Essays on the Economics of Price Transmission

Timothy Walter Helm

Submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy (with coursework component)

July 2012
Department of Economics
The University of Melbourne
Abstract

This thesis explores three topics relating to price transmission in economic theory. The broad aim is to better understand the causes and consequences of price adjustment processes, within a selection of case studies where price transmission plays a central role.

Part I examines asymmetric price transmission, whereby rising prices in upstream markets are more rapidly passed through to downstream markets than are falling upstream prices. A model of oligopolistic firms adjusting to a cost shock is used to predict the patterns of price transmission induced by asymmetry in the costs of adjusting output. The model allows for quantification of results through selection of realistic parameter values, and for this purpose an original form of adjustment cost function is introduced. After arguing that existing methods of measuring asymmetry are inappropriate for cross-market comparisons, several new measures are suggested and then used to analyse the outcomes of the model.

The subject of Part II is the relationship between market power and the speed of adjustment, and the aim is to investigate how theorists can reliably use dynamic models to understand this relationship. It is argued that parameterising market power in such models via the ‘conjectural variations’ technique can be conceptually problematic, and that several attempts to do so have been methodologically unsound. Part II demonstrates a new method of exploring the market power–adjustment speed relationship, based on an alternative interpretation of conjectural variations. It is shown that even within a single model the direction of the predicted relationship may not be unambiguous.

Part III is a case study of a policy argument where the non-transmission
of a supply cost shock plays an important role. The topic is land-value taxation, and the aim is to examine the robustness of the ‘zero price-transmission’ argument as well as to explore several of its implications. One such implication, it is argued, is that the tax effectively acts to pool the risk associated with land-value fluctuations, thereby improving the allocation of a risk that serves little to no social purpose when borne privately. Part III also explores land-value taxation in historical and political context, contrasting the influence of efficiency and equity arguments in the concerns of classical economists, in the motivation for the adoption of land taxes, and in the reasons for subsequent abandonment.
Declaration

This is to certify that

i) the thesis comprises only my original work towards the PhD except where indicated in the preface,

ii) due acknowledgement has been made in the text to all other material used,

iii) the thesis is fewer than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Timothy Walter Helm
Acknowledgements

Above all I would like to thank my wife Mel for her patience and support over the lifetime of my candidature.

Sincere thanks also go to my supervisor, Professor John Creedy, who has provided much encouragement and guidance.

The thesis has benefitted from the comments and assistance of Sisira Jayasuriya, John Freebairn, Russell Hillberry and Jeff Borland, as well as from discussions with other graduate students in the Department of Economics.

Assistance from Imogen Halstead, Philip de la Rue, Ermias Weldemicael, Solmaz Moslehi, Mike Chongussayakul, Samir Al-Busaidi and Luke Wainscoat has been particularly helpful.

Finally I would like to thank my parents for having encouraged me to undertake and complete my doctoral studies, and my parents-in-law for their support throughout the process.
## Contents

<table>
<thead>
<tr>
<th>Contents</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>xv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xvii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
</tbody>
</table>

### I Asymmetric adjustment costs and asymmetry in price transmission 9

#### 2 A dynamic oligopoly model with adjustment costs and conjectures 11

<table>
<thead>
<tr>
<th>2.1 Introduction</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1.1 Background</td>
<td>11</td>
</tr>
<tr>
<td>2.1.2 Outline of Part I</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Modelling method</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Dynamic model setup</td>
<td>19</td>
</tr>
<tr>
<td>2.3.1 The firm’s problem: a differential game</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2 Concepts in differential games</td>
<td>20</td>
</tr>
<tr>
<td>2.3.3 Information, beliefs and strategy space assumptions</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Dynamic model solution</td>
<td>24</td>
</tr>
<tr>
<td>2.4.1 The firm’s problem: an optimal control problem</td>
<td>25</td>
</tr>
<tr>
<td>2.4.2 Optimal quantity strategies</td>
<td>26</td>
</tr>
<tr>
<td>2.4.3 Steady state condition and interpretation of θ</td>
<td>27</td>
</tr>
</tbody>
</table>
## 2.4.4 Quantity adjustment path

