hat{\pi}_{t+H}$, based on inputs $\hat{\pi}_{t+1}, \ldots, \hat{\pi}_{t+H}, \hat{\pi}_{t+H+1}$; some function other than the T-mapping (the analogue of ψ) would be needed to close the algorithm by computing a revised value of $\hat{\pi}_{t+H+1}$. While the extended algorithm would have the advantage of incorporating information about more distant future structural changes, it would also impose added calculation costs, since more computations would be required to obtain each revision of the desired one-period ahead forecast. A balance must thus be struck between the forecast horizon and the computational requirements of the algorithm; in this chapter we take the approach of setting $H=1$ and allowing multiple iterations of the algorithm within a given period.[9]

Default paths, calculation decision rules and calculation paths

Thus far we have specified agents' expectations given that they begin a period with initial inputs β^0, γ^0 and perform k iterations of the algorithm. To complete the specification, we must indicate (1) how initial inputs are determined from the information that agents possess at the start of a period; and (2) how agents determine the number of iterations to perform. While it is possible in principle that initial inputs might depend on any of the variables in agents' information sets, we make the simplifying assumption that initial inputs depend only on the final outputs that were calculated in the preceding period.[10] Let $(\beta_{t-1}^C, \gamma_{t-1}^C)$ denote the outputs from the final iteration of the calculation algorithm in period $t-1$, and let (β_t^0, γ_t^0) denote the initial inputs used in period t. Initial inputs are generated from outputs according to the following *input functions*:

$$\beta_t^0 = \gamma_{t-1}^C \beta_{t-1}^C \tag{7a}$$

$$\gamma_t^0 = \varphi(\beta_{t-1}^C, \gamma_{t-1}^C). \tag{7b}$$

The specification (7a) reflects intertemporal consistency of inflation expectations: $\gamma_{t-1}^C \beta_{t-1}^C$ denotes the final period $t-1$ forecast of $\hat{\pi}_{t+1}$, and it is reasonable to suppose that this gives the initial forecast of $\hat{\pi}_{t+1}$ in period t. (7b) gives a general specification that is used to generate the initial conjecture of $\hat{\pi}_{t+2}$.[11]

The input functions with which agents are endowed play a fundamental role in our calculation framework, as they determine an agent's hypothesis at any point in time about the future path of inflation. To see this, note that for each pair of period 1 initial inputs, (7) may be used to define recursively a path of β and γ through time. This path gives the expectations that an agent would hold if zero iterations of the calculation algorithm were performed in each period; we call this the *default path*. The default path thus fixes the

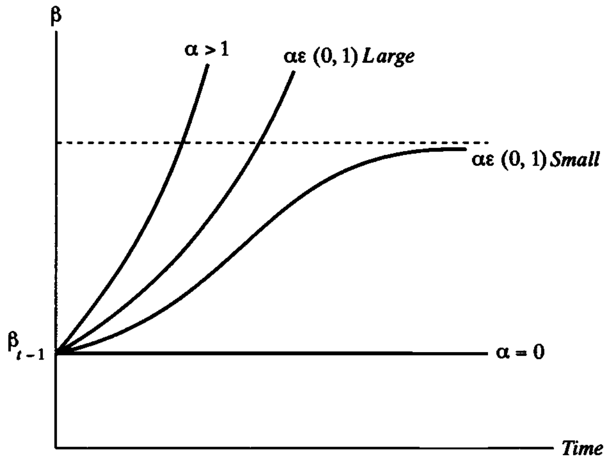


Figure 15.3 Default paths for $\gamma_0 > 1$

hypothesis about future inflation dynamics that agents hold prior to making calculations.

Possibilities for default paths are illustrated by the following example:

$$\varphi(\gamma_{t-1}^C) = \alpha\gamma_{t-1}^C + (1-\alpha) \tag{8}$$

where $\alpha \geq 0$. Here the default path is determined by $\beta_t = \gamma_{t-1}\beta_{t-1}$ and $\gamma_t = \alpha\gamma_{t-1} + (1-\alpha)$ for given (β_0, γ_0). Figure 15.3 illustrates possible default paths for $\gamma_{t-1} > 1$, under various values of α. When $\alpha = 0$, the agent believes, in the absence of further calculation, that the future path of inflation will be a steady-state. If $\alpha = 1$, in contrast, then the last calculated growth coefficient is projected into the infinite future, and prices are conjectured to accelerate upward without limit. For $\alpha \in (0,1)$, price acceleration is conjectured to decrease over time, and the default path converges to a finite limit if α or γ_0 is not too large. Note that default paths are certain to terminate in collapse if α is sufficiently close to unity. More generally, a wide range of hypotheses about inflation dynamics may be admitted through appropriate choices of φ.

Our model of expectation formation may now be described as an interaction between the calculation algorithm and the default paths. In any given period, the calculation algorithm operates on the current default path by checking it against the T-mapping and revising β and γ accordingly; this revision generates another default path that serves as the revised hypothesis. For example, the default path may indicate that inflation will rise for a time, then settle at a steady-state. If the calculation algorithm revises β and γ upward, then the revised default path may have inflation rising at a faster rate in the short run, and then settling at a higher steady-state. The agent may continue to revise his hypothesis by performing further iterations, and the total number of iterations he performs indicates how intensely he labours to form expectations.

As the final ingredient of our calculation framework, we must specify how many iterations of the algorithm an agent performs as a function of his information at the start of a period. We will assume that the number of iterations is determined by the initial inputs for a period, according to the function $k^C(\beta^0, \gamma^0)$ that maps initial inputs into the integers $0, 1, \ldots, n$; k^C is called the *calculation decision rule* (CDR).[12] The CDR represents the decision theoretic component of our framework, and it is possible to choose it from among a wide range of possibilities. For example, a CDR might be chosen to satisfy some notion of individual optimality, or a simple rule of thumb might be specified. In the next section we will investigate one possible choice for the CDR.

The model is completely determined once a CDR is specified, and we may define solution paths as follows. The sequence $\{(\beta_t^C, \gamma_t^C)\}_{t=1}^{\infty}$ is a *calculation path* (CP), relative to initial inputs (β_1^0, γ_1^0), if for all t: (1) (β_t^C, γ_t^C) is determined by performing $k^C(\beta_t^0, \gamma_t^0)$ iterations from the initial inputs (β_t^0, γ_t^0), and (2) (β_t^0, γ_t^0) is formed from $(\beta_{t-1}^C, \gamma_{t-1}^C)$ using the input functions (7). The realized inflation path associated with a CP is determined by $\pi_t = T(\beta_t^C, \beta_{t+1}^C)$.

Optimal calculation decision rules

Optimality criterion

We have defined solutions to the model relative to a given specification of the CDR, and further analysis now hinges on how the CDR is specified. It seems plausible to suppose that the CDR would incorporate some balancing by the agents of the costs and benefits of calculation. Agents' introspective evaluation of their calculation problems may lead them to compare costs and benefits; for example, agents might possess higher order calculation technologies that allow them to compute a CDR that minimizes forecast errors net of calculation costs. Alternatively, agents may be led to an optimal CDR through some adaptive process. We will not investigate any particular selection process in this chapter, however; instead we will consider a simple optimality criterion that may be interpreted as the limiting outcome of some unspecified selection process.

Our optimality criterion begins with a specification of the costs and benefits of calculation. In the hyperinflation model that we are considering, the benefits of calculation are associated with reductions in the losses from forecast errors. The following *loss function* expresses the losses from forecast errors made in period t, where errors are reflected by the distance between actual and desired real balances:

$$h(K|\hat{\pi}_{t+1}^{-a} - \pi_{t+1}^{-a}|). \tag{9}$$

Here $\hat{\pi}_{t+1}$ is the agent's inflation expectation determined by the final iteration of the algorithm in period t. We assume that h satisfies $h(0)=0$, $h'>0$ and $h''>0$.

Calculation is assumed to be costly in terms of both time and resources. The time cost of calculation is reflected by the assumption that no more than n calculations may be made in any period. Further, the resource cost of making $k \leq n$ calculations, expressed in the same units as the loss function h, is given by c_k. We assume that $c_0=0$ and $c_k > c_{k-1}$ for $k \leq n$.

A CDR is optimal from an agent's point of view if it minimizes the losses from forecast errors net of calculation costs, given the calculation decisions made by other agents. For simplicity, we will assume that agents are concerned only with current period net losses, i.e. future losses are fully discounted.[13] Further, we assume that agents take account only of the current period calculations of other agents, and not the calculations that other agents make in future periods. The latter restriction is needed to preserve consistency with agents' cognitive limitations as expressed by the calculation algorithm: in period t, the algorithm incorporates the T-mapping for period $t+1$ only, but future calculations depend on the T-mapping for periods $t+2$ and beyond. Thus any procedure that agents might use in period t to check the future calculations would have to employ structural information that is unavailable according to the original algorithm.

With these restrictions, our optimality criterion takes the following form. Let $\hat{\pi}_{t+1}$ be calculated from the initial inputs β_t^0, γ_t^0 for each possible number of iterations $k=0,1,\ldots,n$. The net loss associated with k iterations is given by (9) minus c_k, where π_{t+1} in (9) is replaced by $T(\beta_t^C, \gamma_t^C \beta_t^C)$ (recall that β_t^C and γ_t^C are the final period t outputs along the CP). Thus the calculated inflation rates are compared to the inflation rate that would be realized given the inflation expectations $\hat{\pi}_{t+1} = \beta_t^C$ and $\hat{\pi}_{t+2} = \gamma_t^C \beta_t^C$ that other agents hold in period t; this constitutes the best forecast of π_{t+1} that can be formed based on other agents' period t calculations. According to this criterion, the best forecast would be obtained by performing one more iteration than do the other agents, i.e. by staying 'one step ahead of the pack'. Expectation formation can be interpreted as a 'calculation race', in which agents strive to get ahead of their rivals, but are unable to do so since agents calculate simultaneously and have equal calculation abilities. An optimal CDR chooses k^C to be a Nash equilibrium of this calculation race.

More formally, our optimality criterion is expressed as follows. We suppress the time subscripts, since the criterion depends only on the initial inputs and not on the particular period. Let the best response function $k^M(k)$ be defined by:

$$k^M(k) \equiv \arg\min_{k'} \{h(K|(\beta^{k'})^{-a} - T(\beta^k, \gamma^k \beta^k)^{-a}|) + c_{k'}\}.$$

k^C is said to be *optimal* if $k^C(\beta^0, \gamma^0) \in k^M(k^C(\beta^0, \gamma^0))$ for all β^0, γ^0. Existence of an optimal CDR is considered in Proposition 1, whose proof is given in the appendix:

Proposition 1: If $x\psi(x|\beta, \gamma) \geq 0$ for all x, β, γ, then there exists an optimal CDR.

Proposition 1 establishes existence of optimal CDRs under the assumption that γ is never revised in the opposite direction as β, which constitutes a weak version of the acceleration property discussed on p.314. This assumption assures that calculation trajectories possess monotonicity of the form exhibited in figure 15.2, and monotonicity is needed in turn to establish existence of fixed points of k^M.

Optimal CDRs for a particular specialization of the model are depicted in figure 15.4.[14] The maximum number of iterations per period in this example is $n=3$, and the cost per calculation is constant at $c_k - c_{k-1} = 1$. Figure 15.4a gives the CDR that minimizes the number of iterations among all optimal CDRs, while Figure 15.4b gives the CDR that maximizes the number of iterations; the structure that underlies the two figures is identical. This illustrates the scope for multiplicity of optimal CDRs, which emerges as a consequence of the strategic complementarities that are inherent in the calculation race. Note that for sufficiently high β^0 and γ^0, the agents calculate $\hat{\pi}_{t+1} \geq \bar{\pi}_{t+1}$, and collapse then occurs in period t.

As mentioned above, our notion of optimal CDRs sidesteps the issue of how agents actually determine the optimal CDR, and instead moves directly to a limiting selection. In essence, we have replaced the initial rational expectations hypothesis with a rationality hypothesis that is imposed after one level of calculation. While this procedure does not create any difficulties from the purely logical standpoint, as argued by Lipman (1991), it does involve a certain philosophical inconsistency. We respond to the latter criticism in two ways. First, our optimal CDR is meant to capture the intuitive and empirically meaningful notion that agents balance costs and benefits in making their calculation decisions. The reasonableness of our specification relative to the alternative hypotheses of rational expectations and passive adaptation becomes a testable issue, and we argue

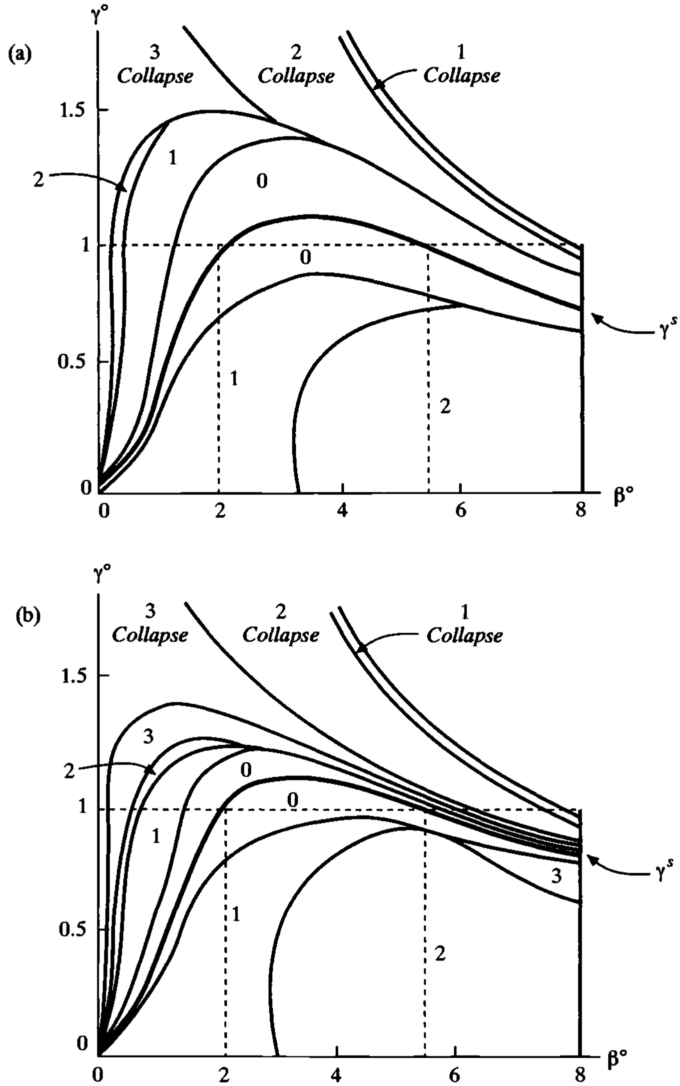


Figure 15.4 Optimal CDRs
(a) Minimum calculation (b) Maximum calculation

below that our theory receives strong empirical support.[15] Second, the optimal CDR may be regarded as a tractable first step in the modelling of calculation behaviour, and the more complex problem of explicitly modelling the CDR selection process is appropriately left for future work.[16]

Behaviour of calculation paths under optimal calculation decision rules

In this subsection we consider the long-run behaviour of CPs that are generated by optimal CDRs. The main result is that under mild restrictions on the calculation algorithm and input functions, when calculation costs are small CPs can settle only at points that are either close to steady-state REE or are in the collapse region. We also consider the stability of the high inflation and low inflation steady-state REE. Readers may ignore this subsection without loss of continuity.

Some new terminology is needed to describe the asymptotic behaviour of CPs. Suppose $D_t = D$ and $\bar{\pi}_{t+1} = \bar{\pi}$ for all t. For a given neighbourhood N of the (β, γ) plane, we say that a CP *terminates in N* if there exists a time T such that $(\beta_t^C, \gamma_t^C) \in N$ for all $t \geq T$. A point (β, γ) with $\beta < \bar{\pi}$ is a *terminal point* if for any neighbourhood N containing (β, γ), there exists a CP that terminates in N. The collapse outcome is said to be a terminal point if there exists a *CP* that terminates in collapse. Thus terminal points give the outcomes at which CPs can settle.

Proposition 2 provides a characterization of the set of terminal points, under weak restrictions on the input function φ. The proofs of this and the following propositions are given in Evans and Ramey (1992b).

Proposition 2: Let ψ and φ be C^1 functions, and suppose:

(a) $x\psi(x|\beta, \gamma) \geq 0$ for all x, β, γ;

(b) $\varphi(\beta, 1) = 1$ for all β;

(c) $0 \leq \varphi_2(\beta, \gamma) < 1$ for all β, γ.[17]

Then for $c_1 < h(D)$, the set of terminal points is characterized as follows.

(I) If $D < \bar{D}$, then the terminal points are the collapse outcome together with:

$$\{(\beta, \gamma) | \beta \in N_L \cup N_U, \gamma = 1\}$$

where N_L and N_U are closed subintervals of $(0, \bar{\pi})$, possibly overlapping, with $\pi_L \in \text{int}\{N_L\}$, $\pi_U \in \text{int}\{N_U\}$. The intervals become strictly smaller as c_1 declines, and:

$$\lim_{c_1 \to 0} N_L = \{\pi_L\}, \lim_{c_1 \to 0} N_U = \{\pi_U\}.$$

(II) If $D \geq \bar{D}$, then the terminal points are the collapse outcome together with:

$$\{(\beta, \gamma) | \beta \in N_M, \gamma = 1\}$$

where for c_1 sufficiently close to $h(D)$, N_M is a closed subinterval of $(0, \bar{\pi})$. N_M becomes strictly smaller as c_1 declines. If $D = \bar{D}$, then N_M

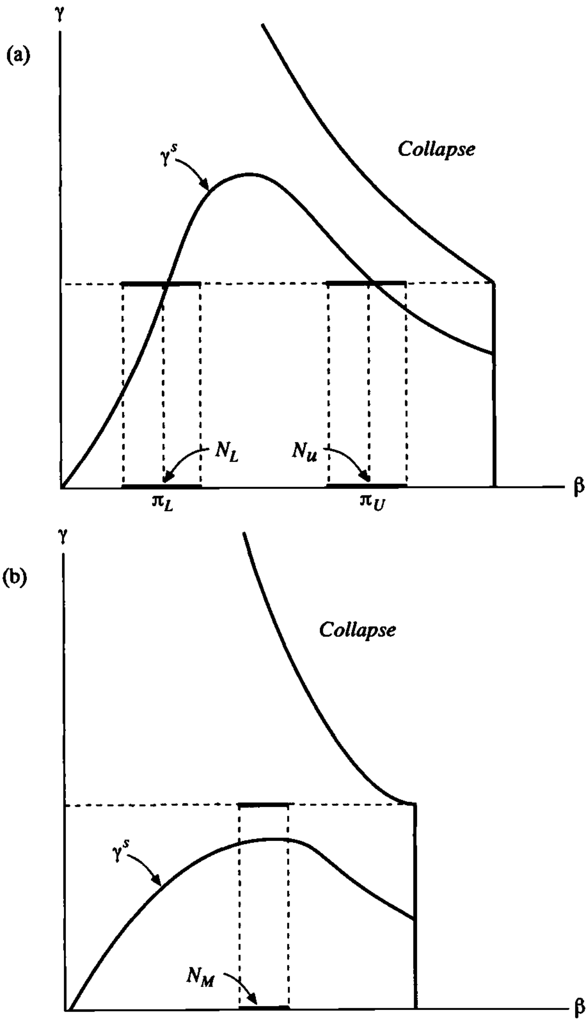


Figure 15.5 Set of terminal points
(a) $D < \bar{D}$ (b) $D > \bar{D}$

shrinks to a single point as c_1 approaches zero, while for $D > \bar{D}$, N_M is empty for c_1 sufficiently small.[18]

The set of terminal points characterized by Proposition 2 is illustrated in figure 15.5. Observe that given our conditions on φ, all non-collapse terminal points must have $\gamma = 1$; in particular, condition (b) ensures that if agents believed they were in a steady-state in the preceding period, then the default path continues to reflect this belief in the current period. It follows that agents must recognize that they are in a steady-state if the economy comes to rest at some point. CPs can settle only at points that are approximately REE when the cost of a single calculation is low, and the set of terminal points shrinks to the steady-state REE as this cost approaches zero.