2.5 Static input choice model

2.5.1 Input cost minimisation

2.5.2 Equilibrium displacement model

2.5.3 Price adjustment path

2.6 Conclusion

### 3 Adjustment cost functions

3.1 Introduction

3.2 Quadratic form

3.2.1 Properties

3.2.2 Interpretation

3.2.3 Adjustment speed parameter

3.3 Logarithmic form

3.3.1 Properties

3.3.2 Interpretation

3.3.3 Derivation of $A_{ln}$

3.3.4 Adjustment speed parameter

3.4 Conclusion

### 4 Analysis: adjustment speeds and asymmetric price transmission

4.1 Introduction

4.2 The speed and magnitude of transmission

4.3 Asymmetric price transmission

4.3.1 Introduction

4.3.2 Asymmetry measures

4.3.3 Model predictions

4.4 Conclusion
CONTENTS

II Market power and the speed of price adjustment 79

5 Conjectural variations, market power and the speed of price
adjustment 81
  5.1 Introduction ....................................................... 81
  5.2 Conjectural variations ........................................... 87
    5.2.1 Introduction ................................................ 87
    5.2.2 Methodological issues ..................................... 91
    5.2.3 Conjectures as ‘shorthand’ ............................... 92
    5.2.4 Dynamic conjectural variations ......................... 93
    5.2.5 Worthington (1989) and Martin (1993) .................. 93
    5.2.6 Critique of Worthington (1989) and Martin (1993) ap-
        proach ......................................................... 96
  5.3 Dockner (1992) model .......................................... 98
    5.3.1 Introduction ................................................. 98
    5.3.2 General model .............................................. 99
    5.3.3 Linear-quadratic case .................................... 101
  5.4 Market power and adjustment speeds ......................... 104
    5.4.1 Adjustment path and adjustment speed measure .......... 104
    5.4.2 Variation of $\alpha$ and $\lambda$ with parameters .......... 105
    5.4.3 Discussion .................................................. 109
  5.5 Conclusion ..................................................... 111

III Taxes on land values: incidence and other top-
ics 115

6 Land-value taxation: Efficiency, development timing, and
risk pooling 117
  6.1 Introduction ...................................................... 117
    6.1.1 Background .................................................. 117
    6.1.2 Outline of Chapter 6 ...................................... 120
  6.2 Incidence and neutrality ..................................... 124
    6.2.1 A tax on the return to land ............................... 125
CONTENTS

6.2.2 A tax on the value of land .......................... 131
6.2.3 Exemptions and tax pass-through .................. 137

6.3 The timing of development ............................... 138
6.3.1 Development timing under certainty ................. 138
6.3.2 Development timing with uncertainty ............... 145

6.4 Asset price effects ...................................... 150
6.4.1 Pre-tax asset values and income ................... 151
6.4.2 Post-tax asset values, incomes and tax receipts .... 152
6.4.3 The government’s ‘equity share’ .................... 153

6.5 Risk pooling ........................................... 155
6.5.1 The source and nature of the risk .................. 156
6.5.2 Consequences of land value risk .................... 160
6.5.3 Pooling risk through land tax ...................... 162
6.5.4 Consequences of risk-pooling ....................... 163
6.5.5 Distributional consequences ......................... 165
6.5.6 Summary ........................................... 168

6.6 Conclusion ............................................. 170

7 Land taxation in historical context ...................... 175

7.1 Introduction ............................................ 175
7.2 A history of economic thought ......................... 176
7.2.1 The Physiocrats and Smith ......................... 176
7.2.2 Ricardo and the Mills .............................. 179
7.2.3 From Henry George onwards ....................... 181
7.2.4 Summary ........................................... 182
7.3 A history of implementation ............................ 183
7.3.1 The contemporary role of taxes on land .......... 183
7.3.2 New Zealand and Australia ......................... 187
7.3.3 The USA and the UK .............................. 195
7.3.4 Summary ........................................... 198

7.4 Conclusion ............................................. 200

8 Conclusion ................................................ 201
CONTENTS

A Solving the dynamic and static games of Chapter 2 209
B Parameter effects from Chapter 4 217
C Solving the dynamic game of Chapter 5 219

Bibliography 223
List of Tables

4.1 Direction of influence of parameters on speed and magnitude of price transmission ................................. 53

4.2 Quantitative influence of parameters on speed of price transmission ....................................................... 57

5.1 Relationships between parameter values in Dockner (1992) model, various discount rates .......................... 106

5.2 Relationships between parameter values in Dockner (1992) model, various adjustment cost values .............. 106

5.3 Relationships between parameter values in Dockner (1992) model, various quadratic cost parameters ........... 107
List of Figures

4.1 Value of adjustment speed parameter for various levels of market power ........................................... 54
4.2 Value of adjustment speed parameter for various demand elasticities .................................................. 56
4.3 Downstream price responses and asymmetry in price responses ......................................................... 59
4.4 Asymmetry in price transmission for various adjustment cost levels ................................................... 70
4.5 Asymmetry in price transmission for various demand elasticities ....................................................... 71
4.6 Asymmetry in price transmission for various market power levels ..................................................... 72
4.7 Asymmetry in price transmission for various discount rates ................................................................. 74

5.1 Market power-adjustment speed relationship in Dockner (1992) model, varying parameters $r$, $k$, and $b$ ...................................................... 108
Chapter 1
Introduction

Much of economic theory consists of models which examine equilibria, and which study economic relationships through comparative statics. Mechanisms of adjustment and disequilibrium situations have traditionally received relatively less attention, although given the complexities of dynamic analysis this is understandable and appropriate. Yet this focus has also potentially obscured some interesting and important facts about how markets adjust to shocks in reality; that is, facts about the nature of ‘price transmission’. There remain important and unanswered questions in the economics of price transmission, particularly regarding the determinants of the magnitude of transmission, speeds of adjustment, and asymmetry in adjustment processes.

These questions matter because price transmission assumptions are a feature of many arguments in macroeconomics, international trade, public finance and other areas. Theories in these disciplines and policy arguments based on them frequently depend on simple assumptions (often left unstated) regarding the pass-through of prices from one market to another.¹ Most common is the implicit assumption that price changes will be immediately and

¹For example, assumptions about exchange rate pass-through (where changes in the domestic-currency level of internationally-set prices feed through to domestic prices) will clearly affect expectations and views on inflation, trade flows, and the profitability of domestic industries. Basic trade theory itself and the normative conclusions typically drawn from it also effectively rely on an assumption about exchange rate pass-through (i.e. that liberalisation results in domestic prices adjusting to international levels). Another example is in tax policy: the extent to which a tax is passed on determines the incidence and ultimately the welfare effects, behavioural changes, and efficiency of the tax.
fully reflected in prices in related markets.

Understanding the microeconomic determinants of price adjustment is therefore important, in order to both refine and appropriately qualify those theories and policy arguments where transmission of price signals plays a central role. In other words, given that policymakers care about the speed and magnitude of transmission of price shocks in many real-life situations, economists ought to understand these processes as much as possible. In those situations of most interest it would be valuable to see how the standard theoretical approximation of instantaneous transmission should be modified, based on a fuller understanding of transmission processes.

There is also a second reason for subjecting price adjustment assumptions in economic arguments to closer analysis. This is that the sensitivity of price transmission to small changes in the underlying characteristics of markets is an important policy consideration, in a world of imperfect information about both the reliability of economic science and the true nature of policy-affected markets. If underlying characteristics of markets – such as their structure, the conduct of players, and the informational environment – greatly affect price adjustment processes, then policy interventions where price transmission is important become riskier. The stock we place in economic theories and related policy arguments which rely on price transmission ought to be a function of their robustness, or universality. Exploration of how price linkages depend on the nature of the markets in question can help establish this.

The essays in this thesis are motivated in part by the broad goal of better understanding price adjustment processes. Each is centred on a particularly interesting or unusual observation or prediction about patterns of price transmission. In different ways, and on quite different topics, the three parts of the thesis address the same basic questions of ‘why do prices adjust in this way?’

---

2 An illustrative example may be the choice of the ‘point of obligation’ in environmental taxation. If a tax is imposed where compliance and collection costs are lowest but the cheapest abatement opportunities are upstream or downstream of this point, then the effectiveness of the tax in delivering efficient abatement relies on transmission of price signals. If assumptions about this are incorrect, due to misunderstanding the nature of the relevant markets and how this bears on price transmission, then the choice of legal incidence may be costlier or less effective than other options and compensation may be directed to the wrong groups.
and ‘what does this imply?’ That is, the material within is themed around explaining the causes and understanding the consequences of price adjustment processes across a selection of ‘case studies’ where price transmission plays a central role.

The thesis is not aimed at comprehensively surveying the literature on the mechanics of price adjustment or discussing all approaches to these questions. Rather, a focus on price transmission is the common element across the three parts of the thesis; the three case studies consider quite distinct applied questions, which all centre in some way on a price adjustment issue. Furthermore, the approach throughout is theoretical, not empirical. Where mathematical modelling is used (in Parts I and II), the choice of method is necessarily selective: particular techniques and assumptions are used to the exclusion of other potentially-insightful approaches.

The first part of the thesis explores the topic of asymmetric price transmission. A growing collection of empirical studies across a wide variety of markets has shown that the pass-through of price changes from one point to another in a vertical supply chain often varies according to the direction of the change. Most commonly, rising prices in upstream markets are transmitted more rapidly to downstream markets than falling upstream prices; that is, ‘prices rise faster than they fall’.

That the static theoretical models explaining price determination do not (and cannot) predict asymmetry in adjustment speeds is obvious. What this means for the continuing emphasis placed on these models and the directions of theory going forward is more debatable. One provocative claim in this debate is even that “the theory [of markets] is wrong, at least insofar as an asymmetric response to costs is not its general implication... in this case, lay prejudice comes close to the truth than our theory does” (Peltzman 2000, p493). Whatever the significance of this sort of claim, it is clear there are opportunities to usefully extend theory to better reflect empirical evidence and to deal with many practical issues relating to price adjustment.

However the significance of widespread asymmetry is not just in the challenge it poses to economic theory, reflected in the quote above, as a previously-overlooked and unexplained ‘stylised fact’ about markets. Asym-
metric price transmission may also have welfare implications, either as a signal of the market power of supply-chain intermediaries or perhaps as the primary opportunity for exercise of that power. Crucially, however, this depends on the cause of asymmetry. If slower pass-through of negative input price shocks reflects higher real costs of output expansion than of output contraction, rather than the exercise of market power, then the welfare effects will be solely redistributive without any net loss.