For larger calculation costs, in contrast, there exist terminal points that are not near any steady-state REE. In particular, non-collapse terminal points exist in the high deficit case, despite the fact that there are no non-collapse steady-state REE. At terminal points of the latter form, β_t^C may be close to π_{t+1}, so that forecast errors are small. Thus non-existence of non-collapse steady-state REE does not rule out the possibility of 'near-rational' steady-states associated with non-negligible calculation costs.

The condition $\varphi_2 < 1$ ensures that in projecting the future path of inflation, agents reduce at least slightly the degree of price acceleration implied by the previously calculated γ. As example (8) illustrates, this is needed to ensure that agents are allowed to entertain conjectures of convergent inflation paths when $\gamma \neq 1$. It is interesting to note that the condition eliminates 'hyperdeflationary' paths in which β_t^C converges to zero; essentially, the existence of such paths requires that the default paths reflect an extremely large degree of downward price acceleration.[19]

We turn now to the issue of the stability of terminal points, and in particular we are interested in the question of which terminal points are robust to low calculation costs. Two related concepts of stability will be considered. The first concept is motivated by the standard notion of local stability: (β,γ) with $\beta < \bar{\pi}$ is said to be *C-stable* if, for any neighbourhood M containing (β,γ), there exists a neighbourhood $N \subset M$ such that the following is true: for any specification of calculation costs, and for any $(\beta_1^0,\gamma_1^0) \in N$, there exists a CP that terminates in M. Thus initial inputs that are close to (β,γ) can give rise to CPs that remain close to (β,γ).

Our second concept is motivated by the usual notion of asymptotic stability: (β,γ) with $\beta < \bar{\pi}$ is *asymptotically C-stable* if there exists M containing (β,γ) with the following property: for all N containing (β,γ), there exists $\bar{c} > 0$ such that for any $(\beta_1^0,\gamma_1^0) \in M$, all ensuing CPs terminate in N if $c_1 < \bar{c}$. This departs from C-stability by requiring that *all* CPs generated by nearby initial inputs must become arbitrarily close to (β,γ) for sufficiently small calculation costs. Asymptotic C-stability is not stronger than C-stability, however, since the latter does not impose any restrictions on calculation costs.

Using Proposition 2, it is clear that $(\pi_L,1)$ and $(\pi_U,1)$ are the only non-collapse points that can satisfy either definition under the given conditions on φ. Proposition 3 further considers these two points.

Proposition 3: Let ψ and φ be C^1 functions that satisfy (a)–(c) of Proposition 2, and suppose $D < \bar{D}$.

(I) $(\pi_L, 1)$ is C-stable and asymptotically C-stable if the following ratio is sufficiently small:[20]

$$\frac{\psi_1(0|\pi_L,1)}{1-\varphi_2(\pi_L,1)}.$$

(II) $(\pi_U, 1)$ can be neither C-stable nor asymptotically C-stable.

Proposition 3 establishes that the high inflation steady-state REE fails to be stable, in that there are arbitrarily close initial inputs giving rise to CPs that lead away from the steady-state when calculation costs are low. The low inflation steady-state REE satisfies the stability criteria when $\psi_1(0|\pi_L,1)$, which measures the strength of the acceleration property near $(\pi_L, 1)$, is small in the sense stated in Proposition 3. Essentially, the algorithm must not accelerate γ away from $(\pi_L, 1)$ faster than the default paths pull γ in.[21]

Finally, although we will not pursue the details, it is worth noting that the collapse state is stable in a certain sense. Consider the subregion of the collapse region on which $\gamma \geq 1$. Initial points sufficiently close to this subregion, with $\gamma_0 \geq 1$, will generate CPs that remain close to the collapse region, and terminate in collapse if c_1 is sufficiently small. On the other hand, the subregion with $\gamma < 1$ can be regarded as unstable; initial expectations that lie close to the collapse region and below the γ^S curve do not generate CPs that remain close to the collapse region.

Examples of calculation paths

In this section we will consider some examples of CPs that demonstrate how our theory offers explanations for the empirical phenomena discussed in the Introduction. The examples are calculated using the specifications of ψ and φ given in (6) and (8).[22] Costs per calculation are constant at $c_k - c_{k-1} = c$, up to the per period limit of n. The examples consider various values of c and n. Calculation decisions are determined by optimal CDRs, as defined in the preceding section; in all cases we select the optimal CDRs that maximize the number of iterations of the algorithm.

Permanent deficit increase: low deficit case

Suppose that the economy begins in a long-run steady state at $D = 0$, with $\pi_t = \pi_L = 1$ for $t < 1$. At the beginning of period 1, the government announces a permanent increase in the deficit to $D = 1.768$, which is to begin at $t = 3$. This deficit level is associated with a new low inflation steady-state

Table 15.1. *Permanent deficit increase: data for low-deficit case*

$n=1, c=0.01$									
t	1	2	3	4	5	6	7	8	9
k	0	1	1	1	0	0	0	1	0
π_{t+1}	1.244	2.123	2.086	2.085	2.059	2.059	2.008	2.029	2.029
β_t^c	1	1.547	1.869	2.016	2.066	2.091	2.104	2.059	2.057
γ_t^c	1	1.054	1.042	1.025	1.012	1.006	1.003	0.999	1.000
$n=3, c=0.01$									
t	1	2	3	4	5	6	7	8	9
k	0	3	0	2	1	1	0	0	0
π_{t+1}	1.44	2.234	2.039	2.034	2.009	2.031	2.031	2.031	2.031
β_t^c	1	2.085	2.27	2.168	2.116	2.066	2.063	2.062	2.061
γ_t^c	1	1.089	1.045	1.013	1.003	0.999	0.999	1.000	1.000
$n=1, c=1$									
t	1	2	3	4	5	6	7	8	9
k	0	1	0	0	0	0	0	0	0
π_{t+1}	1.244	1.873	1.870	1.868	1.867	1.867	1.867	1.867	1.867
β_t^c	1	1.547	1.632	1.676	1.699	1.711	1.717	1.720	1.721
γ_t^c	1	1.054	1.027	1.014	1.007	1.003	1.002	1.001	1.000

of $\pi_L=2$. Table 15.1 and figure 15.6 summarize the calculation behaviour and inflation paths elicited by the policy beginning in period 1.

Initial inputs in period 1 are given by $\beta_1^0=\gamma_1^0=1$. Note from table 15.1 that announcement of the policy generates no effect in period 1; agents are making calculations of period 2 inflation, and these calculations are not affected by a deficit increase that occurs in period 3. Thus agents have no incentive to calculate, and final outputs remain at the initial input levels, with $\pi_1=1$.

Now consider the calculation decisions in period 2. Agents want to calculate $\hat{\pi}_3$, and since $D_3>0$ the algorithm gives outputs $\beta_2^1>\beta_2^0, \gamma_2^1>\gamma_2^0$. Iteration of the algorithm leads agents to revise $\hat{\pi}_3$ upward, which reduces desired money holdings and increases the period 2 inflation rate. Thus calculation in period 2 generates an inflation reaction *one period in advance* of the policy change. This is consistent with Marimon and Sunder's (1988, 1991) finding that anticipated policy changes led to significant reactions only in the periods immediately preceding the changes.

In this case, the one-period ahead anticipation effect emerges from the one-period forecast horizon of the algorithm. More generally, the degree of

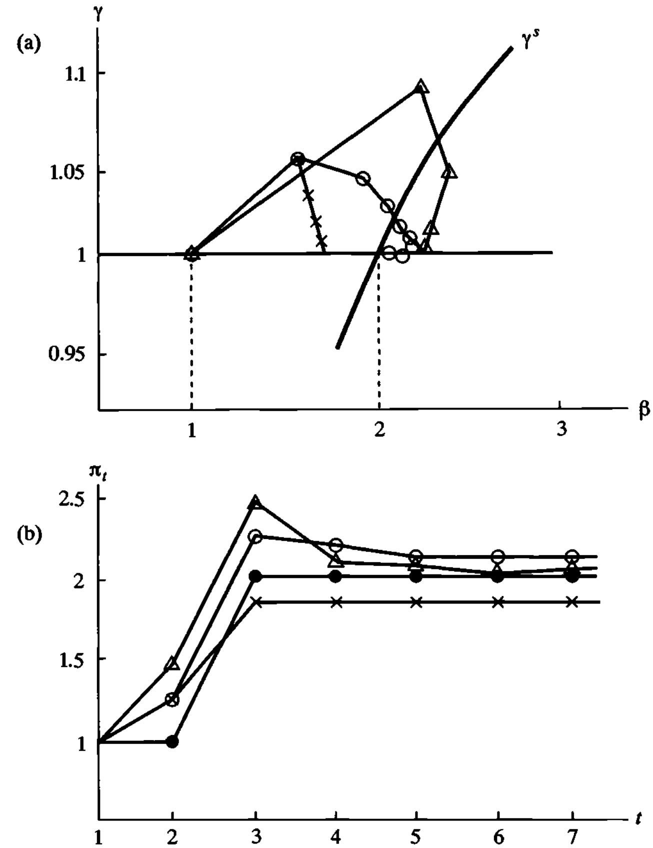


Figure 15.6 Permanent deficit increase: low deficit case
(a) Trajectories of CEPs (b) Inflation paths
○ $n=1, c=0.01$ △ $n=3, c=0.01$
× $n=1, c=1$ ● Path of π_L

anticipation would depend on the length of the forecast horizon built into the algorithm, which in turn would reflect the cognitive abilities of agents. It is important to note that such limited anticipation effects are inconsistent with either rational expectations, which predicts immediate reaction to announced policy changes, or passive adaptive learning, which posits that reactions occur only after the changes are implemented.[23]

In the case of $n=1, c=0.01$, agents calculate again in periods 3 and 4, at which point expectations are sufficiently accurate to dissuade them from calculating in period 5. The default path then carries the agents toward a terminal inflation rate in excess of 2, and by period 8 the default path exceeds 2 by an amount large enough to induce another calculation. Calculation ceases after period 8, and the default path converges to $(\beta, \gamma) = (2.057, 1)$. Qualitatively similar behaviour occurs when n is raised to 3, except that the anticipation effect is much larger and convergence to the new steady-state is more rapid.

Figure 15.6b gives realized inflation along the CPs. For $n = 1, c = 0.01$ and $n = 3$, $c = 0.01$, the actual deficit increase in period 3 has a large impact on period 3 inflation, which overshoots the steady-state as a consequence in part of the rise in inflation expectations between periods 2 and 3.[24] Note finally that when $n = 1$ and costs are raised to $c = 1$, agents calculate once in period 1, and then calculation ceases. The default path approaches $(\beta, \gamma) = (1.72, 1)$, associated with a long-run inflation rate of 1.87. Here the high cost of calculation has dissuaded agents from improving their forecasts, and as a side effect the inflationary impact of the policy is reduced.

Permanent deficit increase: high deficit case

Suppose once again that the economy begins in a long-run steady-state at $D = 0$. At the beginning of period 1, the government announces a permanent increase to the higher deficit level $D = 1.974$, at which no steady-state REE exists. As above, the increase takes effect in period 3. Results are summarized in figure 15.7.

When calculation costs are $n = 1$, $c = 0.1$, agents begin to calculate in period 2, and they perform one iteration per period until collapse occurs in period 10. Note from figure 15.7a that the growth coefficient γ falls for a time along this calculation trajectory, as agents persist in giving weight to the hypothesis that inflation will approach a steady-state. Agents are pulled away from their steady-state hypothesis only after sufficient evidence of impending collapse has accumulated. Figure 15.7b shows the associated inflation path. Observe that a burst of inflation accompanies the deficit increase in period 3, and then inflation accelerates upward until collapse occurs.[25]

For $n = 3$, $c = 0.4$, agents calculate twice in period 2, once in period 3, three times in period 4, and twice more in period 5, at which point they ascertain the occurrence of collapse and demand zero real balances. Here the benefits of calculation increase at later stages of the hyperinflation episode, which gives rise to increasing intensity of calculation as collapse approaches. It is important to note that collapse occurs sooner when n is increased; in essence, calculation capabilities determine the speed at which currency collapse unravels backwards.

The high deficit policy does not inevitably lead to collapse, however.

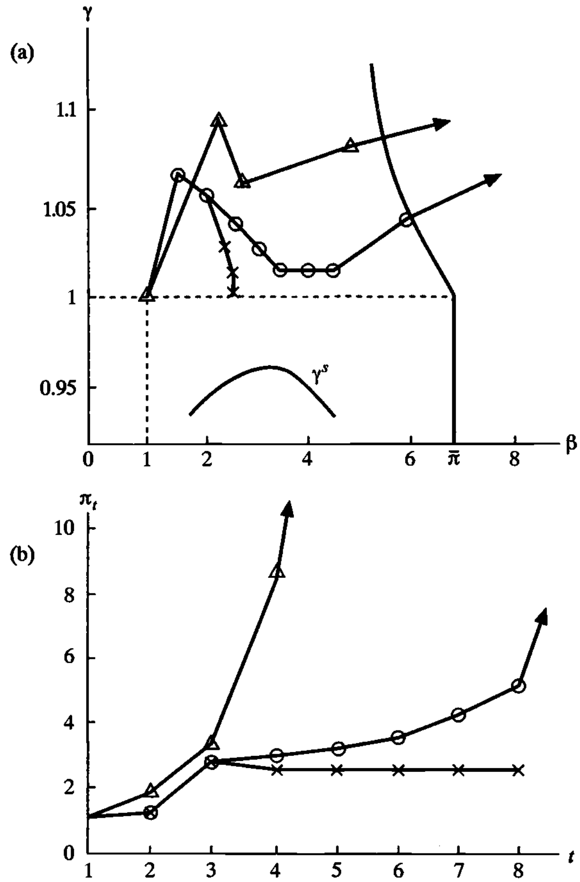


Figure 15.7 Permanent deficit increase: high deficit case
(a) Trajectories of CEPs (b) Inflation paths
○ $n=1, c=0.1$ △ $n=3, c=0.4$ × $n=1, c=1$

When calculation costs are increased to $n=1, c=1$, agents perform one iteration each in periods 2 and 3, and then cease calculation. After this point expectations are sufficiently close to rational that there is insufficient incentive to continue calculation; the CP then evolves toward the terminal point $(\beta, \gamma) = (2.43, 1)$, and the inflation rate reaches a long-run steady-state at 2.60. The occurrence of such a steady-state clearly distinguishes our approach from rational expectations and passive adaptive learning, in that both of the latter predict that no non-collapse steady-state is possible in the high deficit case.

Moreover, steady-states in the high deficit case are most likely to occur where expectations are nearly rational, i.e. in the neighbourhood of the curve γ^S. In fact, Marimon and Sunder did observe steady-states in high deficit cases where no steady-state REE existed, and these steady-states arose precisely where expectation errors were smallest. Our calculation framework provides a simple explanation for this anomalous laboratory finding: subjects' expectations were sufficiently close to rational that rethinking them was not worth the effort.[26]

Policy reversal

We close by considering a slightly more complex policy experiment. The economy begins in a long-run steady-state at a deficit level of $D = 1.768$. In period 1 it is announced that the deficit will be raised to $D = 1.974$ in period 3, and kept at this level until some later period, at which time it will be restored to $D = 1.768$. Figure 15.8 illustrates the inflationary effects for various policy reversal dates.

In figure 15.8a, in which calculation costs are $n=1$, $c=0.01$, collapse occurs in period 10 in the absence of policy reversal. When the policy is reversed at $t=5$, there is a relatively small and transitory inflationary effect, with the inflation rate peaking at 2.81 in period 3. Reversal at $t=8$ generates a higher peak inflation rate of 3.53 in period 6. In period 7 agents anticipate the policy reversal in period 8, and begin calculating downward expectation revisions.

If policy reversal is delayed too long, it becomes impossible to avert currency collapse. Consider the effects of a reversal at $t=11$. Agents calculate in periods 2 through 9, revising upward their conjecture of the inflation path. In period 10, the anticipated deficit reduction makes the default path more accurate, and agents do not calculate. In period 11, however, the default path has advanced sufficiently far toward collapse that calculation once more revises the path upward, despite the lower deficit level. Continued upward acceleration then drives the economy to collapse in period 13. This indicates that the timing of policy reactions to hyperinflationary episodes may be crucial, in that collapse may become inevitable once expectations have been allowed to pass a critical point.

Figure 15.8b illustrates the effects of policy timing under a faster calculation technology, with $n=2$. Now collapse in period 6 becomes inevitable when policy reversal is delayed past $t=7$. It follows that the cost to policy-makers of delay increases when agents can calculate more rapidly.

Conclusion

This chapter has developed a theory of expectation formation in hyperinflationary environments, in which expectation revisions are tied explicitly

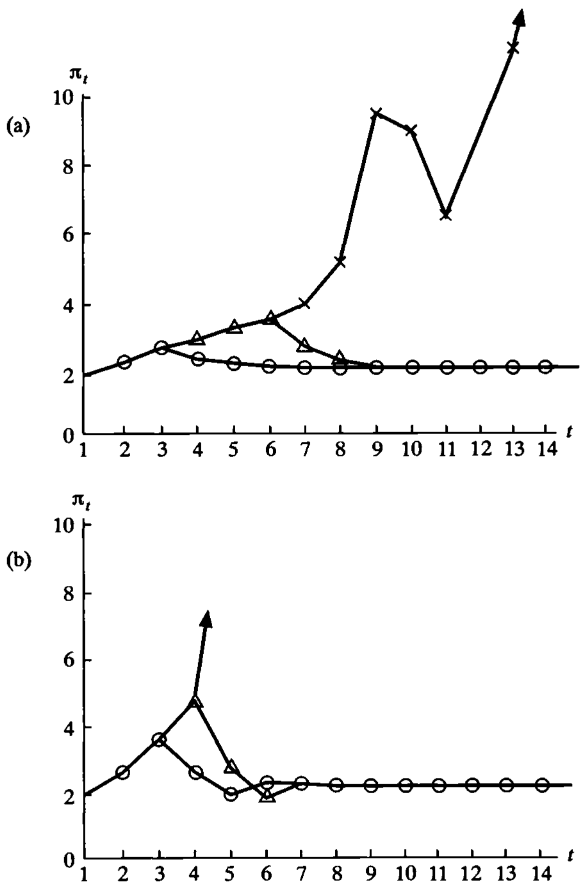


Figure 15.8 Policy reversal
(a) $n=1$, $c=0.01$
○ Reversal at $t=5$ △ Reversal at $t=8$ × Reversal at $t=11$
(b) $n=2$, $c=0.01$
○ Reversal at $t=5$ △ Reversal at $t=6$
Reversal at $t=7$ implies collapse at $t=5$

to agents' trade-offs between losses from inaccurate forecasts and the costs of calculating expectation revisions. Our theory posits a particular form for agents' expectation calculation algorithms, in which they use a true model of the economy to guide their calculations. We show that a wide variety of economic phenomena can be explained in terms of agents' capabilities and incentives to calculate expectations.

Our theory explains how, in cases of unsustainable budget deficits, a pattern of accelerating inflation followed by currency collapse arises naturally from agents' increasing incentives to calculate expectations as the hyperinflationary episode unfolds. Moreover, laboratory findings of one-period ahead anticipation effects, rapid adjustment to new steady-states, and convergence to near-rational outcomes when steady-state REE do not exist, which are difficult to understand in terms of either rational expectations or least squares learning, are easily comprehended when agents' capabilities and incentives to calculate expectations are taken into account.

Finally, our approach has important implications for policy, particularly for policy reversibility. When calculation behaviour is considered, the timing of policy shifts can be crucial not only for the transitional dynamics, but for whether currency collapse can be averted as well.

This chapter has focused on a particular relatively simple specification of the calculation technology, but clearly our framework can be extended in many directions:

Extensions of the calculation algorithm

The algorithm might be based on a model of economic structure that differs from the true structure, where the difference would reflect incomplete knowledge on the part of agents. The algorithm could incorporate a longer forecast horizon, as discussed on pp.315–16, and forecast rules rather than point forecasts might be calculated. The agents could be required to engage in costly calculation of best responses to expectations as well as the expectations themselves (Marimon and Sunder, 1988, 1991, find experimental evidence that agents do not fully optimize their money holdings, given their expectations).