Part I explores the patterns of price adjustment generated by a general asymmetric adjustment cost of this type. The goal is to make a small contribution towards a set of ‘diagnostics’ that could help isolate the causes of asymmetry in any given market. Here the contribution is to describe what asymmetry induced by asymmetric adjustment costs might look like in practice. The approach involves asking how adjustment costs affect the speed of adjustment to shocks, what degree of pricing asymmetry is consistent with realistically-sized asymmetries in adjustment costs, and how other aspects of the relevant markets impact on these relationships.

These questions are examined in the setting of a dynamic model where firms respond to an input price shock whilst facing costs of output adjustment. The model is designed to allow variety in aspects of market structure and in the competitiveness of firm conduct, in order to observe the sensitivity of price-adjustment patterns to these factors. As well, it is deliberately designed in a manner that allows some quantification of results through the selection of realistic parameter values.

Chapter 2 develops this model, and Chapter 3 considers how to mathematically represent adjustment costs so that the cost parameter can be easily interpreted and sized. Since the simplest specification of convex adjustment costs (a quadratic cost function) does not, by itself, contain an economically-meaningful adjustment cost parameter, Chapter 3 suggests two functional forms with parameters that may be straightforwardly interpreted and potentially quantified.

Chapter 4 considers the question of how asymmetric price responses might best be measured, and argues that the standardly-used measure would be inappropriate when comparing the degree of asymmetry across industries.
Various alternative measures are suggested, and then used to examine the patterns of asymmetry arising from the model under conditions of asymmetry in adjustment costs.

The second part of the thesis also examines dynamic models of oligopoly, although from a methodological perspective and with somewhat different goals in mind. Part II is broadly aimed at better understanding how concentration and market power each affect the speed at which prices adjust to cost shocks. The more specific aims, however, are to investigate how theorists can reliably use dynamic models to approach this question, and whether there is any reason to expect general relationships between these variables.

The work here builds on a long-running line of enquiry into these relationships within industrial economics. Amongst the collection of (often-contradictory) theoretical and empirical results in this literature are several hypotheses that arise from simple dynamic oligopoly models, in which a ‘conjectural variation’ parameter is included as a market power measure.

Although the conjectural variations modelling approach appears attractive for various reasons, there are conceptual problems with the technique that make the correct use and interpretation of conjectures crucial. Chapter 5 discusses these concerns and asks whether a conjectural variation, as a useful parameterisation of market power, can in fact be correctly used in the context of a dynamic model in order to investigate the market power – adjustment speed relationship.

It is argued that the methods of employing conjectures in several existing dynamic models are, in fact, not methodologically sound. However an alternative interpretation of conjectures as static ‘shorthand’ for the competitiveness of a dynamic game is available. Based on this interpretation the chapter demonstrates a novel method of using dynamic models to gain insight into the market power – adjustment speed relationship. Moreover, this application shows that even within a single model the direction of the predicted relationship may not be unambiguous.

Part III is a case study of a policy argument where the non-transmission of a supply cost shock plays a crucial role. The topic is land-value taxation, where the efficiency of the tax is a direct consequence of the immobility of
land and the inability of landowners to pass the tax through to prices. As a revenue source which is ‘immune’ to the pressures globalisation places on mobile tax bases, land-value taxation is receiving renewed attention in policy circles, including through recent tax reviews in Australia, New Zealand and the UK.

The chapters in this part of the thesis have two broad goals. The first is to examine the premises of the neutrality argument and better understand when these hold in reality. Secondly, the chapters aim to explore some of the implications of this argument for policy. Chapter 6 first lays out the basic logic of the neutrality of the tax, and summarises a collection of lesser-known points regarding circumstances when the tax might be non-neutral. These circumstances include the presence of uncertainty about future demand for land, or the use of particular methods of land-value assessment, among other situations. In explaining these points the chapter suggests a logical framework for understanding the neutrality of the tax, and for classifying the various situations of non-neutrality.

The chapter then presents a new argument about the implications of the tax that, it is suggested, adds a useful perspective to our understanding and adds another relevant point to the policy case. A land-value tax is well-known to be generally neutral with respect to resource allocation. However Chapter 6 argues it may also act to effectively improve the allocation of a certain undesirable quantity: the risk associated with land-value fluctuations. The argument concludes that land-value taxation can be usefully considered as a policy that acts to pool an otherwise-unavoidable risk, which serves little to no social purpose when borne privately.

The final chapter steps back to look at the efficiency case for land taxation in historical and political context, both as a major topic of debate for the classical economists and as a policy with a long history of implementation (particular in Australia and New Zealand). The surveys presented here ask how classical economists viewed the desirability of land taxes, particularly with regard to the ethics of private ownership of land income and taxation of land wealth. Next, the question of how these ethical (as well as efficiency) arguments influenced the early adopters of land taxation is considered. Finally
the reasons for the gradual disappearance of taxes on land are examined, with a particular focus on how conceptions of fairness in taxation differ today from the views held throughout the 19th and early 20th century.

Ultimately these historical surveys highlight the limited influence over policy choices of efficiency arguments for the tax, in the face of distributional concerns. In the debates of the classical economists regarding land taxes, in the motivation for their introduction, and in the reasons for subsequent abandonment it has been views about fairness, not efficiency, that have dominated throughout the period considered.
Part I

Asymmetric adjustment costs and asymmetry in price transmission
Chapter 2

A dynamic oligopoly model with adjustment costs and conjectures

2.1 Introduction

2.1.1 Background

This chapter constructs a mathematical model designed to explore ‘asymmetric price transmission’ in vertical supply chains. Both the adjustment process and steady-state equilibria are modelled, in order to allow examination of the speed of adjustment as well as the magnitude of price pass-through.

Asymmetry in price transmission refers to a situation where the speed of adjustment of downstream prices to upstream shocks depends on the direction of the upstream change. Interest in this situation arises largely for scientific reasons, as a poorly-understood phenomenon which an expanding body of empirical evidence confirms is widespread, being found in a variety of different supply chains. In these studies, rising input prices are usually passed through to output prices more quickly than are falling input prices, a pattern referred to as ‘positive’ asymmetry (i.e. ‘prices rise faster than they fall’). This has been observed in gasoline supply chains (Borenstein et al. 1997), supermarket prices (Peltzman 2000), and in consumer banking deposits (Neumark and Sharpe 1992) as well as in a number of agricultural supply chains (Ward 1982; Kinnucan and Forker 1987; Miller and Hayenga...
Explaining this tendency towards asymmetry is important for several reasons. First, although it is present in a wide variety of markets, it does not emerge as a simple prediction from any mainstream economic models (Peltzman 2000).

Secondly, since asymmetry in the timing of price movements causes redistribution and can also potentially result in net welfare losses, understanding the causes of asymmetry may be important for government policy. Under positive asymmetry, the (relative) delay in passing through negative cost shocks prevents downstream consumers from taking advantage of lower retail prices, temporarily reducing consumers’ surplus (relative to the symmetric pass-through case). The delayed expansion of retail demand also means the quantity demanded from upstream producers does not rise, reducing producers’ surplus.

Whether asymmetric price transmission is also associated with an overall welfare loss depends on the cause of the asymmetry. If a delay in passing through a negative cost shock reflects a temporary widening of the profit margins of intermediate firms, due somehow to the exercise of market power, then in addition to the redistributive effects there will be a deadweight welfare loss and a potential case for policy intervention. However, temporarily higher price spreads could also result from the presence of real adjustment costs, with no net welfare loss in consequence (Meyer and von Cramon-Taubadel 2004).

Given that asymmetry may or may not signal the exercise of market power, and may or may not be associated with a net welfare loss, it would be valuable to develop a set of empirical tests of supply chain price transmission which could not only detect asymmetry, but also quantify the welfare effects and identify whether the likely cause is anti-competitive conduct or some other factor.

Progress towards this goal is somewhat uneven. The focus of the theoretical literature has been on identifying potential causes of asymmetry, but there has been little discussion of the context and applicability of these ex-

---

1 See Meyer and von Cramon-Taubadel (2004) for a survey.
2.1. INTRODUCTION

planations or of how they could be tested empirically. Consequently "there is little in the literature that could serve as a basis for empirical tests that distinguish between these explanations" (Meyer and von Cramon-Taubadel 2004, p604). The empirical literature in turn has focussed more on identifying asymmetry than on assessing the consequences or deducing the causes, with Azzam (1999) labelling asymmetry tests "more useful in describing how markets look than how they work".

The explanations for asymmetry that have been proposed can be grouped loosely into two categories: ‘market power’ and ‘adjustment costs’. The first group includes ideas based around consumer search costs (Brown and Yücel 2000; Johnson 2002), focal point pricing (Borenstein et al. 1997), and tacit collusion (Balke et al. 1998).\(^2\) Theories based around adjustment costs include several related to inventory management (Reagan and Weitzman 1982; Ward 1982; Blinder 1982; Borenstein et al. 1997). In several other models asymmetry arises from the interaction of symmetric menu costs with some other industry characteristic (trend inflation in Ball and Mankiw 1994, and spatially differentiated markets in Azzam 1999).

Peltzman (2000) also suggests that asymmetry could arise from a simple source of asymmetric adjustment costs related to the recruitment and cancellation of input contracts. Where firms face higher costs of acquiring inputs (through search costs and price premia, for example) than of shedding them, they will expand and contract output at different rates depending on the direction of the upstream shock, thus producing asymmetry in the rate of downstream price responses.

Meyer and von Cramon-Taubadel’s survey (2004, p591) characterises this theoretical literature as a "bouquet of often casual explanations, each of which is able to produce a wide range of asymmetry behaviour". That is, little is known about the pattern of asymmetry associated with each explanation: whether asymmetry is expected to constitute long delays in response to negative input price movements, with symmetrically sized adjustments when

\(^2\)The essential elements of these theories are information asymmetries or bounded rationality on the part of consumers. They differ qualitatively from the modelling approach used in this chapter, which is based on complete information for consumers (alongside bounded rationality amongst firms).
these take place, or large upwards adjustments and incremental downwards adjustments (‘rockets and feathers’, as described by Bacon 1991), or even relatively smooth adjustment to both positive and negative shocks. Meyer and von Cramon-Taubadel (2004) thus argue that "deductive, theoretical work could provide a better indication of the conditions under which asymmetric price transmission would indeed represent a rational use of market power or response to adjustment costs, and exactly what form this asymmetric price transmission could be expected to take" (p605).