Dependence of the algorithm on the information set

In the specification employed by this chapter, the only aspect of the information set that affects calculation is an agent's own previous calculations. This has allowed us to make a sharp distinction between our approach and that of passive adaptive learning. Our framework can in principle be modified, however, to allow a rich dependence between information and calculation. Calculation behaviour might depend on past economic data, as well as past calculations, via the input functions, the CDRs, or direct revisions of the structural model or calculation procedure. For example, agents might compare past calculated inflation rates with realized rates, with larger discrepancies triggering more intensive calculation. An average of past inflation rates may be updated and used to modify the initial input should agents choose to calculate. These and other extensions would make possible a natural combination of passive adaptation and active cognition.[27]

Determination of the CDR

An adaptive process could be introduced that directly adjusts the CDRs in response to their success at minimizing net losses, e.g. some form of evolutionary adaptation might be specified. Optimality of the CDR would then arise only gradually, if at all.

Our overriding goal has been to establish a general framework for analysing expectation formation behaviour that allows agents to engage in a process of active cognition with respect to the economic environment, but places inherent limitations on their cognitive abilities in terms of restrictions on the calculation technology. We believe that the framework and results of this chapter, and the possible extensions outlined above, are indicative of the scope for this approach to capture a rich variety of expectation formation behaviour.

Appendix

Proof of Proposition 1: Note first that, for all k:

$$\text{sign}[T(\beta^k,\gamma^k\beta^k)-\beta^k]=\text{sign}[T(\beta^0,\gamma^0\beta^0)-\beta^0]. \tag{A1}$$

To see this, suppose $T(\beta^0,\gamma^0,\beta^0)>\beta^0$. Then (5a) and (5b) give $\beta^1>\beta^0$ and $\gamma^1\geq\gamma^0$, where the latter invokes the acceleration property. Further, since $T(\beta,\gamma\beta)$ is strictly increasing in β and γ, we have $T(\beta^1,\gamma^1\beta^1)>T(\beta^0,\gamma^0\beta^0)$ $=\beta^1$. The inequality extends to all k by induction, and the argument applies at once to $T(\beta^0,\gamma^0\beta^0)\leq\beta^0$.

Note next that $k^M(k)$ is non-decreasing in k, that is, $k<k'$, $k_v\in k^M(k)$, $k_w\in k^M(k')$ imply $k_v\leq k_w$. Suppose $T(\beta^0,\gamma^0\beta^0)>\beta^0$ and choose $k_v=\max k^M(k)$, $k'>k$. By definition we have, for all $k_x<k_v$:

$$\begin{aligned} h(K|(\beta^k x)^{-a}-T(\beta^k,\gamma^k\beta^k)^{-a}|)+c_{k_x}\geq h(K|(\beta^k v)^{-a} \\ -T(\beta^k,\gamma^k\beta^k)^{-a}|)+c_{k_v}. \end{aligned} \tag{A2}$$

Using (A1) it follows that $\beta^{k'}>\beta^k$ and $\gamma^{k'}\geq\gamma^k$, and so $T(\beta^{k'},\gamma^{k'}\beta^{k'})$ $>T(\beta^k,\gamma^k\beta^k)$. Similarly, we have $\beta^k x<\beta^k v$. Define the following:

$$\begin{aligned} A&=(\beta^k x)^{-a}-(\beta^k v)^{-a} \\ B&=(\beta^k v)^{-a}-T(\beta^k,\gamma^k\beta^k)^{-a} \\ C&=T(\beta^k,\gamma^k\beta^k)^{-a}-T(\beta^{k'},\gamma^{k'}\beta^{k'})^{-a}. \end{aligned}$$

Note that $A>0$ and $C>0$. Now (A2) may be written:

$$h(K|A+B|)-h(K|B|)\geq c_{k_v}-c_{k_x}>0 \tag{A3}$$

which implies $|A+B|>|B|$; combining this with $A>0$ implies $A+B>0$. To have $k_x\notin k^M(k')$ it is sufficient that:

$$h(K|A+B+C|)-h(K|B+C|)>c_{k_v}-c_{k_x}. \quad \text{(A4)}$$

If $B+C \geq 0$, we may differentiate the left-hand side of (A4) with respect to C to obtain:

$$[h'(K(A+B+C))-h'(K(B+C))]K>0$$

while if $B+C<0$, the derivative becomes:

$$[h'(K(A+B+C))+h'(-K(B+C))]K>0.$$

Thus (A4) follows from (A3) and $C>0$.

Finally, fix $k=n$. If $n \in k^M(n)$, then we are done. Otherwise, $k_v<n$ for all $k_v \in k^M(n)$. Now consider $k=n-1$. Again we are done if $n-1 \in k^M(n-1)$. Otherwise, note that we cannot have $n \in k^M(n-1)$, or we would have $n \in k^M(n)$ due to the fact that k^M is non-decreasing, and so we would have finished in the first step. Thus if $n-1 \notin k^M(n-1)$, then $k_v<n-1$ for all $k_v \in k^M(n-1)$. Proceeding by induction, for each k we have $k \in k^M(k)$ or $k_v<k$ for all $k_v \in k^M(k)$, and at $k=0$ we must have $0 \in k^M(0)$. □

Notes

Part of this chapter was originally included in 'Expectation Calculation and Macroeconomic Dynamics' STICERD *Discussion Paper*, TE/89/202, London School of Economics (1989). Earlier versions were presented at the NBER Conference on Economic Fluctuations in Cambridge, MA (October 1991), the Economic Science Association meetings in Tucson (October 1991), the OFCE conference in Paris (January 1992), and the Econometric Society North American Winter Meetings in New Orleans (January 1992). We have benefited from numerous comments, and we are particularly grateful for the comments and criticisms of John Conlisk, Charles Goodhart, Herschel Grossman, Rudolfo Manuelli, Neil Rankin, Mark Salmon and Michael Woodford. The authors are responsible for all errors.

1. Historical evidence is given in Bental and Eckstein (1990), who also explicate the difficulties of duplicating the historical pattern in a rational expectations equilibrium.
2. In experimental environments with unchanging deficit levels, Marimon and Sunder (1993) did find some agreement between subjects' forecasts and least squares predictions.
3. Note that agents can observe P_t and P_{t-1}, and thus they know π_t, when they choose their period-t money holdings. The future realization π_{t+1}, which depends on P_{t+1}, is what agents are seeking to forecast in this model.
4. Essentially, currency becomes worthless once the government defaults. Equivalently, we can view the government as issuing infinite amounts of currency in the collapse period, which drives P_t to infinity. Alternative assumptions are possible for the period after default, and these might modify calculation behaviour in the periods immediately preceding default.
5. Bental and Eckstein (1990) show that to obtain an accelerating pattern of

inflation that terminates in an anticipated stabilization requires not high but low (i.e. sustainable) deficits in the prestabilization period, combined with an appropriate shift in taxes (in addition to any cut in government consumption) at the stabilization date itself. Another possible route for reconciling rational expectations with historical hyperinflation experience might be to incorporate a regime shift of uncertain existence or timing. See Flood and Garber (1980a) for an example of this kind of approach in a model with an exogeneous money supply process.

6. When $\lambda = 0$ and $\gamma^0 = 1$, the calculation algorithm is equivalent to the notional-time process of *expectational stability* used in Evans (1985) to analyse stability of steady-state REE. Here expectational stability becomes an algorithm for calculating expectations in real time.
7. The qualitative features shown in figure 15.2 hold for $a < 1$, which generates the initial convex region of the γ^S curve, and $\lambda \in (0,1)$, which yields the concave calculation trajectories.
8. Calculating a fixed point of the algorithm in period t would not imply that all forecast errors are eliminated, however, since the algorithm does not incorporate structural changes that occur beyond the forecast horizon, nor does it account for calculations that are made in future periods.
9. One could further extend the algorithm by allowing the choice of H to be part of the algorithm; more generally, agents might be allowed to choose between competing algorithms. We leave the analysis of the more general and complex cases for future work. Other possible extensions of the calculation framework are discussed in the Conclusion.
10. A number of alternative possibilities are discussed in the Conclusion.
11. The default path in Evans and Ramey (1992a) is simply $\beta_t^0 = \beta_{t-1}^C$; agents in that model seek to predict only current inflation, and the calculation algorithm has a single input and output.
12. It is also possible that the CDR would depend on other variables, such as past observed inflation rates; such possibilities are discussed further in the Conclusion (pp.332–3).
13. Discounting of future losses is considered in Evans and Ramey (1992a).
14. This specialization puts $a = 0.5$, $K = 5$, $D = 1.768$, ψ is given by (6) with $\lambda = 0.1$, the loss function is $h(z) = 3z^2$, $n = 3$, and $c_k - c_{k-1} = 1$ for $k = 1,2,3$.
15. Recent experimental work of Smith and Walker (1993a, 1993b) on decision costs gives further support to the notion that calculation behaviour derives from a consistent balancing of benefits and costs: undergraduate experimental subjects were found to commit smaller decision errors when the payments for a given experiment were scaled upward. Here the benefits of calculation are increased, while the costs are presumably constant across subjects, and the results are consistent with the notion that calculation becomes more intensive.
16. In a single-agent model, Conlisk (1988) takes a similar approach to optimal decisions under bounded rationality, and he gives further discussion of the justifications for this approach.
17. φ_2 denotes the derivative with respect to the second argument, γ.
18. For $c_1 \geq h(D)$ the set of terminal points may have other forms. In particular, if

$a \geq 1$ the intervals may be of the form $[\beta, \bar{\pi})$, i.e. there are terminal points arbitrarily close to the collapse region. Further, in the low deficit case a third interval of this form may be added to N_L and N_U.

19. The condition also serves to rule out cases in which the input functions and calculation algorithm exactly offset one another from period to period, which it does by ensuring that current calculations have a persistent effect on the default path.
20. ψ_1 denotes the derivative with respect to the first argument, x.
21. Our findings of potential stability of the low inflation steady-state and instability of the high inflation steady-state are in agreement with Marimon and Sunder's experimental results; see especially Marimon and Sunder (1993).
22. Values for the parameters a, K and λ are as in n. 14 above, and also $\alpha = 0.5$.
23. Limited anticipation could arise even if the calculation algorithm incorporated a very long forecast horizon, provided that calculation costs are not too small.
24. To understand this overshooting effect, let us rewrite (1) as $\Delta m_t + m_{t-1}(\pi_t - 1)/\pi_t = D$, where $m_t = H_t/P_t$. The second term on the left-hand side is the inflation tax component of seigniorage, while the first term is an additional source of seigniorage generated by the willingness of agents to hold higher real balances. When $\beta_t > \beta_{t-1}$, Δm_t is negative, and a higher inflation tax is needed to finance D.
25. Although behaviour is describable as a speculative bubble in this instance, it is not an explosive REE path like those described by Obstfeld and Rogoff (1983), since the adjustment is driven by calculation behaviour rather than REE dynamics. This indicates a source of price level bubbles that has not been addressed by standard empirical tests along the lines developed, for example, by Flood and Garber (1980b).
26. These near-rational steady-states differ from the near-rational equilibria of Akerlof and Yellen (1985a), as well as the steady-state outcomes described in table 15.1, in not being close to any steady-state REE.
27. Further, if agents are heterogeneous, then the information flows available to an agent may include the calculations of other agents. This would generate calculation externalities, as discussed in Evans and Ramey (1992a).

16 Menu costs and aggregate price dynamics

Alan Sutherland

Introduction

The idea that small costs of adjusting prices can create significant nominal rigidities is now a well known result in macroeconomics. The basic implications of menu costs are easily derived and demonstrated in a simple static model of a monopoly firm (as in Akerlof and Yellen, 1985b; Mankiw, 1985). However, there are now a number of studies which generalize this analysis, first, to consider the *dynamic* behaviour of an individual representative firm (Sheshinski and Weiss, 1977, 1983), and secondly, to consider the dynamic behaviour of the *aggregate* price level when firms are heterogeneous (Caplin and Spulber, 1987; Caplin and Leahy, 1991). This chapter brings together the main results of this literature and presents them in a unified framework which is accessible to non-specialist readers.[1]

The chapter begins by deriving and explaining a *dynamic* menu cost model of an individual firm. In this context 'dynamic' means that the firm aims to maximize the discounted value of its profit stream while experiencing serially correlated shocks to the demand for its product. The firm therefore faces an ongoing price adjustment problem such that at every instant it must decide whether and by how much it should adjust its price level. This contrasts with the models of Akerlof and Yellen (1985b) and Mankiw (1985) where the firm experiences a single shock and is only concerned with maximizing profits within the current period. In the first section it is shown how, in the dynamic model, a firm chooses optimal trigger values of the deviation of its price level from the frictionless optimum. While the value of this deviation is small enough to be between the trigger points the firm holds its price fixed. But when the deviation hits one of the trigger points, prices are adjusted by a discrete amount. The basic insight of the static models therefore survives in a dynamic framework.

If all firms are identical and face identical shocks to demand it is obvious that the nominal stickiness displayed by each individual firm will also be displayed by the aggregate price level. But if firms are heterogeneous in some way, so that each firm is potentially at a different position relative to its price trigger points, it becomes less clear that the menu cost effect will survive at the aggregate level. When firms are heterogeneous it becomes theoretically possible for there always to be some firms at or near their trigger point so that any aggregate shock (such as a monetary shock) will always trigger some price adjustment. To investigate this issue it clearly becomes necessary to construct a framework where there are many firms subject to menu costs in the presence of both idiosyncratic and common shocks. Such a framework is constructed in the second section of the chapter.[2]

The third and fourth sections of the chapter then use this framework to analyse two important and contrasting cases. In the first case (due to Caplin and Spulber, 1987) the shocks the firms face are all of the same sign so that prices move in only one direction. It is shown that, in this case, firms are distributed in such a way that any common shock (i.e. money supply shock) induces just enough firms to change their prices to keep the real money supply constant. In other words, the stickiness evident in prices at the level of the individual firm is lost at the aggregate level and monetary shocks are neutral despite the presence of menu costs. The assertion that menu costs at the individual firm level must give rise to nominal stickiness at the aggregate level is thus revealed to be a 'fallacy of composition'.

The fourth section, however, considers an alternative case (due to Caplin and Leahy, 1991) which shows that this fallacy of composition is by no means a general result. The crucial difference in this case is that shocks can be either positive or negative. This apparently innocuous change to the structure of the model is sufficient to produce a very different distribution of firms. The equilibrium distribution of firms is such that there need not be any firms close to a price level trigger point. It is therefore possible for there to be a monetary shock which does not induce any price adjustment. Thus aggregate price inertia is generated and money is non-neutral.

Dynamic optimization and the individual firm

In this section a dynamic model of an individual firm's pricing decision is constructed. The firm's problem is analysed in a continuous-time stochastic framework where the shocks hitting the firm take the form of a 'compound Poisson' process.[3] The compound Poisson process has the advantage that it encompasses both the one-sided and two-sided cases. Thus, by a suitable choice of parameter values, it is possible to generate the Sheshinski and Weiss (1983) case where the firms' optimal price only ever rises, and the Caplin and Leahy (1991) case where the firm's price can both rise and fall.[4]

Suppose that an individual firm faces a stochastic demand schedule such that, in the absence of adjustment costs, its profit maximizing price, p^*, follows a compound Poisson process.[5] Thus if p_t^* is the profit maximizing price at time t then p^* at time $t+dt$ is given by the following

$$p_{t+dt}^* = \begin{cases} p_t^* - \varepsilon \text{ with probability } \alpha\lambda dt \\ p_t^* \text{ with probability } (1-\lambda dt) \\ p_t^* + \varepsilon \text{ with probability } (1-\alpha)\lambda dt \end{cases} \tag{1}$$

where $0<\alpha<1$ and where the innovation, ε, is exponentially distributed with density given by $\gamma e^{-\gamma\varepsilon}$ (ε is therefore non-negative). The average drift (per unit time) in p^* is given by $\lambda(2\alpha-1)/\gamma$. And the variance (per unit time) is $2\lambda/\gamma^2$.

The one-sided and two-sided cases are generated by suitable choice of the parameter α (which has a central role in determining the degree of drift in the p^* process). If $\alpha=1$ the process is asymmetric and prices only ever rise. This corresponds to the case analysed by Sheshinski and Weiss (1983).[6] If, on the other hand, $\alpha=1/2$ the p^* process is exactly symmetric and prices have no tendency to either rise or fall. This is the model of an individual firm that corresponds to the aggregate analysis of Caplin and Leahy (1991).

Denote the actual price that the firm sets at time t by p_t and define x_t to be the deviation between the actual and optimal price, i.e. $x_t=p_t-p_t^*$. Thus, between price changes x follows the following process

$$x_{t+dt} = \begin{cases} x_t - \varepsilon \text{ with probability } (1-\alpha)\lambda dt \\ x_t \text{ with probability } (1-\lambda dt) \\ x_t + \varepsilon \text{ with probability } \alpha\lambda dt \end{cases} \tag{2}$$

Assume the firm's instantaneous profit function is quadratic in x in the region of maximum profits such that

$$\Pi_t = -b[p_t-p_t^*]^2 = -b[x_t]^2 \tag{3}$$

where b is a fixed parameter. Here Π_t is the difference between the actual flow of profit that is earned at time t and the maximum that could be earned if p_t is set equal to p_t^*. As p_t moves away from p_t^* in either direction Π_t becomes negative.

The firm has to pay a menu cost of k every time the price level is altered. The firm's objective is to maximise the discounted value of profits net of adjustment costs. The firm's objective at time t is therefore to maximize W_t where W_t is given by the following

$$W_t = E_t\left\{\int_t^{\infty}[\Pi(\tau)-C(\tau)]e^{-\delta(\tau-t)}d\tau\right\} \tag{4}$$

where δ is the firm's discount rate and E_t is the expectations operator conditional on information available at time t. $C_\tau = k$ if prices are adjusted at time τ, otherwise $C_\tau = 0$. By substituting from (3), (4) can be rewritten as

$$W_t = E_t\left\{\int_t^\infty [-bx_\tau^2 - C_\tau]e^{-\delta(\tau-t)}d\tau\right\}. \tag{5}$$

The firm's problem is to choose x at each moment of time in order to maximize W_t.

The optimal form of rule for the firm is to choose upper and lower trigger points for price changes, denoted u and l, and upper and lower return points, denoted U and L. Thus when x reaches the value u the firm adjusts its price (and therefore incurs the lump sum cost k) so as to return x to U. Likewise at l the price level is adjusted so as to return x to L. Proving that such a trigger point rule is optimal within a wider class of rules is a difficult problem which is not tackled here. For the purposes of the exercise conducted below it is taken as given that a rule of this form is optimal and the procedure for deriving the optimal values of u, l, U and L is outlined. It should be noted, however, that this rule is sufficiently general to encompass a number of important special cases. In particular the case where the firm continuously maintains its price at p^* can be represented as $u = l = U = L = 0$. The fact that this does not turn out to be optimal for a firm that faces menu costs shows that price stickiness is not an artefact of the assumption that a trigger point rule is followed.

The procedure for deriving the optimal values of u, l, U, L involves first deriving an explicit expression for the value of the firm's objective function. It is possible to deduce from the structure of the model that W_t will be some function of x_t, i.e. $W_t = V(x_t)$.[7] Consider a small interval of time, dt, during which price adjustment does not take place (i.e. $l < x_t < u$ and $l < x_{t+dt} < u$). The function $V(\cdot)$ satisfies the following[8]

$$V(x_t) = -bx_t^2dt + (1-\delta dt)E_t[V(x_{t+dt})]. \tag{6}$$

The definition of the process followed by x allows the last term of this to be expanded to yield

$$V(x_t) = -bx_t^2dt + (1-\delta dt)\begin{Bmatrix}(1-\lambda dt)V(x_t) + \\ (1-\alpha)\lambda dt E[V(x_t - \varepsilon)] \\ \alpha\lambda dt E[V(x_t + \varepsilon)]\end{Bmatrix}. \tag{7}$$

The first term in the large bracket is the value of the objective function at $t + dt$ if there is no change in x multiplied by the probability of there being no change. The second term is the expected value of the objective function given that there is a negative jump in x multiplied by the probability of a negative jump. The third term is the expected value of the objective function given that there is a positive jump in x multiplied by the probability of a positive jump.

Multiplying through the large bracket by $(1-\delta dt)$ and eliminating terms in $(dt)^2$ allows (7) to be rewritten as follows[9]

$$(\delta+\lambda)V(x_t) = -bx_t^2 + (1-\alpha)\lambda E[V(x_t-\varepsilon)] + \alpha\lambda E[V(x_t+\varepsilon)]. \tag{8}$$

It is possible to obtain an explicit solution for the function V conditional on values for l, u, L and U by using a two-step procedure. In the first step the existence of the trigger points is ignored and (8) is expanded to yield

$$\begin{aligned}(\delta+\lambda)V(x_t) = -bx_t^2 + (1-\alpha)\lambda \int_0^\infty \gamma e^{-\gamma\varepsilon} V(x_t-\varepsilon)d\varepsilon \\ + \alpha\lambda \int_0^\infty \gamma e^{-\gamma\varepsilon} V(x_t+\varepsilon)d\varepsilon.\end{aligned} \tag{9}$$

This is an integral equation in the function $V(\cdot)$. It has a general solution of the form

$$\begin{aligned}V(x_t) = -\frac{2b\lambda}{\delta^3\gamma^2}[(2\alpha-1)^2\lambda+\delta] - \frac{(2\alpha-1)2b\lambda}{\gamma\delta^2}x_t - \frac{b}{\delta}x_t^2 \\ + A_1 e^{\theta_1 x_t} + A_2 e^{\theta_2 x_t}\end{aligned} \tag{10}$$

where θ_1 and θ_2 are the roots of $-(\delta+\lambda)\theta^2 + (1-2\alpha)\lambda\gamma\theta + \delta\gamma^2 = 0$ and A_1 and A_2 are arbitrary coefficients.

Recall that in deriving (10) the existence of the trigger points was ignored. The trigger points enter the solution process by providing two boundary conditions to tie down the two arbitrary coefficients, A_1 and A_2. The two conditions are called 'value matching' conditions because they specify necessary relationships between $V(U)$, $V(L)$, $V(u)$ and $V(l)$, i.e. the value of the firm's objective function at the different trigger points. To derive the first value matching condition consider a slight rearrangement of (9) with V evaluated at u

$$\begin{aligned}V(u) = -\frac{b}{(\delta+\lambda)}u^2 + \frac{(1-\alpha)\lambda}{(\delta+\lambda)}\int_0^\infty \gamma e^{-\gamma\varepsilon}V(u-\varepsilon)d\varepsilon \\ + \frac{\alpha\lambda}{(\delta+\lambda)}\int_0^\infty \gamma e^{-\gamma\varepsilon}V(u+\varepsilon)d\varepsilon.\end{aligned} \tag{11}$$

After making due allowance for the trigger and return points in the integrals on the right-hand side of this equation, it is possible to rewrite (11) as

$$V(u) = -\frac{b}{(\delta+\lambda)}u^2 + \frac{(1-\alpha)\lambda}{(\delta+\lambda)}\left\{\int_0^{u-l} \gamma e^{-\gamma\varepsilon} V(u-\varepsilon)d\varepsilon + e^{-\gamma(u-l)}[V(L)-k]\right\} + \frac{\alpha\lambda}{(\delta+\lambda)}[V(U)-k]. \quad (12)$$

This follows because any shock which takes x above u results in x being reset to U at a cost of k. While any shock which takes x below l results in x being reset to L at a cost of k. A corresponding equation can be derived at l as follows

$$V(l) = -\frac{b}{(\delta+\lambda)}l^2 + \frac{(1-\alpha)\lambda}{(\delta+\lambda)}[V(L)-k] + \frac{\alpha\lambda}{(\delta+\lambda)}\left\{\int_0^{u-l} \gamma e^{-\gamma\varepsilon} V(l+\varepsilon)d\varepsilon + e^{-\gamma(u-l)}[V(U)-k]\right\}. \quad (13)$$

(12) and (13) are two independent relationships that must hold between $V(U)$, $V(L)$, $V(u)$ and $V(l)$. After substituting for $V(U)$, $V(L)$, $V(u)$ and $V(l)$ using the functional form given in (10), (12) and (13) can be solved to yield values for A_1 and A_2 conditional on values for l, u, L and U.

(10), together with the values of A_1 and A_2 that come from the value matching conditions, provide an explicit solution for the firm's objective function conditional on arbitrary values of the trigger and return points, i.e. it is possible to write $W_t = V(x_t; l, u, L, U)$. The next stage is to derive the values of l, u, L and U that maximize W_t. This stage of the process is perfectly standard. The usual first order conditions can be derived from the explicit solution for W_t (by setting the derivatives of V with respect to l, u, L and U equal to zero). The optimal values for l, u, L and U are then derived in the usual way by solving the first order conditions. The first order conditions are stated below without any explicit derivation. (Note that even though V is a function of x_t it turns out that x_t can be eliminated from the first order conditions. The optimal values of the trigger points are therefore independent of the current value of the state variable.) After considerable rearrangement and simplification it is possible to write the four first order conditions as follows

$$V'(U) = 0 \quad (14)$$

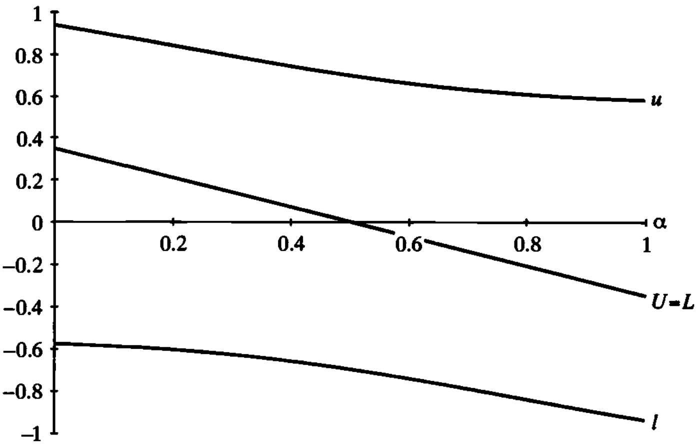


Figure 16.1 Optimal trigger points

$$V'(L)=0 \tag{15}$$

$$V'(u)=\frac{2b}{(\delta+\lambda)}u+\frac{(1-\alpha)\lambda}{(\delta+\lambda)} \times\left\{\int_0^{u-l}\gamma e^{-\gamma\varepsilon}V'(u-\varepsilon)d\varepsilon-\gamma e^{-\gamma(u-l)}[V(L)-V(l)-k]\right\} \tag{16}$$

$$V'(l)=\frac{2b}{(\delta+\lambda)}l+\frac{\alpha\lambda}{(\delta+\lambda)} \times\left\{\int_0^{u-l}\gamma e^{-\gamma\varepsilon}V'(l+\varepsilon)d\varepsilon+\gamma e^{-\gamma(u-l)}[V(U)-V(u)-k]\right\} \tag{17}$$

where $V'(U)$, $V'(L)$, $V'(u)$ and $V'(l)$ should be read as the derivative of $V(x)$ (with respect to x) evaluated at, respectively, $x=U$, L, u and l. The four first order conditions can be solved to yield the optimal values of U, L, u and l.

Figure 16.1 plots optimal values of U, L, u and l for different values of α. Remember, α determines the direction and degree of drift in the p^* process. Thus if $\alpha=1/2$ there is no drift in p^*, and if $\alpha=0$ there are only ever positive shocks to p^* so there is positive average drift. The values of the other parameters are $\lambda=20$, $\gamma=4$, $\delta=1$ and $k=0.05$.

At $\alpha=1/2$ the solution is symmetric, with $U=L=0$ and $u=-l=0.693$. But as α declines towards zero the solution becomes steadily more asymmetric. U and L rise above the frictionless optimal point ($x=0$) and u and l also rise. The rise in U and L reflects the negative drift in x induced by the positive drift in p^*. When resetting the price level it is optimal to set a price above p^* to allow for future anticipated price rises. The lower trigger point, l, rises because when $p^*>p$ the positive drift means there is less chance of p^* falling back towards the current level of p. It is therefore better to reset prices sooner. A symmetric argument explains why the upper trigger point rises.

In the extreme case where $\alpha=0$, so that p^* never falls, only the lower trigger point, l, and the return point are relevant for ongoing price changes. The upper trigger point is relevant only if the initial level of x is above U. In this case the firm will adjust x to U if $x>u$. Otherwise the firm will simply allow the positive drift in p^* to make the required adjustment. After this initial adjustment phase x will never again be in the region $x>U$.

Aggregate price dynamics: the basic framework

The previous section illustrated the optimizing problem facing an individual firm in choosing its price. This was illustrated in a continuous-time continuous-state–space framework. In this section, and the two that follow, the implications of impulse control of prices for the aggregate economy are illustrated. Here the optimal choice of trigger points will not be explicitly represented. Instead it will simply be assumed that firms follow trigger strategies of the type illustrated in the previous section. In addition it will prove easier to work in a discrete-time discrete-state–space framework. As far as possible the notation established in the previous section will be retained.

There are assumed to be n price-setting firms in the economy where n is a large number. Each firm's optimal price is subject to common shocks and idiosyncratic shocks. It is convenient to parameterize the model in such a way as to allow the relative balance of common and idiosyncratic shocks to be varied.[10] Common shocks take the form of money growth shocks. Money growth can take on one of two values with equal probability as follows

$$\Delta m=\begin{cases}\mu+S/2\\ \mu-S/2\end{cases}. \tag{18}$$

The parameter μ measures the average growth rate of money and S determines the level of money growth shocks. It is natural to assume that the aggregate optimal price level (denoted p^*) will change proportionately with money, i.e. $\Delta p^*=\Delta m$.

At the level of the individual firm, the change in the optimal price of firm i can take one of two values

$$\Delta p_i^* = \begin{cases} a \\ b \end{cases} \quad \text{where } a > b. \tag{19}$$

The actual realization of Δp_i^* will depend on a combination of common and idiosyncratic shocks.[11] But from the point of view of the individual firm the source of the shock is unimportant. In what follows it is assumed that the combination of common and idiosyncratic shocks results in a and b being equiprobable. Obviously for the model to be consistent, the parameters a, b, μ and S cannot be chosen independently. (19) implies that the expected growth rate of the aggregate optimal price level, p^*, is $(a+b)/2$. Thus for (18) and (19) to be consistent it must be true that $\mu = (a+b)/2$. Also, if there are no idiosyncratic shocks, so that all firms have the same optimal price level, it must be true that $S = (a-b)$. Thus $S = (a-b)$ defines the maximum value of S.

In this model idiosyncratic shocks imply that firms do not all experience the same shock to p_i^* in each period. If there are no common shocks (i.e. $S = 0$) then, given that there are a large number of firms, half of the firms will experience $\Delta p_i^* = a$ and half will experience $\Delta p_i^* = b$. But if there are common shocks the proportions experiencing a and b will vary according to the rate of money growth. Thus in a period where the high money growth rate occurs the proportion experiencing $\Delta p_i^* = a$ is assumed to be π (where $\pi > 1/2$) and the proportion experiencing $\Delta p_i^* = b$ is $(1-\pi)$, where the proportion π satisfies the following equation

$$\pi a + (1-\pi)b = \frac{(a+b)}{2} + \frac{S}{2}. \tag{20}$$

The left-hand side of this equation is the change in the aggregate optimal price and the right-hand side is the change in the money supply given that a positive monetary shock has occurred. Rearranging (20) yields the following expression for π

$$\pi = \frac{1}{2} + \frac{1}{2}\left(\frac{S}{a-b}\right). \tag{21}$$

π is also the probability of an individual firm experiencing $\Delta p_i^* = a$ *conditional on a high money growth shock having occurred.* To make this consistent with there being an unconditional probability of $\Delta p_i^* = a$ occurring of $1/2$ for an individual firm it must be the case that π is also the proportion of firms experiencing $\Delta p_i^* = b$ in periods when a low money growth shock occurs.

The framework can be summarized as follows. In each period a money growth shock of size $(a+b)/2 + S/2$ occurs with probability $(1/2)$ and of size $(a+b)/2 - S/2$ with probability $(1/2)$. If high money growth occurs a proportion π (given by (21)) of firms experience $\Delta p_i^* = a$ and $(1-\pi)$ experience $\Delta p_i^* = b$. If low money growth occurs then the proportions are reversed so that $(1-\pi)$ experience $\Delta p_i^* = a$ and π experience $\Delta p_i^* = b$. The net result for an individual firm is that $\Delta p_i^* = a$ and $\Delta p_i^* = b$ each occur with an unconditional probability of 1/2. The balance between common and idiosyncratic shocks can be varied by varying S between 0 and $(a-b)$.

Aggregate output is determined by

$$y = m - p \tag{22}$$

where p is the aggregate actual price level (as distinct from the aggregate optimal price level p^*). The change in output is therefore given by

$$\Delta y = \Delta m - \frac{1}{n} \sum_{i=1}^{n} \Delta p_i \tag{23}$$

where n is the number of firms. But since $\Delta m = \Delta p^*$ this can be rewritten as

$$\Delta y = \frac{1}{n} \sum_{i=1}^{n} \Delta p_i^* - \frac{1}{n} \sum_{i=1}^{n} \Delta p_i = -\frac{1}{n} \sum_{i=1}^{n} \Delta x_i \tag{24}$$

where $x_i = (p_i - p_i^*)$ is the price deviation for firm i. Thus the dynamics of output are entirely determined by the dynamics of the average price deviation. The effect of menu costs on the dynamics of output can therefore be studied by considering the average of the x_is. In particular, the effect of monetary shocks on output in the presence of menu costs can be investigated by considering the effect of monetary shocks (i.e. common shocks) on the average of the x_is.

In order to investigate the dynamics of the average of the x_is it is necessary to derive and investigate the dynamics of the 'cross-sectional distribution' of firms over the different possible values of x.[12] Thus, much of the analysis in the next two sections is directed at deriving the dynamics of the cross-sectional distribution of the x_is. The following section considers the case of one-sided shocks while the subsequent section considers the case of two-sided shocks.

One-sided price shocks

This section considers the case as analysed by Caplin and Spulber (1987) where only positive p^* shocks occur. The appropriate p^* process can be generated by setting $a = 1$ and $b = 0$. This is a discrete time equivalent of the one-sided case illustrated on p.343. There it was shown that the optimal policy for an individual firm is to choose a minimum level for x, denoted l, and a return point, denoted L. For the sake of illustration, in this section set

$l = -3$ and $L = 2$. For an individual firm x can therefore only take on the values -2, -1, 0, 1 and 2. When x falls to -3 it is immediately reset to 2.

Notice that these assumptions imply that average money growth is $\mu = 1/2$ and that $0 \leq S \leq 1$. At $S = 0$ there are no monetary shocks while at $S = 1$ monetary shocks are the only source of noise.

The dynamics of the cross-sectional distribution of x can be investigated by considering the behaviour of x for two particular firms randomly chosen from the total population of firms.[13] Call these firms i and j. In particular, consider the joint probability distribution of x_i and the difference between x_i and x_j, i.e. $x_i - x_j$. x_i can take on values $-2, -1, 0, 1, 2$ while $x_i - x_j$ can take on values from -4, where $x_i = -2$ and $x_j = 2$, to 4, where $x_i = 2$ and $x_j = -2$. The joint distribution at time t, conditional on information available at time 0, can therefore be represented by a 5×9 matrix denoted $Q(t)$. Here, for instance, the element $q_{-1,-2}(t)$ of $Q(t)$ is the probability that at time t $x_i = -1$ and $x_i - x_j = -2$. In what follows, the steady-state solution for $Q(t)$ (i.e. the solution as t tends to infinity) will be derived.[14]

It may, at first sight, seem odd to study the dynamics of the joint probability distribution of x_i and $x_i - x_j$. It will be shown below, however, that this distribution can be used to determine the cross-sectional distribution of firms. In particular the steady-state Q can be used to determine the steady-state cross-sectional distribution.[15] This can then be used to show the effects of monetary shocks on the average value of x across the population of firms.

Suppose firms i and j are chosen randomly at time 0 and suppose that at time 0 the position of the two firms is known with certainty. It therefore follows that one element of $Q(0)$ is unity and the others are all zero. But, from the standpoint of time 0, the future position of the two firms is unknown, so Q must evolve as periods further and further into the future are considered. As t is increased Q evolves according to rules which can be derived from the basic probability rules governing the individual firms. The first stage in considering the evolution of Q is therefore to consider these basic probability rules.

If there is a high money growth shock the probability of x_i falling is $(1+S)/2$. The probability of x_j falling is also $(1+S)/2$. These are independent events, so the joint probability of both x_i and x_j falling in a high money growth period is $(1+S)^2/4$. Likewise the joint probability of x_i falling and x_i remaining unchanged is $(1+S)(1-S)/4$. The joint distribution of the changes in x_i and x_j is therefore given by table 16.1.

In a low money growth period the corresponding joint distribution is as shown in table 16.2.

High and low money growth are equi-probable so the overall joint distribution is the average of the previous two tables, as shown in table 16.3.

Table 16.1. *Joint distribution of changes in x_i and x_j: high money growth period*

	$\Delta x_j = -1$	$\Delta x_j = 0$
$\Delta x_i = -1$	$(1+S)^2/4$	$(1+S)(1-S)/4$
$\Delta x_i = 0$	$(1+S)(1-S)/4$	$(1-S)^2/4$

Table 16.2. *Joint distribution of changes in x_i and x_j: low money growth period*

	$\Delta x_j = -1$	$\Delta x_j = 0$
$\Delta x_i = -1$	$(1-S)^2/4$	$(1+S)(1-S)/4$
$\Delta x_i = 0$	$(1+S)(1-S)/4$	$(1+S)^2/4$

Table 16.3. *Overall joint distribution of changes in x_i and x_j*

	$\Delta x_j = -1$	$\Delta x_j = 0$
$\Delta x_i = -1$	$(1+S^2)/4$	$(1-S^2)/4$
$\Delta x_i = 0$	$(1-S^2)/4$	$(1+S^2)/4$

Notice that as the degree of common shocks rises the more correlated the two firms become. It the extreme when $S=1$ the two firms are perfectly correlated. And when $S=0$ the two firms are completely independent.

Table 16.3 can be used to construct the dynamics of Q. Suppose that $q_{0,0}(0)=1$ and that all other elements of $Q(0)$ are zero. This implies $x_i=x_j=0$. If $\Delta x_j=\Delta x_i=-1$ then in period 1 $x_i=-1$ and $x_i-x_j=0$. Table 16.3 shows that the probability of $\Delta x_j=\Delta x_i=-1$ is $(1+S^2)/4$ so $q_{-1,0}(1)=(1+S^2)/4$. Similar reasoning shows that $q_{0,0}(1)=(1+S^2)/4$, $q_{-1,-1}(1)=(1-S^2)/4$ and $q_{0,1}(1)=(1-S^2)/4$. It is convenient to represent this with a transition matrix $T_{0,0}$ which summarizes the transition of probability mass from $q_{0,0}$ as follows

$$T_{0,0}=\tfrac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-S^2) & (1+S^2) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1+S^2) & (1-S^2) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (25)$$

It is then possible to write

$$Q(1) = q_{0,0}(0)T_{0,0}. \tag{26}$$

It is clearly possible to derive a transition matrix for each element of Q. Had it been the case, for instance, that in period 0 $x_i = x_j = -2$ so that $q_{-2,0}(0) = 1$ the relevant transition matrix is $T_{-2,0}$ which is given as follows

$$T_{-2,0} = \frac{1}{4}\begin{bmatrix} (1-S^2) & 0 & 0 & 0 & (1+S^2) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1+S^2) & 0 & 0 & 0 & (1-S^2) \end{bmatrix}. \tag{27}$$

In this case the transition matrix takes into account the potential resetting of x_i and/or x_j. $Q(1)$ is now given by

$$Q(1) = q_{-2,0}(0)T_{-2,0}. \tag{28}$$

Thus, from any initial starting point it is possible to write down a rule linking $Q(0)$ and $Q(1)$.