The work in this chapter and in the two that follow aims to contribute to this task, by exploring the effects of adjustment-cost asymmetry within a simple model environment. The model incorporates a convex adjustment cost, and allows the patterns of price adjustment generated by asymmetry in this cost to be predicted, whilst also considering how other market characteristics (including the perceived ‘price-making’ power of each firm) bear on price transmission. A convex adjustment cost is seen as a not-unrealistic approximation to the (potentially asymmetric) costs of ‘hiring and firing’ inputs, which Peltzman (2000) suggests could lead to asymmetry. It is also a useful simplification for modelling purposes.

The structure of the model is representative of the characteristics of oligopolistic supply-chain intermediaries, on which much discussion around asymmetric pass-through centres. The concern commonly expressed is that these firms – large food processors, for instance, or distributors within agricultural or non-agricultural (e.g. gasoline) supply chains – are able to withhold cost decreases due to collusion (or the exercise of market power in some other manner). Plausibly, however, capacity constraints and the costs of sourcing new inputs mean that higher costs of expanding output than contracting output could explain asymmetric responses to input price movements. By including the key features of supply-chain intermediaries in the model set-up – fewness of firms, a homogeneous product, quantity-setting strategies with a market-clearing price – the aim is to uncover some idea of what degree of asymmetry would be a ‘reasonable’ response to plausible asymmetries in adjustment costs in these sorts of intermediate markets.

The convex adjustment cost assumption adopted is standard within the
large body of work examining the costs of capital and labour input adjustment – their source, structure, and size (see Hamermesh and Pfann 1996 for a survey). This work lends substantial support to the idea that output adjustment costs can be a significant driver of firm behaviour, and moderate support to hypotheses for asymmetric price transmission based around various asymmetries in costs of output adjustment such as proposed by Peltzman (2000) and modelled here. Note that apparently nowhere within this literature, however, have the costs of adjusting levels of non-factor (i.e. material) inputs been considered. Peltzman’s suggestion that asymmetric price responses could be driven by asymmetries in the costs of acquiring and firing inputs could equally apply to non-factor inputs as to labour inputs.

Hamermesh and Pfann (1996) note basic evidence from accounting-based surveys that adjustment costs for labour demand are both ‘large’, and asymmetric, in that costs of hiring exceed costs of separations (p1268). Econometric studies confirm the significance of these costs, although evidence for the direction of asymmetry is less clear: some studies find hiring costs exceed separation costs (e.g. Pfann and Verspagen 1989, and references in Hamermesh and Pfann 1996, p1280), while others find the opposite (studies using French data by Abowd and Kramarz 1993, Goux et al. 2001, and Kramarz and Michaud 2010).³

In the literature on capital adjustment, the significance of adjustment costs (in accounting for investment behaviour) is less clear. Cooper and Haltiwanger (2006), for instance, conclude that investment costs are economically significant (while also emphasising the variation in estimates of these costs within the empirical literature), while Hall (2004) and Groth and Khan (2010) reach the opposite conclusion. The standard assumption that the cost structure is convex is found by Cooper and Haltiwanger (2006, p613) to fit actual behaviour at the aggregate (industry) level quite well.

³Since many separation costs arise from government legislation and thus differ widely across jurisdictions, generalisations about the direction of asymmetry are likely to be impossible. Labour adjustment costs also likely vary by industry and worker type: for instance, Alonso-Borrego (1998) finds (tentative) evidence from Spanish manufacturing firms that hiring costs exceed firing costs for non-production (white-collar) workers, but that the opposite holds for production (blue-collar) workers.
even while discrete costs of adjustment are found to be necessary to account for the lumpiness of investment at the firm level. Asymmetry in capital investment costs has been investigated less, partly because little disinvestment is observed in practice (Cooper and Haltiwanger 2006, p614). However it is worth noting that at least one study (Lansink and Stefanou 1997) finds that adjustment costs associated with investment are higher than for disinvestment, within a sample of farm businesses.

2.1.2 Outline of Part I

By imposing the assumption that there is asymmetry in output adjustment costs along the lines suggested by Peltzman (2000), the model developed in this chapter naturally predicts asymmetry in price adjustment. However the questions of interest here, which the modelling strategy is chosen to answer, are about how much asymmetry in price adjustment speeds is consistent with a given asymmetry in adjustment costs, and whether other characteristics of the firm and market bear on this relationship.