Now consider a slightly more general problem. Suppose the positions of the two firms are unknown at time 0 but that their probability distribution is defined by $Q(0)$. (In other words, suppose that there is more than one non-zero element in $Q(0)$.) There is now a general rule that links $Q(1)$ to $Q(0)$ as follows

$$Q(1) = \sum_{h=-2}^{2} \sum_{k=-4}^{4} q_{h,k}(0)T_{h,k} \tag{29}$$

where account is taken of all the possible starting points for x_i and x_j. (29) defines the transitional dynamics from period 0 to period 1. The transition matrices are, however, time invariant so (29) can be used to define the transition from any period, t, to the next period, $t+1$, i.e.

$$Q(t+1) = \sum_{h=-2}^{2} \sum_{k=-4}^{4} q_{h,k}(t)T_{h,k} \tag{30}$$

This defines the dynamics of the distribution of the xs.

In particular (30) can be used to derive the steady-state distribution $\bar{Q}$ which is defined by

$$\bar{Q} = \sum_{h=-2}^{2} \sum_{k=-4}^{4} \bar{q}_{h,k}T_{h,k}. \tag{31}$$

(31) provides 45 linear equations in the 45 elements of $\bar{Q}$. However,

(provided $S < 1$) only 44 of these equations are linearly independent. A 45th independent equation is provided by the requirement that the elements of $\bar{Q}$ should sum to unity. Solving for the elements of $\bar{Q}$ is therefore a matter of solving 45 linear equations in 45 unknowns.

The case where $S = 1$, so that there are no idiosyncratic shocks, requires a slightly different approach. In this case the 45 equations are not linearly independent. There are thus multiple solutions for the steady-state distribution. But in this case the solution to the model becomes very simple since the lack of idiosyncratic shocks implies that the relationship between firms is fixed throughout time. The long-run distribution of $x_i - x_j$ can therefore be determined from the initial conditions. Conversely it must be the case that there is a solution for $\bar{Q}$ that corresponds to each possible starting point for the two firms.

In the case where $S < 1$ the following solution for $\bar{Q}$ is obtained

$$\bar{Q} = \begin{bmatrix} 1/25 & 1/25 & 1/25 & 1/25 & 1/25 & 0 & 0 & 0 & 0 \\ 0 & 1/25 & 1/25 & 1/25 & 1/25 & 1/25 & 0 & 0 & 0 \\ 0 & 0 & 1/25 & 1/25 & 1/25 & 1/25 & 1/25 & 0 & 0 \\ 0 & 0 & 0 & 1/25 & 1/25 & 1/25 & 1/25 & 1/25 & 0 \\ 0 & 0 & 0 & 0 & 1/25 & 1/25 & 1/25 & 1/25 & 1/25 \end{bmatrix}. \quad (32)$$

It is simple to calculate from this (by summing the rows) that the unconditional distribution for firm i is uniform across the five states.[16] It is also simple to use $\bar{Q}$ to calculate the distribution for x_j conditional on the position of x_i. When $x_i = -2$ the first row of $\bar{Q}$ implies that $x_j = -2$ with probability 1/5, $x_j = 1$ with probability 1/5, etc. The conditional distributions of x_j for the other values of x_i can be calculated from the other rows $\bar{Q}$. These distributions are all uniform across the five states.

The distribution $Q(t)$ was introduced as the joint probability distribution of two individual firms, i and j, conditional on information available at time 0. But since the number of firms is large it is possible to use $Q(t)$ to say things about the proportion of firms in different states[17] (provided that there is some element of idiosyncratic shocks). Assume that there is a large number of firms which are initially arbitrarily distributed across the xs. Interpret $Q(0)$ as the joint distribution of x_i and $x_i - x_j$ where firms i and j are randomly selected from the total population of firms. Each row of $Q(0)$ is therefore a simple transformation of the cross-sectional distribution of firms conditional on x_i.[18] Furthermore, the matrices which define the transitional dynamics of Q (i.e. the T matrices) can be interpreted as defining the transitional dynamics of the cross-sectional distribution. In this case, each T matrix is interpreted as recording the movement of firms between states rather than the movement of probability mass. The transitional dynamics of the cross-sectional distribution can thus be traced using the same techniques as described above. And, in particular, the steady-state cross-sectional distribution can be obtained from $\bar{Q}$.

It has already been shown that each row of $\bar{Q}$ implies that the distribution of x_j conditional on a given value of x_i is uniform across the five possible values of x. It must therefore follow that the cross-sectional distribution is uniform across the five states. In addition, it is important to note that this result is true for any degree of common shocks less than $S=1$. Thus it only requires some degree of idiosyncrasy in shocks to lead to a uniform cross-sectional distribution.

In order to see more clearly that a uniform distribution is indeed a steady-state, consider in detail the effect of some specific shocks. In a high money growth period the proportion of firms experiencing a rise in p^* is $(1+S)/2$. The other $(1-S)/2$ firms experience no change in p^*. Thus of the firms previously at $x=1$ $(1+S)/2$ move to $x=0$ and $(1-S)/2$ stay at $x=1$. Similarly, of those firms previously at $x=0$, $(1+S)/2$ move to $x=-1$ and $(1-S)/2$ stay at $x=0$. The proportion of firms left at $x=0$ at the end of the period is therefore $(1+S)/10+(1-S)/10=1/5$, i.e. there is no change. The interaction between $x=-2$ and $x=2$ is slightly more complicated, but the result is the same. Of the firms previously at $x=-2$, $(1+S)/2$ experience a rise in p^* which takes them past the trigger value of x. They therefore reset x to $x=2$. The proportion of firms left at $x=2$ at the end of the period therefore also remains at $1/5$. The same exercise conducted for the case of a low money growth period yields exactly the same result. It is thus clear that a uniform cross-sectional distribution is indeed a steady-state.

The implications of this for the dynamics of the average value of x, and hence for the effect of monetary shocks on output, are now also clear. If the distribution of firms across the xs is always uniform the average value of the xs will always be zero. (24) therefore implies that output will be completely unaffected by monetary shocks. In other words, there is complete monetary neutrality at the aggregate level.

This result obviously depends heavily on the uniform cross-sectional distribution of firms. It is therefore worth considering situations in which the uniform distribution does not hold. First consider the case where there are no idiosyncratic shocks, i.e. $S=1$. In this case the dynamics of $Q(t)$ become path-dependent and the steady-state cross-sectional distribution is determined by the initial cross-sectional distribution. Thus, if firms are all concentrated on one value of x, that distribution will obviously also be the steady-state cross-sectional distribution and the dynamics of the average value of x will obviously be identical to the dynamics of an individual firm. In this case the non-neutralities exhibited at the individual firm level will also be reflected at the average level. So for instance if all firms are at $x=0$ a high money growth shock will result in all firms moving to $x=-1$. From (24) it can be seen that output expands ($\Delta y=-\Delta x=1$).

There is a sense, however, in which even this case can be said to exhibit monetary neutrality. To see this, consider what happens when all firms are at $x=-2$. If a high money growth shock occurs, all firms will reset their prices so that x moves to $x=2$, i.e. $\Delta x=4$. From (24) it can be seen that this causes a large fall in output ($\Delta y=-\Delta x=-4$). The probability distribution over values of x for each individual firm, and thus for all firms, is uniform. There is therefore a probability of 1/5 that a positive money growth shock will cause a contraction of output of 4 and a probability of 4/5 that a money growth shock will cause an expansion of output of 1. The unconditional expectation of changes in output is thus zero. In other words, it is still possible to argue that monetary shocks are neutral even when there are no idiosyncratic shocks.

The second case in which a uniform cross-sectional distribution of firms will not be present is during the transition from some arbitrary initial distribution towards the steady-state. If, for instance, all firms start with $x=0$ it will take some time for idiosyncratic shocks to spread the firms into the steady-state uniform distribution. This transitional phase will be particularly lengthy in the case where idiosyncrasies are small. During this phase there will clearly be scope for monetary shocks to affect output. This non-neutrality will, however, be subject to the same criticism as just discussed, namely that the unconditional expectation of changes in output will be zero.[19]

Two-sided price shocks

The previous section considered the case where price shocks are all positive, so prices only ever rise. This section considers the alternative case, where prices can both rise and fall. The two-sided case is captured by setting $a=1$ and $b=-1$ so that the optimal price level follows a random walk (which is the discrete-time equivalent of the Brownian motion process adopted in Caplin and Leahy, 1991). It will be assumed that when x falls past -2 it is reset to zero and when it rises above $+2$ it is also reset to zero. There are thus five possible values for x, i.e. -2, -1, 0, 1 and 2. The parameter S which measures the size of idiosyncratic shocks can now vary between 0 and 2. When $S=2$ there are no idiosyncratic shocks.

It is again possible to define $Q(t)$ as the joint probability distribution of x_i and x_i-x_j where i and j are two individual firms. The transitional dynamics of $Q(t)$ will differ from those described in the previous section because the dynamics of p^* are different and because the optimal policy followed by firms involves two-sided impulse control of x. Note that the probability of both firms i and j moving in the same direction is $(1+S^2/4)/4$ and the probability of firms i and j moving in different directions is $(1-S^2/4)/4$. Thus for instance the transition matrix for $q_{0,0}(t)$ is now

$$T_{0,0}=\tfrac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-S^2/4) & 0 & (1+S^2/4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1+S^2/4) & 0 & (1-S^2/4) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (33)$$

This takes account of the possibility that prices may fall as well as rise.

The transition matrix for $q_{-2,-4}(t)$ is

$$T_{-2,-4}=\tfrac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & (1-S^2/4) & (1+S^2/4) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & (1+S^2/4) & (1-S^2/4) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (34)$$

This is where firm i and firm j are at opposite extremes of the range of xs. $T_{-2,-4}$ therefore takes account of the possibility of each firm having to reset its price to zero.

Given the appropriate modifications to the transition matrices the dynamics of $Q(t)$ are again defined by (30) and the steady-state is defined by equation (31). The solution for $\bar{Q}$ is as follows

$$\bar{Q}=\tfrac{1}{27}\begin{bmatrix} \frac{\alpha}{(1-\alpha)} & \frac{2\alpha}{(1-\beta)} & 1 & \frac{2\beta}{(1-\alpha)} & \frac{\beta}{(1-\beta)} & 0 & 0 & 0 & 0 \\ 0 & \frac{2\alpha}{(1-\beta)} & \frac{1}{(1-\alpha)} & 2 & \frac{1}{(1-\beta)} & \frac{2\beta}{(1-\alpha)} & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{2\beta}{(1-\alpha)} & \frac{1}{(1-\beta)} & 2 & \frac{1}{(1-\alpha)} & \frac{2\alpha}{(1-\beta)} & 0 \\ 0 & 0 & 0 & 0 & \frac{\beta}{(1-\beta)} & \frac{2\beta}{(1-\alpha)} & 1 & \frac{2\alpha}{(1-\beta)} & \frac{\alpha}{(1-\alpha)} \end{bmatrix} \quad (35)$$

where $\alpha=(1-S^2/4)/4$ and $\beta=(1+S^2/4)/4$. This solution is unique for $S<2$, i.e. provided there is some idiosyncratic component to shocks. The unconditional distribution of x_i (the row sums of $\bar{Q}$) is

$$\begin{bmatrix} \frac{1}{9} & \frac{2}{9} & \frac{3}{9} & \frac{2}{9} & \frac{1}{9} \end{bmatrix} \quad (36)$$

and the unconditional distribution of $x_i - x_j$ (the column sums of $\bar{Q}$) is

$$\frac{1}{27}\begin{bmatrix} \frac{\alpha}{(1-\alpha)} & \frac{4\alpha}{(1-\beta)} & \frac{(3-2\alpha)}{(1-\alpha)} & \frac{4(1+\beta-\alpha)}{(1-\alpha)} & \frac{(5-\beta)}{(1-\beta)} & \frac{4(1+\beta-\alpha)}{(1-\alpha)} & \frac{(3-2\alpha)}{(1-\alpha)} & \frac{4\alpha}{(1-\beta)} & \frac{\alpha}{(1-\alpha)} \end{bmatrix}. \quad (37)$$

It can immediately be seen that the distribution of x for an individual firm (given by (36)) is no longer uniform. It can also be seen from the definition of $\bar{Q}$ that in general there is no simple relationship between the position of x_i and the conditional distribution of $x_i - x_j$.

A pattern does emerge, however, if the degree of idiosyncratic shocks is reduced to zero. In this case the solution for $\bar{Q}$ is not unique but the insights gained will also apply for the case where idiosyncrasies are very small. Taking the limit of (37) as S tends to 2 yields the following for the unconditional distribution of $x_i - x_j$

$$\begin{bmatrix} 0 & 0 & \tfrac{1}{9} & \tfrac{2}{9} & \tfrac{1}{3} & \tfrac{2}{9} & \tfrac{1}{9} & 0 & 0 \end{bmatrix}. \tag{38}$$

This immediately reveals a very important property of the two-sided model, namely, that there is a zero probability of $x_i - x_j$ being greater than 2 or less than -2 when there are no idiosyncratic shocks. Since firm i and firm j can be any two firms it follows that all firms must have xs which are within a range $+/-2$ of each other. Thus firms must be clustered with $x = \{-2, -1, 0\}$ or $x = \{-1, 0, 1\}$ or $x = \{0, 1, 2\}$.

Notice that, as in the previous case, $\bar{Q}$ can be used to say things about the cross-sectional distribution of firms. Further insights into the cross-sectional distribution can be gained by considering the conditional distribution of $x_i - x_j$ for each value of x_i. When $x_i = -2$ the conditional distribution of $x_i - x_j$ is[20]

$$\begin{bmatrix} 0 & 0 & \tfrac{1}{3} & \tfrac{1}{3} & \tfrac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{39}$$

In other words, there is 1/3 probability of $x_j = x_i$. When $x_i = -1$ the conditional distribution is

$$\begin{bmatrix} 0 & 0 & \tfrac{1}{6} & \tfrac{1}{3} & \tfrac{1}{3} & \tfrac{1}{6} & 0 & 0 & 0 \end{bmatrix} \tag{40}$$

so again there is 1/3 probability of $x_j = x_i$. And finally when $x_i = 0$ the conditional distribution is

$$\begin{bmatrix} 0 & 0 & \tfrac{1}{9} & \tfrac{2}{9} & \tfrac{1}{3} & \tfrac{2}{9} & \tfrac{1}{9} & 0 & 0 \end{bmatrix} \tag{41}$$

which also shows 1/3 probability of $x_j = x_i$. (The same result holds for $x_i = 1$ and $x_i = 2$.) Thus, whatever level of x_i is chosen there is always 1/3 probability that $x_j = x_i$. This result, combined with the previous result that firms are clustered within a range $+/-2$ of each other, suggests that firms are in fact uniformly distributed on that range. Thus firms are either uniformly distributed on the range $\{-2, -1, 0\}$ or $\{-1, 0, 1\}$ or $\{0, 1, 2\}$ (with each of these distributions occurring with probability 1/3).[21] It is simple to check that this conjectured behaviour for the cross-sectional distribution is consistent with the conditional distributions listed above.

As was emphasized above, these results are obtained by assuming that there are no idiosyncratic shocks, i.e. $S = 2$. But in this case the solution for

$\bar{Q}$ is not unique. There are a number of different $\bar{Q}$s, each corresponding to a different initial distribution of firms. It is easy to see, however, that the cross-sectional distribution described above is only slightly changed by the introduction of small idiosyncrasies. For instance for $S=1.99$ the probability of $x_i - x_j$ being outside the range $+/-2$ is less than 0.2 per cent. Thus the cross-sectional distribution will still involve 99.8 per cent of firms clustered with the range $+/-2$ of each other.

It is easy to check that the cross-sectional distribution described above is a steady-state by considering a few examples of shocks. Suppose the cross-sectional distribution of firms is given by

$$[0 \ \tfrac{1}{3} \ \tfrac{1}{3} \ \tfrac{1}{3} \ 0]. \tag{42}$$

If there is a high money growth shock (and assuming there are no idiosyncrasies) all firms will find that their xs will fall by 1. Thus the new distribution is given by

$$[\tfrac{1}{3} \ \tfrac{1}{3} \ \tfrac{1}{3} \ 0 \ 0]. \tag{43}$$

If there is a further high money growth shock the 1/3 of firms with $x=-2$ will rest their prices so that $x=0$. The other 2/3 of firms will simply let x fall by 1. It is clear that this leaves the cross-sectional distribution unchanged. A low money growth shock would cause x to rise by 1 for all firms and shift the cross-sectional distribution to the right. This can continue until the rightmost group of firms hits the upper barrier and resets x to 0. This again does not affect the shape of the distribution.

It is now possible to consider the implications of this model for the neutrality of money. (24) states that output is determined by the average of the xs. The average of the xs in the case considered in this section is given by the mid-point of the cluster of firms. Thus in (42) the average is 0. The positive monetary shock which shifts the distribution from (42) to (43) shifts the average of the xs from 0 to -1. Thus the monetary shock causes an expansion of output. Any further positive monetary shocks, however, do not produce any further reduction in the average x so output does not increase beyond 1. Negative shocks to money shift the cluster of firms to the right and thus increase the average x and reduce output. This can continue until the cluster of firms hits the upper barrier. At this point further negative monetary shocks become neutral.

The overall result is that the two-sided model displays non-neutralities at the aggregate level. The neutrality displayed by the one-sided model only becomes apparent when the cluster of firms hits the upper or lower barrier. And even then monetary shocks are only neutral in one direction.

Conclusion

This chapter has summarized a number of the main models and results in the literature on menu costs and aggregate price dynamics. The main results that emerge from this literature are as follows:

1. The trigger point strategy shown to be optimal in a static framework has its equivalent in a dynamic framework for an individual firm that maximizes the discounted value of profits in the face of serially correlated shocks. Such a firm holds its nominal price constant while the deviation of its actual price from the optimal (frictionless) price is within an optimally determined range. When shocks get sufficiently extreme the actual price is adjusted discretely and then again held constant.

2. If the shocks hitting the economy are all in one direction (so, for instance, optimal prices never fall) it is found that the nominal rigidity at the individual firm level disappears at the aggregate level. This is because the cross-sectional distribution of firms is such that there are always some firms that are close to a price adjustment trigger.

3. If, however, shocks can be both positive and negative (so prices may fall and rise) it is found that menu costs can give rise to aggregate price rigidity. In this case the cross-sectional distribution of firms is more concentrated so there will not necessarily be firms close to a trigger point.

This chapter demonstrates each of these three results in turn in models which are generally representative of the literature. It is apparent from the descriptions given that these models are highly restricted and stylized. This is particularly true in the aggregate model where many simplifying assumptions are necessary to allow progress to be made. One of the main criticisms that can be made of this literature is that the models are in fact so restricted that they have no direct grounding in the microeconomics of price-setting. There is nothing in the aggregate model described in this chapter, for instance, that distinguishes it as a model of price-setting in particular. It could, with minimal alteration, be regarded as a model of any variable that is subject to lump sum costs of adjustment (e.g. employment or inventories).

One of the main avenues of future research in this area must therefore be to build in more of the microeconomic foundations that are appropriate to a price-setting problem. Thus, for instance, in the aggregate model it would be interesting to incorporate explicitly the optimizing decisions of firms. In particular it would be productive to base this optimizing framework on a profit function that is explicitly derived from a model of an imperfectly competitive firm. Such a framework would allow many important issues to be explored. For instance, it would be possible to consider explicitly the interaction between menu costs and the degree of market power of an individual firm and the implications of this interaction for aggregate price dynamics. Caballero and Engel (1993) explore some issues related to this point. A second line of inquiry could be to investigate the interaction between menu costs and real rigidities, thus extending the work of Ball and Romer (1990) to a dynamic aggregate setting.

Notes

1. Caplin (1993) also surveys the main results in this literature. The purpose of this chapter, in contrast to Caplin's, is to provide a sufficiently detailed derivation of the key results to allow readers to pursue further work in this area. Consequently the emphasis in this chapter is on technical description rather than on discussion or assessment.
2. Ideally the aggregate framework would be built as an extension of the model of the individual optimizing firm presented in the first section. This, however, would be very complex, so this chapter follows the pattern of many other studies in this area by suppressing the optimizing decisions of firms when considering the aggregate price level. This effectively means that the second–fourth sections of the chapter are independent of the material presented in the first section.
3. The analysis in this section is technically quite difficult. However, as has already been pointed out in n. 2, the second–fourth sections of this chapter are largely independent of the material contained in this section. Readers who are not interested in the technical details of the individual firm's case may therefore wish to go directly to p.344 where the aggregate analysis begins.
4. It is also possible to model the two-sided problem using Brownian motion. This is, in some ways, simpler than the compound Poisson process, but it is not possible to generate a single-sided equivalent. A simple exposition of the optimal adjustment rule in the Brownian motion case can be found in Dixit (1991, 1993) and Bertola and Caballero (1990). Dixit (1991) generates an analytical approximation to the solution of the firm's problem in the presence of Brownian motion shocks and shows that a fourth order menu cost can generate first order price stickiness.
5. For example suppose the firm faces a linear demand function of the form $q_t = \psi_t - p_t/2$ where q_t is quantity demanded and ψ_t is a demand shock variable. Assume that marginal costs are constant at 2c. The profit maximizing price for this firm in absence of menu costs is $p_t^* = \psi_t + c$. If the demand shock is assumed to follow a compound Poisson process then p^* will also follow such a process.
6. Sheshinski and Weiss in fact use a more general formulation than (1) in their model, but only in the context of a single-sided shock. The more general formulation has the advantage that prices follow a continuous process (unlike in (1)) but this comes at the expense of greater complexity.
7. W will not be a direct function of time because the probability structure of the x process is not a function of time and the trigger and return points are not functions of time (by assumption).

8. From the definition of W in (5)

$$V(x_t)=E_t\left\{\int_t^{\infty}[-bx_\tau^2-C_\tau]e^{-\delta(\tau-t)}d\tau\right\}.$$

The integral on the right can be split into two parts as follows

$$V(x_t)=E_t\left\{\int_t^{t+dt}[-bx_\tau^2-C_\tau]e^{-\delta(\tau-t)}d\tau\right.$$

$$\left.+e^{-\delta dt}\int_{t+dt}^{\infty}[-bx_\tau^2-C_\tau]e^{-\delta(\tau-t-dt)}d\tau\right\}$$

which can then be rewritten as (6).

9. The interval of time, dt, is assumed to be infinitesimally small so terms of order $(dt)^2$ can be treated as being equal to zero.
10. The basic framework described here is an adaptation of similar structures in, among others, Bertola and Caballero (1990), Caballero (1992) and Caplin (1993).
11. In common with all menu cost models it is assumed that firms have some degree of monopoly power. This is necessary if firms are to be price-setters. In an imperfectly competitive market one important determinant of a firm's optimal price is the level of prices set by other firms. In order to keep the analysis simple this cross-effect between firms is ignored in the framework being outlined here. Caplin and Leahy (1991) make a similar assumption, and show that it amounts to setting to unity the elasticity with respect to the aggregate price level of each individual firm's profit function.
12. The 'cross-sectional distribution' is defined as the observed proportion of the total population of firms at each value of x in a given period. Thus if x can take on values over the range -2 to 2, then the cross-sectional distribution would be a vector $[c_1, c_2, c_3, c_4, c_5]$ where c_1 is the proportion of firms with $x=-2$, c_2 is the proportion of firms with $x=-1$, and so on for c_3, c_4 and c_5. The average of x over all firms would thus be given by $(-2\times c_1)+(-1\times c_2)+(0\times c_3)+(1\times c_4)+(2\times c_5)$.
13. Note that the assumption of a large total population of firms is maintained throughout this and the following section.
14. The steady-state Q could also be described as the 'unconditional distribution', in the sense that it is not conditional on information at time 0. To avoid confusion, however, the term 'unconditional' is not used for this purpose in this chapter.
15. In the one-sided case it is not strictly necessary to consider x_i-x_j. However, in the two-sided case it turns out that cross-sectional distribution does not possess a steady-state over the absolute values of the xs whereas the distribution of

firms relative to each other does reach a steady-state. In that case it is therefore necessary to consider the probability distribution of $x_i - x_j$.

16. To be precise, the term 'unconditional distribution' in this context means the distribution of x_i unconditional on the value of x_j.
17. This can be justified on the basis of the Glivenko–Cantelli theorem. See Bertola and Caballero (1990) and Caballero (1992) for some further discussion of this point.
18. Given x_i the probability distribution of $x_i - x_j$ (where firm j is chosen at random from the population of firms) must be determined by the cross-sectional distribution of firms.
19. Caballero and Engel (1991, 1993) analyse the transition phase and its specific implications for macroeconomic policy.
20. This is obtained from the first row of $\bar{Q}$ by taking the limit as S tends to 2 and dividing by the unconditional probability of $x_t = -2$ (i.e. 1/9).
21. The alternating nature of the cross-sectional distribution reveals the fact that there is no steady-state in the two-sided case in terms of the absolute position of the firms. However, as is apparent, firms do reach a steady-state distribution relative to each other, hence the need to approach the problem by considering the distribution of $x_i - x_j$.

Bibliography

Abreu, D., 1988. 'Towards a Theory of Discounted Infinitely-Repeated Games', *Econometrica* 56, pp.383–96

Aitchison, J. and Brown, J.A.C., 1957. *The Lognormal Distribution*, New York: Cambridge University Press

Akerlof, G.A., 1970. 'The Market for Lemons: Quality Uncertainty and the Market Mechanism'. *Quarterly Journal of Economics* 84, pp.488–500

Akerlof, G.A. and Yellen, J.L., 1985a. 'A Near-Rational Model of the Business Cycle with Wage and Price Inertia', *Quarterly Journal of Economics* 100 (Supp.), pp.823–38

1985b. 'Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?', *American Economic Review* 75, pp.708–21

Andersen, T.M., 1994. *Price Rigidity – Causes and Macroeconomic Implications*, Oxford: Oxford University Press

Andersen, T.M. and Hviid, M., 1990. 'Price Rigidities in Dynamic Models of Imperfect Competition', in N.M. Christodoulakis (ed.), *Dynamic Modelling and the Control of National Economies 1989*, Oxford: Pergamon

1993. 'Strategic Implications of Private Information for Price Adjustment', Institute of Economics, University of Aarhus, mimeo

1994. 'Acquisition and Dissemination of Information in Imperfectly Competitive Markets', *Economic Inquiry* 32, pp.498–510

Arnsperger, C. and de la Croix, D., 1993. 'Bargaining and Equilibrium Unemployment: Narrowing the Gap Between New Keynesian and Disequilibrium Theories', *European Journal of Political Economy* 9, pp.163–90

Arrow, K.J., 1959. 'Towards a Theory of Price Adjustment', in M. Abramowitz, (ed.), *The Allocation of Economic Resources*, Stanford: Stanford University Press

1963. 'Uncertainty and the Welfare Economics of Medical Care', *American Economic Review* 53, pp.941–73

1968. 'The Economics of Moral Hazard: Further Comment', *American Economic Review* 58, pp.537–9

1969. 'The Organisation of Economic Activity', in *The Analysis and Evaluation of Public Expenditures: The PPB System*, Joint Economic Committee, 91st

Congress, First Session, Washington, DC

Arrow, K.J. and Hahn, F.H., 1971. *General Competitive Analysis*, San Francisco: Holden-Day

Azariadis, C., 1975. 'Implicit Contracts and Underemployment Equilibria', *Journal of Political Economy* 83, pp.1183–1202

Baily, M.N., 1974. 'Wages and Employment Under Uncertain Demand', *Review of Economic Studies* 41, pp.37–50

Ball, L., 1988. 'Is Equilibrium Indexation Efficient?', *Quarterly Journal of Economics* 103, pp.299–311

Ball, L. and Romer, D., 1987. 'Sticky Prices as Coordination Failure', NBER, *Working Paper* 2327

1989a. 'Are Prices Too Sticky?', *Quarterly Journal of Economics* 104, pp.507–24

1989b. 'The Equilibrium and Optimal Timing of Price Changes', *Review of Economic Studies* 56, pp.179–98

1990. 'Real Rigidities and the Non-Neutrality of Money', *Review of Economic Studies* 57, pp.183–203

1991. 'Sticky Prices as Coordination Failure', *American Economic Review* 81, pp.539–52

Ball, L., Mankiw, N.G. and Romer, D., 1988. 'The New Keynesian Economics and the Output–Inflation Trade-Off', *Brookings Papers on Economic Activity* 1, pp.1–82

Ball, R. and Bodkin, R., 1963. 'Income, the Price Level and Generalised Multipliers in a Keynesian Economy', *Metroeconomica* 15, pp.59–81

Barnett, A.H., 1980. 'The Pigouvian Tax Rule Under Monopoly', *American Economic Review* 70, 1037–41

Barro, R.J. and Grossman, H.I., 1971. 'A General Disequilibrium Model of Income and Employment', *American Economic Review* 61, pp.82–93

1976. *Money, Employment and Inflation*, Cambridge: Cambridge University Press

Baumol, W. and Oates, W., 1975. *The Theory of Environmental Policy*, Cambridge: Cambridge University Press

Baye, M. and Cosimano, T., 1990. 'Choosing Sides in Matching Games: Nash Equilibria and Comparative Statics', *Economica* 57, pp.283–94

Bean, C., 1984. 'Targeting Nominal Income: An Appraisal', *Economic Journal* 93, pp.806–19

Bénassy, J.-P., 1973. 'Disequilibrium Theory', Ph.D. dissertation, Department of Economics, University of California at Berkeley

1975. 'Neo-Keynesian Disequilibrium Theory in a Monetary Economy', *Review of Economic Studies* 42, pp.503–23

1976. 'A Disequilibrium Approach to Monopolistic Price Setting and General Monopolistic Equilibrium', *Review of Economic Studies* 43, pp.69–81

1978. 'A Neo-Keynesian Model of Price and Quantity Determination in Disequilibrium', in G. Schwodiauer (ed.), *Equilibrium and Disequilibrium in Economic Theory*, Dordrecht: Reidel

1982. *The Economics of Market Disequilibrium*, New York: Academic Press

1987. 'Imperfect Competition, Unemployment and Policy', *European Economic Review* 31, pp.417–26
1988. 'The Objective Demand Curve in General Equilibrium with Price Makers', *Economic Journal* 98 (Supp.), pp.37–49
1990. 'Non-Walrasian Equilibria, Money and Macroeconomics', in B. Friedman and F.H. Hahn (eds.), *Handbook of Monetary Economics*, Amsterdam: North-Holland
1991a. 'Microeconomic Foundations and Properties of a Macroeconomic Model with Imperfect Competition', in K.J. Arrow (ed.), *Issues in Contemporary Economics Volume 1: Markets and Welfare*, London: Macmillan
1991b. 'Optimal Government Policy in a Macroeconomic Model with Imperfect Competition and Rational Expectations', in W. Barnett *et al.* (eds.), *Equilibrium Theory and Applications*, Cambridge: Cambridge University Press
1991c. 'Monopolistic Competition', in W. Hildenbrand and H. Sonnenschein (eds.), *Mathematical Economics*, vol. 4, Amsterdam: North-Holland
1993. 'Non-Clearing Markets: Microeconomic Concepts and Macroeconomic Applications', *Journal of Economic Literature* 31, pp.732–61
Bennett, J.S., 1990. *A Theory of Central Planning*, Oxford: Blackwell
Bennett, J.S. and Dixon, H.D., 1993. 'The Macroeconomics of a Transitional Economy', *European Economic Review*, forthcoming
Bental, B. and Eckstein, Z., 1990. 'The Dynamics of Inflation With Constant Deficit and Expected Regime Change', *Economic Journal* 100, pp.1245–60
Bentolila, S., 1988. *Labor Demand and Adjustment Costs*, Ph.D. Dissertation, Cambridge, MA: MIT
Bentolila, S. and Bertola, G., 1990. 'Firing Costs and Labor Demand: How Bad is Eurosclerosis?', *Review of Economic Studies* 57, pp.381–402
Bentolila, S. and Saint-Paul, G., 1994. 'A Model of Labor Demand With Linear Adjustment Costs', *Labour Economics*, forthcoming
Bernanke, B.S., 1983. 'Non-Monetary Effects of the Financial Crisis in the Propagation of the Great Depression', *American Economic Review* 73, pp.257–76
Bernanke, B.S. and Gertler, M., 1989. 'Agency Costs, Net Worth and Business Fluctuations', *American Economic Review* 79, pp.14–31
1990. 'Financial Fragility and Economic Performance', *Quarterly Journal of Economics* 105, pp.87–114
Bertola, G. and Caballero, R., 1990. 'Kinked Adjustment Costs and Aggregate Dynamics', *NBER Macroeconomics Annual* 5, pp.237–88
Bester, H., 1985. 'Screening vs. Rationing in Credit Markets With Imperfect Information', *American Economic Review* 75, pp.850–5
Bils, M., 1987. 'The Cyclical Behaviour of Marginal Cost and Price', *American Economic Review* 77, pp.838–55
Blanchard, O.J. and Diamond, P., 1989. 'The Beveridge Curve', *Brookings Papers on Economic Activity*, 1, pp.1–60
Blanchard, O.J. and Fischer, S., 1989. *Lectures on Macroeconomics*, Cambridge, MA: MIT Press

Blanchard, O.J. and Kiyotaki, N., 1987. 'Monopolistic Competition and the Effects of Aggregate Demand', *American Economic Review* 77, pp.647–66

Blanchard, O.J. and Summers, L., 1986. 'Hysteresis and the European Unemployment Problem', *NBER Macroeconomics Annual*, 1, pp.15–78

Blinder, A.S., 1987. 'Credit Rationing and Effective Supply Failures', *Economic Journal* 97, pp.327–52

Bohm, P., 1987. 'External Economies', in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, Macmillan: London

Bohn, H. and Gorton, G., 1993. 'Coordination Failure, Multiple Equilibria and Economic Institutions', *Economica* 60, pp.257–80

Bovenberg, A. L. and van der Ploeg, F., 1992. 'Environment Policy, Public Finance and the Labour Market in a Second-Best World', CEPR, *Discussion Paper* 745, London: CEPR

1993. 'Green Policies in a Small Open Economy', CEPR, *Discussion Paper* 785, London: CEPR

Boyd, J. and Prescott, E., 1986. 'Financial-Intermediary Coalitions', *Journal of Economic Theory* 38, pp.211–32

Bresnahan, T.F., 1989. 'Empirical Studies of Industries with Market Power', in R. Schmalensee and R.D. Willig (eds.), *Handbook of Industrial Organisation*, vol. 2, Amsterdam: North-Holland

Bruno, M., 1989. 'Econometrics and the Design of Economic Reform', *Econometrica* 57, pp.275–306

Bryant, J., 1983. 'A Simple Rational Expectations Keynes-Type Model', *Quarterly Journal of Economics* 98, pp.525–9

Buchanan, J.M., 1969. 'External Diseconomies, Correction Taxes and Market Structure', *American Economic Review* 59, pp.174–7

Buchanan, J.M. and Stubblebine, W., 1962. 'Externality', *Economica* 29, pp.371–84

Bulow, J.I., Geanakoplos, J.D. and Klemperer, P.D., 1985. 'Multimarket Oligopoly: Strategic Substitutes and Complements', *Journal of Political Economy* 93, pp.488–511

Caballero, R., 1992. 'A Fallacy of Composition', *American Economic Review* 82, pp.1279–92

Caballero, R. and Engel, E., 1991. 'Dynamic S-s Economies', *Econometrica* 59, pp.1659–86

1993. 'Heterogeneity and Output Fluctuations in a Dynamic Menu-Cost Economy', *Review of Economic Studies* 60, pp.95–119

Cagan, P., 1956. 'The Monetary Dynamics of Hyperinflation', in M. Friedman (ed.), *Studies in the Quantity Theory of Money*, Chicago: University of Chicago Press

Calmfors, L. and Driffill, J., 1988. 'Bargaining Structure, Corporatism and Macroeconomic Performance', *Economic Policy* 6, pp.13–61

Caplin, A., 1993. 'Individual Inertia and Aggregate Dynamics', in E. Sheshinski and Y. Weiss (eds.), *Optimal Pricing, Inflation and the Cost of Price Adjustment*, Cambridge, MA: MIT Press

Caplin, A. and Leahy, J., 1991. 'State-Dependent Pricing and the Dynamics of

Money and Output', *Quarterly Journal of Economics* 106, pp.683–708

Caplin, A. and Spulber, D., 1987. 'Menu Costs and the Neutrality of Money', *Quarterly Journal of Economics* 102, pp.703–26

Card, D., 1983. 'Cost of Living Estimation in Major Union Contracts', *Industrial and Labor Relations Review* 37, pp.34–48

Carmichael, L., 1985. 'Can Unemployment Be Involuntary? Comment', *American Economic Review* 75, pp.1213–14

Chatterjee, S. and Cooper, R., 1989. 'Multiplicity of Equilibria and Fluctuations in Dynamic Imperfectly Competitive Economies', *American Economic Review* 79 (Papers and Proceedings), pp.353–7

Clower, R.W., 1965. 'The Keynesian Counter-Revolution: A Theoretical Reappraisal', in F.H. Hahn and F.P.R. Brechling (eds.), *The Theory of Interest Rates*, London: Macmillan

Colander, D., 1986. *Macroeconomics*, Glenview, IL: Scott-Foresman

Conlisk, J., 1988. 'Optimisation Cost', *Journal of Economic Organisation and Behaviour* 9, pp.213–28

Cooper, R., 1986. 'Optimal Labour Contracts, Imperfect Competition and Underemployment Equilibria: A Framework for Analysis', NBER, *Working Paper* 2060

1987. 'Dynamic Behaviour of Imperfectly Competitive Economies with Multiple Equilibria', NBER, *Working Paper* 2388

1988a. 'Will Share Contracts Increase Economic Welfare?', *American Economic Review* 78, pp.138–54

1988b. 'Labor Contracts and the Role of Monetary Policy in an Overlapping-Generations Model', *Journal of Economic Theory* 44, pp.231–50

1990. 'Predetermined Wages and the Impact of Expansionary Government Policy', *Review of Economic Studies* 57, pp.205–14

Cooper, R. and Haltiwanger, J., 1990. 'Inventories and the Propagation of Sectoral Shocks', *American Economic Review* 80, pp.170–90

Cooper, R. and John, A., 1988. 'Coordinating Coordination Failures in Keynesian Models', *Quarterly Journal of Economics* 83, pp.441–63

Cropper, M. and Oates, W., 1992. 'Environmental Economics: A Survey', *Journal of Economic Literature* 30, pp.675–740

Danthine, J.P. and Donaldson, J., 1990. 'Efficiency Wages and the Real Business Cycle Puzzle', *European Economic Review* 34, pp.1275–1301

d'Aspremont, C., Dos Santos Ferreira, R. and Gérard-Varet, L.-A., 1984. 'Oligopoly and Involuntary Unemployment', CORE, *Discussion Paper* 8408, Louvain-la-Neuve: Catholic University of Louvain

1989a. 'Unemployment in a Cournot Oligopoly Model with Ford Effects', *Recherches Economiques de Louvain* 55, pp.33–60

1989b. 'Unemployment in an Extended Cournot Oligopoly Model', *Oxford Economic Papers* 41, pp.490–505

1990. 'On Monopolistic Competition and Involuntary Unemployment', *Quarterly Journal of Economics* 85, pp.895–919

1991a. 'Imperfect Competition, Rational Expectations and Unemployment', in

W. Barnett *et al.* (eds.), *Equilibrium Theory and Applications*, Cambridge: Cambridge University Press

1991b. 'Pricing Schemes and Cournotian Equilibria', *American Economic Review* 81, pp.666–73

1992. 'General Equilibrium Concepts Under Imperfect Competition: A Cournotian Approach', CORE, *Discussion Paper* 9257, Louvain-la-Neuve: Catholic University of Louvain

1994. 'On the Dixit–Stiglitz Model of Monopolistic Competition', CORE, *Discussion Paper* 9468, Louvain-la-Neuve: Catholic University of Louvain; forthcoming in *American Economic Review*