In the next section of this chapter the modelling strategy used and the rationale for this are described, and Sections 2.3, 2.4, and 2.5 work through the various parts of the model. The emphasis throughout is on retaining the ability to express characteristics of the firm and industry in easily-understood, quantifiable parameters. This allows the asymmetry predicted by the model to be examined and roughly quantified via numerical exercises, helping to build up a picture of which aspects of a market are most significant in determining the degree of asymmetry.

The method used to achieve this goal – adapting a technique used in static price transmission modelling to a dynamic setting – is the innovative aspect of the modelling in this chapter. It does, however, require compromises from an ideal modelling approach (as explained below) that make the conclusions drawn more ‘suggestive’ than ‘definitive’. Regardless, the exercise usefully demonstrates a new approach – and highlights some obstacles – to using dynamic models to explore the pattern and degree of adjustment-cost induced asymmetry.
2.2. MODELLING METHOD

The contribution of Chapter 3, which introduces several original forms of adjustment cost function, is complementary to the modelling strategy described here. To obtain some idea of the size of the adjustment speed and asymmetry associated with particular adjustment cost values requires a parameterisation of adjustment costs which is intuitive and quantifiable (and which could potentially be measured in empirical studies). Chapter 3 suggests two new functional forms for adjustment costs in which the parameter is a form of elasticity, which can be easily interpreted and assigned a realistic range of values.

This allows the relationships between adjustment cost asymmetry and price transmission, and the influence of market power and various other aspects of the industry, to be explored via numerical exercises in Chapter 4. In line with the goal of producing ‘stylised facts’ about asymmetry that may be compared with the results of empirical studies, Chapter 4 also critiques the measures typically used to gauge the asymmetry of price responses in a particular market. Several new summary measures of the degree of asymmetry in price transmission are suggested and discussed, and the relationships emerging from the theoretical model of Chapter 2 are expressed in terms of these alternate measures of asymmetry.

2.2 Modelling method

The model presented below is not intended as a new theory about the causes of asymmetric price transmission. Rather, the aim is to formalise an existing hypothesis, the intuitive idea that where costs of contraction and expansion of output differ, the rates of contraction and expansion (and hence the rates of pass-through of positive and negative cost shocks) will also differ. Modelling this allows predictions to be made about the extent of pricing asymmetry consistent with a certain degree of asymmetry in adjustment costs. The influence and interaction of other market characteristics (such as the discount rate and the degree of market power) can also be observed.

Given these goals, it is desirable to use a modelling strategy involving easily-interpretable parameters with known ranges of values. The method
used here borrows from the ‘equilibrium displacement modelling’ technique which has been used in static models of price transmission (for instance, McCorriston et al. 1998, 2001) as well as in other applications. In this approach functional forms are left unspecified, and responses to a shock are expressed in percentage terms. The technique is useful because the only modelling parameters used are elasticities, which can often be assigned realistic values, thereby allowing the use of numerical exercises to assess the quantitative significance of these parameters (that is, to predict the size as well as the sign of the effect of a change in parameter values).

McCorriston et al.’s (1998) model uses this approach to examine the magnitude of transmission of input price shocks to output prices in a static, symmetric setting. The model presented here borrows the basic structure of this model, and can be viewed as a dynamic extension of it in order to examine the speed as well as the magnitude of price transmission.

The basic structure involves firms that choose output levels at each point in time whilst also minimising the costs associated with the two inputs required for production of the final good (an agricultural input and a marketing input). A convex adjustment cost slows adjustment to shocks and makes each firm’s optimisation problem dynamic. To simplify the modelling the dynamic ‘output choice’ problem and the ‘input choice’ problem are each solved separately and the solutions combined afterwards, rather than both decisions being embedded in a single, more-complex optimisation problem.

The dynamic problem, which is introduced in Section 2.3 and solved in Section 2.4, generates an adjustment path equation and a steady-state markup condition. The adjustment path is a function of a known adjustment speed measure and the unknown proportional change in the steady state price level (i.e. the long-run magnitude of price transmission).

The steady-state markup condition is subsequently used in the (static) input choice model of Section 2.5, which uses the equilibrium displacement modelling technique to derive an expression for the percentage change to the steady state price in terms of the size of the exogenous shock to the agricultural input price. Thus the static model, including the dynamic model steady state as one equilibrium condition, yields an expression for the magnitude of
price transmission which enables an adjustment path to be fully specified. This adjustment path equation, which incorporates information about both the speed and the magnitude of price transmission in a single expression, forms the basis of the analysis in Chapter 4.

2.3 Dynamic model setup

This section introduces the dynamic optimisation problem, in which firms choose rates of change of output over an infinite time horizon whilst facing adjustment costs. The usual solution concepts for this type of ‘differential game’ are discussed, and the particular informational and behavioural assumptions used in this article are introduced in the final subsection.

2.3.1 The firm’s problem: a differential game

Suppose that all firms in the industry are identical quantity-setting oligopolists who maximise discounted profits over an infinite horizon. Each firm has discount rate \( r \), so that the problems for all firms \( i = 1, \ldots, N \) are:

\[
\max_{x_i} \int_{t=0}^{\infty} e^{-rt} \left[ p(Q)q_i - C(q_i) - A(x_i, q_i) \right] dt
\]  

(2.1)

The industry output \( Q \) is equal to \( Q = \sum_{i=1}^{N} q_i \), and the firm’s choice variable in this game is \( x_i \), which is the rate of change of output (i.e. \( x_i = \dot{q}_i \)).