Davidson C., Martin, L. and Matusz, S., 1987. 'Search, Unemployment and the Production of Jobs', *Economic Journal* 97, pp.857–76

1988. 'The Structure of Simple General Equilibrium Models with Frictional Unemployment', *Journal of Political Economy* 96, pp.1267–93

1994. 'Jobs and Chocolate: Samuelsonian Surpluses in Dynamic Models of Unemployment', *Review of Economic Studies* 61, pp.173–92

Davidson, P., 1972. *Money in the Real World*, London: Macmillan

Davis, O. and Whinston, A., 1962. 'Externalities, Welfare and the Theory of Games', *Journal of Political Economy* 70, pp.881–894

Dechert, W.D. and Yamamoto, K., 1992. 'Asset Valuation and Production Efficiency in an Overlapping-Generations Model With Production Shocks', *Review of Economic Studies* 59, pp.389–405

Dehez, P., 1985. 'Monopolistic Equilibrium and Involuntary Unemployment', *Journal of Economic Theory* 36, pp.160–5

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J., 1990. 'Positive Feedback Investment Strategies and Destabilising Rational Expectations', *Journal of Finance* 44, pp.379–96

De Meza, D. and Webb, D.C., 1987. 'Too Much Investment: A Problem of Asymmetric Information', *Quarterly Journal of Economics* 102, pp.281–92

Dertouzos, J. and Pencavel, J., 1984. 'Wage-Unemployment Determination Under Trade Unionism: The International Typographical Union', *Journal of Political Economy* 89, pp.1162–81

Diamond, D., 1984. 'Financial Intermediation and Delegated Monitoring', *Review of Economic Studies* 51, pp.393–414

Diamond, P.A., 1965. 'National Debt in a Neoclassical Growth Model', *American Economic Review* 55, pp.503–11

1982. 'Aggregate Demand Management in Search Equilibrium', *Journal of Political Economy* 90, pp.881–94

Diamond, P.A. and Fudenberg, D., 1989. 'Rational Expectations Business Cycles in Search Equilibrium', *Journal of Political Economy* 97, pp.606–19

Diebold, F.X. and Rudebusch, G.D., 1990. 'A Nonparametric Investigation of Duration Dependence in the American Business Cycle', *Journal of Political Economy* 98, pp.596–616

Dixit, A., 1991. 'Analytical Approximations in Models of Hysteresis', *Review of Economic Studies* 58, 141–51

1993. 'A Simplified Treatment of the Theory of Optimal Regulation of Brownian Motion', in E. Sheshinski and Y. Weiss (eds.), *Optimal Pricing, Inflation and the Cost of Price Adjustment*, Cambridge, MA: MIT Press

Dixit, A. and Stiglitz, J. 1977. 'Monopolistic Competition and Optimum Product Diversity', *American Economic Review* 67, pp.297–308

Dixon, H., 1987. 'A Simple Model of Imperfect Competition with Walrasian Features', *Oxford Economic Papers* 39, pp.134–60

1988. 'Unions, Oligopoly and the Natural Range of Employment', *Economic Journal* 98, pp.1127–47

1990a. 'Macroeconomic Policy with a Floating Exchange Rate and a Unionised Non-Traded Sector', *Economic Journal* 100 (Supp.), pp.78–90

1990b. 'Imperfect Competition, Unemployment Benefit and the Non-Neutrality of Money: An Example in the Spirit of Hart', *Oxford Economic Papers* 42, pp.402–13

1991. 'Macroeconomic Equilibrium and Policy in a Large Unionised Economy', *European Economic Review* 35, pp.1427–48

1992. 'Nominal Wage Flexibility in a Partly-Unionised Economy', *The Manchester School* 60, pp.295–306

1994a. 'Imperfect Competition and Open-Economy Macroeconomics', in F. van der Ploeg (ed.), *Handbook of International Macroeconomics*, Oxford: Blackwell

1994b. 'Macroeconomic Price and Quantity Responses with Heterogeneous Product Markets', *Oxford Economic Papers* 46, pp.385–402

1994c. 'Of Coconuts, Decomposition and a Jackass: The Genealogy of the Natural Rate', in R. Cross (ed.), *The Natural Rate of Unemployment*, Cambridge: Cambridge University Press, forthcoming

Dixon, H. and Lawler, P., 1993. 'Fiscal Policy in a Monopolistic Economy', *Discussion Paper* 1993/17, Department of Economics, York University

Dixon, H. and Rankin, N., 1991. 'Macroeconomics and Imperfect Competition: A Survey' (long version), CEPR, *Discussion Paper* 636, London: CEPR

Domowitz, I., Hubbard, G. and Petersen, B., 1988. 'Market Structure and Cyclical Fluctuations in U.S. Manufacturing', *Review of Economics and Statistics* 70, pp.55–66

Dornbusch, R., 1987. 'Exchange Rates and Prices', *American Economic Review* 77, pp.93–106

Dow, J.C.R., 1964. *The Management of The British Economy 1945–60*, Cambridge: Cambridge University Press

Dowd, K., 1993. 'Optimal Financial Contracts', *Oxford Economic Papers* 44, pp.672–93

Drazen, A., 1985. 'A Rational Expectations Keynesian Model with Price Flexibility', *Working Paper* 2–85, Foerder Institute for Economic Research, Tel-Aviv University

1986. 'Involuntary Unemployment and Aggregate Demand Spillovers in an Optimal Search Model', *Working Paper* 32–86, Foerder Institute for Economic Research, Tel-Aviv University

Dubey, P., 1980. 'Nash Equilibria of Market Games: Finiteness and Efficiency', *Journal of Economic Theory* 22, pp.363–76

Dubey, P., Mas-Colell, A. and Shubik, M., 1980. 'Efficiency Properties of Strategic Market Games: An Axiomatic Approach', *Journal of Economic Theory* 22, pp.339–62

Engel, E., 1992. 'A Fallacy of Composition', *American Economic Review* 82, pp.1279–92

Espinosa, M.P. and Rhee, C., 1989. 'Efficient Wage Contracts as a Repeated Game', *Quarterly Journal of Economics* 104, pp.565–88

Evans, G.W., 1985. 'Expectational Stability and the Multiple Equilibrium Problem in Linear Rational Expectations Models', *Quarterly Journal of Economics* 100, pp.1217–33

Evans, G.W. and Ramey, G., 1992a. 'Expectation Calculation and Macroeconomic Dynamics', *American Economic Review* 82, pp.207–24

1992b. 'Expectation Calculation, Hyperinflation and Currency Collapse', Financial Markets Group, *Discussion Paper* 141, London School of Economics

Farmer, R.E.A., 1984. 'A New Theory of Aggregate Supply', *American Economic Review* 74, pp.920–30

1985. 'Implicit Contracts With Asymmetric Information and Bankruptcy: The Effect of Interest Rates on Layoffs', *Review of Economic Studies* 52, pp.427–42

Farrell, J., 1988. 'Puzzles: Sylvia, Ice Cream and More', *Journal of Economic Perspectives* 2, pp.175–82

Farrell, J. and Maskin, E., 1989. 'Renegotiation in Repeated Games', *Games and Economic Behaviour* 1, pp.327–60

Fazzari, S.M. and Petersen, B.C., 1993. 'Working Capital and Fixed Investment: New Evidence on Financing Constraints', *Rand Journal of Economics* 24, pp.328–42

Fazzari, S.M., Hubbard, R.G. and Petersen, B.C., 1988. 'Financing Constraints and Corporate Investment', *Brookings Papers on Economic Activity* 1, pp.141–96

Fender, J. and Yip, C.Y., 1993. 'Monetary Policies in an Intertemporal Macroeconomic Model with Imperfect Competition', *Journal of Macroeconomics* 15, pp.439–53

Fischer, S., 1977. 'Long-Term Contracts, Rational Expectations and the Optimal Money Supply Rule', *Journal of Political Economy* 85, pp.163–90

1984. 'The Economy of Israel', in K. Brunner and A.H. Meltzer (eds.), *Monetary and Fiscal Policy and Their Applications*, Supplement to *Journal of Monetary Economics* 20, pp.7–52

1987. 'Comment on Thomas Sargent and Neil Wallace, "Inflation and the Government Budget Constraint"', in A. Razin and E. Sadka (eds.), *Economic Policy in Theory and Practice*, London: Macmillan

Fisher, I., 1933. 'The Debt–Deflation Theory of Great Depressions', *Econometrica* 1, pp.337–57

Flood, R.P. and Garber, P.M. (1980a) 'An Economic Theory of Monetary Reform', *Journal of Political Economy* 88, pp.24–58

1980b. 'Market Fundamentals vs. Price Level Bubbles: The First Test', *Journal of*

Political Economy 88, pp.24–58
Frank, J., 1986. *The New Keynesian Economics: Unemployment, Search and Contracting*, Brighton: Wheatsheaf
1989. 'Equilibrium Recessions and Keynesian Economics', *Economies et Sociétés*, F31, pp.219–32
1990. 'Monopolistic Competition, Risk Aversion and Equilibrium Recessions', *Quarterly Journal of Economics* 85, pp.921–38
1994, 'Competitive Equilibrium Recessions', Birkbeck College, London, mimeo
Froot, K.A. and Klemperer, P.D., 1989. 'Exchange Rate Pass-Through When Market Share Matters', *American Economic Review* 79, pp.637–54
Fudenberg, D. and Maskin, E., 1986. 'The Folk Theorem in Infinitely-Repeated Games With Discounting or With Imperfect Information', *Econometrica* 54, pp.533–66
Gabel, H.L. and Sinclair-Desgagné, B., 1993. 'Managerial Incentives and Environmental Compliance', *Journal of Environmental Economics and Management* 24, pp.229–40
Gabszewicz, J.J. and Vial, J.-P., 1972. 'Oligopoly 'à la Cournot' in General Equilibrium Analysis', *Journal of Economic Theory* 4, pp.381–400
Gale, D., 1978. 'A Note on Conjectural Equilibrium', *Review of Economic Studies* 45, pp.33–8
Gale, D. and Hellwig, M., 1985. 'Incentive-Compatible Debt Contracts I: The One-Period Problem', *Review of Economic Studies* 52, pp.647–64
Gertler, M., 1988. 'Financial Structure and Aggregate Economic Activity: An Overview', *Journal of Money, Credit and Banking* 20, pp.559–88
1992. 'Financial Capacity and Output Fluctuations in an Economy With Multi-Period Financial Relationships', *Review of Economic Studies* 59, pp.455–72
Gertler, M. and Gelchrist, S., 1994. 'Monetary Policy, Business Cycles and the Behaviour of Small Manufacturing Firms', *Quarterly Journal of Economics* 109, pp.309–40
Giovannini, A., 1988. 'Exchange Rates and Traded Goods Prices', *Journal of International Economics* 24, pp.45–68
Gordon, R.J., 1990. 'What is New Keynesian Economics?', *Journal of Economic Literature* 28, pp.1115–71
Graaff, J. de V., 1987. 'Pecuniary and Non-Pecuniary Economies', in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, London: Macmillan
Grandmont, J.-M., 1983. *Money and Value*, Cambridge: Cambridge University Press
1985. 'On Endogenous Competitive Business Cycles', *Econometrica* 53, pp.995–1045
Grandmont, J.-M. and Laroque, G., 1976. 'On Keynesian Temporary Equilibria', *Review of Economic Studies* 43, pp.327–34
Gray, J., 1976. 'Wage Indexation: A Macroeconomic Approach', *Journal of Monetary Economics* 3, pp.221–35
1978. 'On Indexation and Contract Length', *Journal of Political Economy* 96,

pp.1–18

Green, R. and Weale, M.R., 1990. 'Macroeconomic Policy with Increasing Returns to Scale', Department of Applied Economics, University of Cambridge, mimeo

Greenwald, B. and Stiglitz, J., 1986. 'Externalities in Economies with Imperfect Information and Incomplete Markets', *Quarterly Journal of Economics* 100, pp.229–64

1987. 'Imperfect Information, Credit Markets and Unemployment', *European Economic Review* 31, pp.444–56

1988. 'Information, Finance Constraints and Business Fluctuations', in M. Kohn and S.C. Tsiang (eds.), *Finance Constraints, Expectations and Macroeconomics*, Oxford: Clarendon Press

1993. 'Financial Market Imperfections and Business Cycles', *Quarterly Journal of Economics* 108, pp.77–114

Greenwald, B., Stiglitz, J. and Weiss, A., 1984. 'Informational Imperfections in the Capital Markets and Macroeconomic Fluctuations', *American Economic Review* 74, pp.194–9

Greenwood, J. and Williamson, S.D., 1989. 'International Financial Intermediation and Aggregate Fluctuations Under Alternative Exchange Rate Regimes', *Journal of Monetary Economics* 23, pp.401–31

Grossman, G.M. and Helpman, E., 1991. *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press

Grossman, S.D. and Hart, O.D., 1981. 'Implicit Contracts, Moral Hazard and Unemployment', *American Economic Review* 71, pp.301–7

Grossman, S.D. and Stiglitz, J.E., 1976. 'Information and Competitive Price Systems', *American Economic Review* 66 (Papers and Proceedings), pp.246–53

1980. 'The Impossibility of Informationally Efficient Markets', *American Economic Review* 70, pp.393–409

Guesnerie, R. and Woodford, M., 1992. 'Endogenous Fluctuations' in J.-J. Laffont (ed.), *Advances in Economic Theory: Proceedings of the 6th World Congress of the Econometric Society*, vol. 2, Cambridge: Cambridge University Press

Gurley, J. and Shaw, E., 1955. 'Financial Aspects of Economic Development', *American Economic Review* 45, pp.515–38

Hahn, F.H., 1978. 'On Non-Walrasian Equilibria', *Review of Economic Studies* 45, pp.1–18

Hairault, J.-O. and Portier, F., 1993. 'Money, New Keynesian Macroeconomics and the Business Cycle', *European Economic Review* 37, pp.1533–68

Hall, R., 1986. 'Market Structure and Macroeconomic Fluctuations', *Brookings Papers on Economic Activity* 2, pp.285–322

1987. 'A Non-Competitive Equilibrium Model of Fluctuations', Department of Economics, Stanford University, mimeo.

1988. 'The Relation Between Price and Marginal Cost in U.S. Industry', *Journal of Political Economy* 96, pp.921–47

Haltiwanger, J. and Waldman, M., 1989. 'Limited Rationality and Strategic Complements: The Implications for Macroeconomics', *Quarterly Journal of*

Economics 104, pp.463–83
Hamilton, J.D., 1989. 'A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle', *Econometrica* 57, pp.357–84
Hansen, G., 1985. 'Indivisible Labour and the Business Cycle', *Journal of Monetary Economics* 16, 309–28
Hansen, G. and Wright, R., 1992. 'The Labor Market in Real Business Cycle Theory', *Federal Reserve Bank of Minneapolis Quarterly Review* Spring, pp.2–12
Harris, M. and Raviv, A., 1990. 'Capital Structure and the Informational Role of Debt', *Journal of Finance* 45, pp.321–49
Hart, O.D., 1980. 'A Model of Imperfect Competition with Keynesian Features', *Economic Theory Discussion Paper* 29, Department of Applied Economics, University of Cambridge, February
1982. 'A Model of Imperfect Competition with Keynesian Features', *Quarterly Journal of Economics* 97, pp.109–38
1983. 'Optimal Labour Contracts Under Asymmetric Information: An Introduction', *Review of Economic Studies* 50, pp.3–35
1985a. 'Monopolistic Competition in the Spirit of Chamberlin', *Review of Economic Studies* 52, pp.529–46
1985b. 'Imperfect Competition in General Equilibrium: An Overview of Recent Work', in K.J. Arrow and S. Honkapohja (eds.), *Frontiers of Economics*, Oxford: Blackwell
Hart, O.D. and Moore, J., 1989. 'Default and Renegotiation: A Dynamic Model of Debt', Cambridge, MA: MIT, mimeo
Heller, W., 1986. 'Coordination Failure Under Complete Markets with Applications to Effective Demand', in W. Heller, R. Starr and D. Starrett (eds.), *Equilibrium Analysis: Essays in Honour of Kenneth J. Arrow, Volume II*, Cambridge: Cambridge University Press
Heller, W. and Starrett, D., 1976. 'On the Nature of Externalities', in S. Lin (ed.), *Theory and Measurement of Economic Externalities*, New York: Academic Press
Hillier, B., 1991. *The Macroeconomic Debate*, Oxford: Blackwell
Hillier, B. and Ibrahimo, M.V., 1993. 'Asymmetric Information and Models of Credit Rationing', *Bulletin of Economic Research* 45, pp.271–304
Hillier, B., Lambert, P. and Turner, R., 1982. 'Macroeconomics with Non-Perfect Competition: A Comment', *Economic Journal* 92, pp.701–5
Hillier, B. and Worrall, T., 1994. 'The Welfare Implications of Costly Monitoring in the Credit Market', *Economic Journal* 104, pp.350–62
Horstein, A., 1993. 'Monopolistic Competition, Increasing Returns to Scale and the Importance of Productivity Shocks', *Journal of Monetary Economics* 31, pp.299–316
Hosios, A., 1990. 'On the Efficiency of Matching and Related Models of Search and Unemployment', *Review of Economic Studies* 57, pp.279–98
Howitt, P., 1985. 'Transactions Costs in the Theory of Unemployment', *American Economic Review* 75, pp.88–101

Howitt, P. and McAfee, P., 1988. 'Stability of Equilibria with Trade Externalities', *Quarterly Journal of Economics* 103, pp.261–78
Hurwicz, L., 1979. 'Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points', *Review of Economic Studies* 46, pp.217–25
Jacobsen, H.-J. and Schultz, C., 1990. 'A General Equilibrium Macromodel with Wage Bargaining', *Scandinavian Journal of Economics* 92, pp.379–98
1991. 'Undominated Wage Rates in a Unionised Overlapping-Generations Economy', *European Economic Review* 35, pp.1255–75
1994. 'On the Effectiveness of Economic Policy When Competition is Imperfect and Expectations are Rational', *European Economic Review* 38, pp.305–27
Jaffee, D. and Russell, T., 1976. 'Imperfect Information, Uncertainty and Credit Rationing', *Quarterly Journal of Economics* 90, pp.651–66
John, A., 1988. *Macroeconomic Externalities. Coordination Failures and the Microfoundations of Keynesian Economics*, Ph.D. dissertation, Yale University
Jones, S.R.G. and Stock, J.H., 1987. 'Demand Disturbances and Aggregate Fluctuations: The Implications of Near-Rationality', *Economic Journal* 97, pp.49–64
Kahn, C., 1979. 'An Extension of the Axiomatic Characterisation of Non-Cooperative Equilibria in Economies with a Continuum of Traders', Harvard University, mimeo
Kahn, C. and Mookherjee, D., 1988. 'A General Equilibrium Contracting Model with Keynesian Features', *Quarterly Journal of Economics* 103, pp.609–46
Katz, L., 1986. 'Efficiency Wages: A Partial Evaluation', *NBER Macroeconomics Annual* 1, pp.235–76
Keeton, W.R., 1979. *Equilibrium Credit Rationing*, New York and London: Garland Publishing
Keynes, J.M., 1936. *The General Theory of Employment, Interest and Money*, London: Macmillan
Kiyotaki, N., 1985. *Macroeconomics of Imperfect Competition*, Ph.D. dissertation, Harvard University
1988. 'Multiple Expectational Equilibria under Monopolistic Competition', *Quarterly Journal of Economics* 103, pp.695–713
Kiyotaki, N. and Moore, J., 1993. 'Credit Cycles', London School of Economics, mimeo
Koford, K., Miller, J. and Colander, D., 1988. 'Macroeconomic Externalities: New Keynesian Theory and Policy', University of Delaware, mimeo
Kydland, F. and Prescott, E., 1982. 'Time to Build and Aggregate Fluctuations', *Econometrica* 50, pp.1345–70
Laffont, J.-J., 1987. 'Externalities', in J. Eatwell, M. Milgate and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, London: Macmillan
Lambert, J.-P., 1988. *Disequilibrium Macroeconomic Models: Theory and Estimation of Rationing Models Using Business Survey Data*, Cambridge: Cambridge University Press
Lambin, J.-J., 1993. *Strategic Marketing*, New York: McGraw-Hill
Layard, R. and Nickell, S., 1985. 'The Causes of British Unemployment', *National*

Institute Economic Review 111, pp.62–85

1986. 'Unemployment in the UK', *Economica* 53 (Supp.), pp.121–66

Layard, R., Nickell, S. and Jackman, R., 1991. *Unemployment: Macroeconomic Performance and the Labour Market*, Oxford: Oxford University Press

Lee, D.R., 1975. 'Efficiency of Pollution Taxes and Market Structure', *Journal of Environmental Economics and Management* 2, pp.69–72

Leibenstein, H., 1957. 'The Theory of Underemployment in Densely-Populated Backward Areas', in H. Leibenstein, *Economic Backwardness and Economic Growth*, New York: Wiley

Leijonhufvud, A., 1968. *On Keynesian Economics and the Economics of Keynes*, New York: Oxford University Press

Lerner, A., 1960. 'On Generalising the General Theory', *American Economic Review* 50, pp.121–43

Licandro, O., 1992. 'A Non-Walrasian General Equilibrium Model with Monopolistic Competition and Wage Bargaining', *Working Paper*, Louvain-la-Neuve: Catholic University of Louvain

Lindbeck, A. and Snower, D.J., 1989. *The Insider–Outsider Theory of Employment and Unemployment*, Cambridge, MA: MIT Press

Lipman, B., 1991. 'How to Decide How to Decide How to . . .: Modeling Limited Rationality', *Econometrica* 59, pp.1105–25

Lucas, R.E., 1972. 'Expectations and the Neutrality of Money', *Journal of Economic Theory* 4, pp.103–24

Lucas, R.E. and Rapping, L., 1969. 'Real Wages, Employment and Inflation', *Journal of Political Economy* 77, pp.721–54

MacLeod, B. and Malcomson, J.M., 1993. 'Employment and Contract Inefficiency', paper presented at conference on 'Labour Market Institutions and Contracts', Florence (May)

Madden, P. and Silvestre, J., 1991. 'Imperfect Competition and Fixprice Equilibria when Goods are Gross Substitutes', *Scandinavian Journal of Economics* 93, pp.479–94

1992. 'Imperfect Competition and Fixprice Equilibria under Consumer Aggregation and Net Substitutes', *Scandinavian Journal of Economics* 94, pp.103–11

Malinvaud, E., 1977. *The Theory of Unemployment Reconsidered*, Oxford: Blackwell

Mankiw, N.G., 1985. 'Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly', *Quarterly Journal of Economics* 100, pp.529–37

1988. 'Imperfect Competition and the Keynesian Cross', *Economics Letters* 26, pp.7–13

1989. 'Real Business Cycles: A New Keynesian Perspective', *Journal of Economic Perspectives* 3, pp.79–90

Mankiw, N.G. and Romer, D., 1991. 'Introduction' in N.G. Mankiw and D. Romer (eds.), *The New Keynesian Economics*, vol. 1, Cambridge, MA: MIT Press

Manning, A., 1987. 'An Integration of Trade Union Bargaining Models in a Sequential Bargaining Framework', *Economic Journal* 97, pp.121–39

1990. 'Imperfect Competition, Multiple Equilibria and Unemployment Policy',

Economic Journal 100, pp.151–62
1992. 'Multiple Equilibria in the British Labour Market: Some Empirical Evidence', *European Economic Review* 36, pp.1333–65
Marcet, A. and Sargent, T.J., 1989. 'Least Squares Learning and the Dynamics of Hyperinflation', in W.A. Barnett, J. Geweke and K. Shell (eds.), *Economic Complexity: Chaos, Sunspots, Bubbles and Nonlinearity*, Cambridge: Cambridge University Press
Marimon, R. and Sunder, S., 1988. 'Rational Expectations vs. Adaptive Behaviour in a Hyperinflationary World: Experimental Evidence', *Discussion Paper* 247, Center for Economic Research, University of Minnesota
1991. 'Some Puzzles From Experimental Monetary Economics', Carnegie-Mellon University and IAE Universitat Autonoma de Barcelona, mimeo
1993. 'Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence', *Econometrica* 61, pp.1073–1107
Marris, R., 1991. *Reconstructing Keynesian Economics with Imperfect Competition: A Desk-Top Simulation*, London: Edward Elgar
Marshak, T. and Selten, R., 1974. *General Equilibrium with Price-Setting Firms*, Berlin: Springer-Verlag
Matsuyama, K., 1992. 'Modelling Complementarity in Monopolistic Competition', Northwestern University, mimeo
McDonald, I. and Solow, R., 1981. 'Wage Bargaining and Employment', *American Economic Review* 71, pp.896–908
Meade, J.E., 1952. 'External Economies and Diseconomies in a Competitive Situation', *Economic Journal* 62, pp.54–67
1973. *The Theory of Economic Externalities*, Geneva: A.W. Sijthoff-Leiden
Minsky, H.P., 1976. *John Maynard Keynes*, London: Macmillan
Mishan, E., 1971. 'The Post-War Literature on Externalities', *Journal of Economic Literature* 9, pp.1–28
Mishkin, F., 1978. 'The Household Balance Sheet and the Great Depression', *Journal of Economic History* 38, pp.918–37
Modigliani, F. and Miller, M.H., 1958. 'The Cost of Capital, Corporation Finance and the Theory of Investment', *American Economic Review* 48, pp.261–97
Molana, H. and Moutos, T., 1992. 'A Note on Taxation, Imperfect Competition and the Balanced-Budget Multiplier', *Oxford Economic Papers* 44, pp.68–74
Mortensen, D., 1982. 'The Matching Process as a Non-Cooperative Bargaining Game', in J.J. McCall (ed.), *The Economics of Information and Uncertainty*, Chicago: University of Chicago Press
1986. 'Job Search and Labor Market Analysis', in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, vol. 2, Amsterdam: North-Holland
Moutos, T., 1991. 'Turnover Costs, Unemployment and Macroeconomic Policies', *European Journal of Political Economy* 7, pp.1–16
Murphy, K.M., Shleifer, A. and Vishny, R.W., 1989. 'Industrialisation and the Big Push', *Journal of Political Economy* 97, pp.1003–26
Myers, S.C. and Majluf, N.S., 1984. 'Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have',

Journal of Financial Economics 13, pp.187–221

Neftci, S.N., 1984. 'Are Economic Time Series Asymmetric Over the Business Cycle?', *Journal of Political Economy* 92, pp.307–28

Negishi, T., 1961. 'Monopolistic Competition and General Equilibrium', *Review of Economic Studies* 28, pp.196–201

1978. 'Existence of an Underemployment Equilibrium', in G. Schwodiauer (ed.), *Equilibrium and Disequilibrium in Economic Theory*, Dordrecht: Reidel

1979. *Microeconomic Foundations of Keynesian Macroeconomics*, New York: North-Holland

Ng, Y.-K., 1980. 'Macroeconomics with Non-Perfect Competition', *Economic Journal* 90, pp.598–610

1982a. 'A Micro–Macroeconomic Analysis Based on a Representative Firm', *Economica* 49, pp.121–39

1982b. 'Macroeconomics with Non-Perfect Competition: A Reply to Hillier, Lambert and Turner', *Economic Journal* 92, pp.706–7

1986. *Mesoeconomics: A Micro–Macro-Analysis*, Brighton: Wheatsheaf

Nickell, S., 1986. 'Dynamic Models of Labour Demand', in O. Ashenfelter and R. Layard (eds.), *Handbook of Labor Economics*, vol. 1, Amsterdam: North-Holland

Obstfeld, M. and Rogoff, K., 1983. 'Speculative Hyperinflations in Maximising Models: Can We Rule Them Out?', *Journal of Political Economy* 91, pp.675–87

Oi, W., 1962. 'Labor as a Quasi-Fixed Factor', *Journal of Political Economy* 80, pp.538–55

Orphanides, A., 1993. 'Labour Hoarding When Unemployment is a Worker Discipline Device', *Scandinavian Journal of Economics* 95, pp.111–18

Oswald, A.J., 1985. 'The Economic Theory of Trade Unions: An Introductory Survey', *Scandinavian Journal of Economics* 87, pp.160–93

Pagano, M., 1990. 'Imperfect Competition, Underemployment Equilibria and Fiscal Policy', *Economic Journal* 100, pp.440–63

Parkin, M., 1986. 'The Output–Inflation Trade-Off When Prices Are Costly to Change', *Journal of Political Economy* 94, 200–24

Patinkin, D., 1965. *Money, Interest and Prices*, New York: Harper & Row

Pencavel, J.H., 1984. 'The Trade-Off Between Wages and Employment in Union Objectives', *Quarterly Journal of Economics* 99, pp.215–32

Phelps, E. *et al.* (eds.), 1970. *The Microeconomic Foundations of Macroeconomics*, New York: Norton

Picard, P., 1993, *Wages and Unemployment: A Study in Non-Walrasian Macroeconomics*, Cambridge: Cambridge University Press

Pissarides, C., 1984. 'Search Intensity, Job Advertising and Efficiency', *Journal of Labor Economics* 2, pp.128–43

1986. 'Unemployment and Vacancies in Britain', *Economic Policy* 3, pp.473–93

Plosser, C., 1989. 'Understanding Real Business Cycles', *Journal of Economic Perspectives* 3, pp.51–77

Portier, F., 1994. 'Business Formation and Cyclical Markups in the French Business Cycle', Paris: CEPREMAP and MAD, mimeo

Prescott, E., 1986. 'Theory Ahead of Business Cycle Measurement', *Federal Reserve Bank of Minneapolis Quarterly Review*, Fall, pp.9–22

Quandt, R., 1988. *The Econometrics of Disequilibrium*, New York: Blackwell

Radner, R., 1981. 'Equilibrium Under Uncertainty', in K.J. Arrow and M.D. Intriligator (eds.), *Handbook of Mathematical Economics*, Amsterdam: North-Holland

Rankin, N., 1992. 'Imperfect Competition, Expectations and the Multiple Effects of Monetary Growth', *Economic Journal* 102, pp.743–53

1993. 'Monetary and Fiscal Policy in a "Hartian" Model of Imperfect Competition', in M. Bacharach, M. Dempster and J. Enos (eds.), *Mathematical Models in Economics*, Oxford: Oxford University Press, forthcoming

1995. 'Money in Hart's Model of Imperfect Competition', *European Journal of Political Economy* 11

Rivera-Campos, R., 1991. 'Short-Run and Long-Run Effects of Devaluation in a Macromodel of Imperfect Competition', *Discussion Paper* 388, Department of Economics, University of Warwick

Roberts, J., 1987. 'General Equilibrium Analysis of Imperfect Competition: An Illustrative Example', in G. Feiwel (ed.), *Arrow and the Ascent of Economic Man*, London: Macmillan

1988. 'Involuntary Unemployment and Imperfect Competition: A Game-Theoretic Macro Model', in G. Feiwel (ed.), *The Economics of Imperfect Competition and Employment: Joan Robinson and Beyond*, London: Macmillan

Romer, P., 1986. 'Increasing Returns and Long-Run Growth', *Journal of Political Economy* 94, pp.1002–73

Rosa, R.V., 1951. 'Interest Rates and the Central Bank', in R.V. Rosa, *Money, Trade and Economic Growth: Essays in Honour of John H. Williams*, New York: Macmillan

Rosen, S., 1986. 'Implicit Contracts: A Survey', *Journal of Economic Literature* 23, pp.1144–75

Rotemberg, J.J., 1987. 'The New Keynesian Microeconomic Foundations', *NBER Macroeconomics Annual* 2, pp.69–114

Rotemberg, J.J. and Saloner, G., 1986. 'A Supergame-Theoretic Model of Price Wars During Booms', *American Economic Review* 76, pp.390–407

Rotemberg, J.J. and Woodford, M., 1991. 'Markups and the Business Cycle', *NBER Macroeconomics Annual* 6, pp.63–129

1992. 'Oligopolistic Pricing and the Effects of Aggregate Demand on Economic Activity', *Journal of Political Economy* 100, pp.1153–1207

1993. 'Dynamic General Equilibrium Models with Imperfectly Competitive Product Markets', in T.F. Cooley (ed.), *Frontiers in Business Cycle Research*, Princeton: Princeton University Press, forthcoming

Saint-Paul, G., 1990, 'Efficiency Wages, Dualism and the Dynamics of Labor Demand', DELTA *Working Paper* 90–29

Salop, S.C., 1979. 'Monopolistic Competition with Outside Goods', *Bell Journal of Economics* 10, pp.141–56

Samuelson, P.A., 1958. 'An Exact Consumption-Loan Model of the Rate of

Interest With or Without the Social Contrivance of Money', *Journal of Political Economy* 66, pp.467–82

Sandmo, A., 1975. 'Optimal Taxation in the Presence of Externalities', *Swedish Journal of Economics* 77, pp.86–98

Sargent, T.J., 1987. *Dynamic Macroeconomic Theory*, Cambridge, MA: Harvard University Press

1993. *Bounded Rationality in Macroeconomics*, Oxford: Clarendon Press

Sargent, T.J. and Wallace, N., 1987. 'Inflation and the Government Budget Constraint', in A. Razin and E. Sadka (eds.), *Economic Policy in Theory and Practice*, London: Macmillan

Scheinkman, J.A., 1978. 'Notes on Asset Trading in an Overlapping-Generations Model', University of Chicago, mimeo

Scheinkman, J.A. and Weiss, L., 1986. 'Borrowing Constraints and Aggregate Economic Activity', *Econometrica* 54, pp.23–45

Schultz, C., 1992. 'The Impossibility of Involuntary Unemployment in an Overlapping Generations Model with Rational Expectations', *Journal of Economic Theory* 58, pp.61–76

1993. 'Wages and Employment in Repeated Games With Revenue Fluctuations', University of Copenhagen, mimeo

Shapiro, C. and Stiglitz, J., 1984. 'Equilibrium Unemployment as a Worker Discipline Device', *American Economic Review* 74, pp.433–44

Shapiro, M., 1987. 'Measuring Market Power in U.S. Industry', Cowles Foundation, Yale University, mimeo

Sheshinski, E. and Weiss, Y., 1977. 'Inflation and the Costs of Price Adjustment', *Review of Economic Studies* 44, pp.287–303

1983. 'Optimum Pricing Under Stochastic Inflation', *Review of Economic Studies* 50, pp.513–29

Shleifer, A., 1986. 'Implementation Cycles', *Journal of Political Economy* 94, pp.1163–90

Shleifer, A. and Vishny, R.W., 1980. 'The Role of the Stock Market in Coordinating Macroeconomic Activity', Department of Economics, University of Chicago, mimeo

1988. 'The Efficiency of Investment in the Presence of Aggregate Demand Spillovers', *Journal of Political Economy* 96, pp.1221–31

Shubik, M., 1984. *A Game-Theoretic Approach to Political Economy*, Cambridge, MA: MIT Press

Silvestre, J., 1990. 'There May be Unemployment When the Labour Market is Competitive and the Output Market is Not', *Economic Journal* 100, pp.899–913

1991. 'Inefficiency Without Externality', University of California at Davis, mimeo

1993. 'The Market-Power Foundations of Macroeconomic Policy', *Journal of Economic Literature* 31, pp.105–41

1994. 'Market Power in Macroeconomic Models: New Developments', *Working Paper* 94–07, University of California at Davis; forthcoming in *Annales* *d'Economie et de Statistique*

Smith, V.L. and Walker, J.M., 1993a. 'Rewards, Experience and Decision Costs in First Price Auctions', *Economic Inquiry* 31, pp.237–44

1993b. 'Monetary Rewards and Decision Costs in Experimental Economies', *Economic Inquiry* 31, pp.245–61

Sneessens, H., 1987. 'Investment and the Inflation–Unemployment Trade-Off in a Macroeconomic Rationing Model with Monopolistic Competition', *European Economic Review* 31, pp.781–815

Snower, D.J., 1983. 'Imperfect Competition, Underemployment and Crowding Out', *Oxford Economic Papers* 35 (Supp.), pp.245–70

Solow, R., 1979. 'Another Possible Source of Wage Stickiness', *Journal of Macroeconomics* 1, pp.79–82

1986. 'Monopolistic Competition and the Multiplier', in W. Heller, R. Starr and D. Starrett (eds.), *Equilibrium Analysis: Essays in Honour of K.J. Arrow*, vol. 2, Cambridge: Cambridge University Press

Startz, R., 1984. 'Prelude to Macroeconomics', *American Economic Review* 74, pp.881–92

1989. 'Monopolistic Competition as a Foundation for Keynesian Macroeconomic Models', *Quarterly Journal of Economics* 104, pp.737–52

1990. 'Dynamic Aggregate Demand and New Keynesian Multipliers', *Working Paper*, University of Washington

Stiglitz, J.E., 1976. 'The Efficiency Wage Hypothesis, Surplus Labor and the Distribution of Income in LDC's', *Oxford Economic Papers* 28, pp.185–207

1988. 'Money, Credit and Business Fluctuations', *Economic Record* 64, pp.307–22

1992. 'Capital Markets and Economic Fluctuations in Capitalist Economies', *European Economic Review* 36, pp.269–306

Stiglitz, J.E. and Weiss, A., 1981. 'Credit Rationing in Markets With Imperfect Information', *American Economic Review* 71, pp.393–410

Strand, J., 1991. 'Unemployment and Wages Under Worker Moral Hazard With Firm-Specific Cycles', *International Economic Review* 32, pp.601–12

1992a. 'Business Cycles With Worker Moral Hazard', *European Economic Review* 36, pp.1291–1304

1992b. 'Environmental Policy Under Worker Moral Hazard', Memorandum 24/92, Department of Economics, University of Oslo

1992c. 'Environmental Policy Under Labour Unions With Endogenous Environmental Effort', Memorandum 25/92, Department of Economics, University of Oslo

1993a. 'Optimal Environmental Taxes in a Distorted Economy With Averting and Avoiding Behavour', Department of Economics, University of Oslo, mimeo

1993b. 'Moral Hazard With Firm-Specific Uncertainty', Department of Economics, Stanford University

Svensson, L., 1986. 'Sticky Goods Prices, Flexible Asset Prices, Monopolistic Competition and Monetary Policy', *Review of Economic Studies* 53, pp.385–405

Taylor, J., 1979. 'Staggered Wage Setting in a Macro Model', *American Economic Review* 69, pp.108–13

Thomas, J.P., 1990. 'The Ineffectiveness of Fiscal and Monetary Policy under Imperfect Competition', Department of Economics, Warwick University, mimeo

Tobin, J., 1972. 'Inflation and Unemployment', *American Economic Review* 62, pp.1–18

Townsend, R.M., 1979. 'Optimal Contracts and Competitive Markets With Costly State Verification', *Journal of Economic Theory* 21, pp.265–93

1980. 'Models of Money with Spatially Separated Agents', in J. Kareken and N. Wallace (eds.), *Models of Monetary Economies*, Minneapolis: Federal Reserve Bank of Minneapolis

Turvey, R., 1963. 'On Divergences Between Social Cost and Private Cost', *Economica* 30, pp.309–13

Van Huyck, J., 1989. 'Decentralised Competition, Pecuniary Externalities and Allocative Efficiency', Texas A&M University, mimeo

Weiss, A., 1980. 'Job Queues and Layoffs in Labor Markets with Flexible Wages', *Journal of Political Economy* 88, pp.526–38

Weitzman, M.L., 1982. 'Increasing Returns and the Foundations Unemployment Theory', *Economic Journal* 92, pp.787–804

1983. 'Some Macroeconomic Implications of Alternative Compensation Schemes', *Economic Journal* 93, pp.763–83

1985. 'The Simple Macroeconomics of Profit-Sharing', *American Economic Review* 75, pp.937–53

Williamson, S.D., 1986. 'Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing', *Journal of Monetary Economics* 18, pp.159–79

1987. 'Financial Intermediation, Business Failures and Real Business Cycles', *Journal of Political Economy* 95, pp.1196–1216

Yellen, J., 1984. 'Efficiency Wage Models of Unemployment', *American Economic Review* 74 (Papers and Proceedings), pp.200–5

Index of authors

Index of subjects