Profits in each period equal revenue minus costs, including an adjustment cost \( A \) which depends on both the level and the rate of change of output. Some restrictions are placed on these costs: the production technology is assumed to exhibit constant returns to scale; the adjustment cost function is strictly convex in the rate of change of output so as to cause adjustment to be smooth (\( A_{x_i}(x_i, q_i) > 0 \)); finally, \( A \) also has the properties \( A(0, q_i) = A_{q_i}(0, q_i) = A_{x_i}(0, q_i) = 0 \), where the subscripts denote partial derivatives.

The set of optimisation problems in (2.1) constitute a differential game, with the adjustment cost generating the ‘structural dynamics’ which make
movement to the steady state time-dependent rather than instantaneous.\footnote{See Dockner et al. (2000) for a discussion of properties and solution methods for differential games. Similar oligopoly games include Driskill and McCafferty (1989) and Dockner (1992), which also use an adjustment cost as the friction, and Fershtman and Kamien (1987) which uses a sticky price to slow adjustment instead.}

A differential game consists of a set of optimal control problems where the payoffs for each player and the law of motion of the state variable depend on the actions taken by all players. Because the payoff function in \( (2.1) \) depends on \( Q = \sum_{i=1}^{N} q_i \), and the state \( q = (q_1, \ldots, q_N) \) depends on other player’s actions via \( \dot{q} = (x_1, \ldots, x_N) \), this is a differential game rather than a single-player dynamic optimisation problem.

### 2.3.2 Concepts in differential games

An equilibrium in a differential game consists of a set of strategies, one for each player, such that no player wishes to deviate given the strategies played by the others. There are several options available for the choice of strategy space, with the appropriateness of each dependent on the informational assumptions concerning the players.\footnote{See Dockner et al. (2000, s2.3, 3.5 & 4.1).}

The two most commonly employed in differential games are the ‘open-loop’ and ‘closed-loop’ strategy spaces.

When all players are aware of the structure of the game and the payoff functions of other firms but are either mutually forced to commit to an output path from the outset, or unable to observe or infer the value of the state at any \( t > 0 \), then they select strategies from the open-loop strategy space. An open-loop or time-dependent strategy is a function \( x_i(t) \) which determines the value of player \( i \)’s control variable according only to the time, not to the value of the state or any other information existing at \( t \). Assuming open-loop strategies in a dynamic game reduces the true time-dimension of the game – it essentially becomes a single-shot game, since the firm’s output plans are fixed from the outset, although the outcome is still a dynamic process.

If commitment is impossible and the state is observable throughout time then each firm plays a state-dependent strategy \( x_i(q, t) \) known as a closed-loop (or Markovian) strategy. Such strategies are also referred to as ‘memory-
less’ since they condition action on the current state but not on the history of the game (thus excluding trigger strategies). Open-loop and closed-loop strategy spaces are also not the only possibilities available. For example, another potential strategy space is that which allows players a choice between time-dependent and state-dependent strategies. This would be appropriate for a game context where players had the option of precommitment available to them.

In the context of a dynamic quantity-setting oligopoly, the choice of strategy space effectively determines the intensity of competition between firms, since it dictates whether firms are able to react to other firms’ choices or not (particularly since the variety of other actions that generate collusion or competition in reality are excluded from the model). The choice of strategy space and the choice of market structure (cost and demand function) together prescribe the nature of the interactions between firms, and ultimately determine the markup of price over cost in steady state equilibrium. That is, the long-run competitiveness of the industry is actually endogenous to each dynamic oligopoly model rather than being a parameter under the control of the modeller.

For example, the quantity-setting open-loop game has the Cournot outcome as the steady state, implying that Cournot competition can be considered as a static ‘shorthand’ for the dynamic interactions that occur when firms play time-dependent strategies. Similarly, a closed-loop steady state corresponds to a static ‘conjectural variations’ equilibrium with negative conjectures (Dockner 1992). In general, the long-run outcome of any type of dynamic interaction can potentially be summarised in a static model with a particular choice of conjectural variation; see Figuières et al. (2004, ch2) for discussion.

For the application here – exploring the pattern of asymmetry consistent with asymmetric adjustment costs under a wide range of market structures – the competitiveness of the industry would ideally be parameterised. Otherwise, the endogeneity of competitiveness within each model means that the setup of the model must be changed to explore the effect of market power on price adjustment paths. That is, since each choice of strategy space essen-
tially fixes the level of market power in the model, the sensitivity of the price adjustment process to market power cannot be observed by simply varying a parameter. However one of the strengths of static models of price transmission, that this chapter attempts to replicate in a dynamic setting, is the ability to use a conjectural variations parameter as a proxy for the degree of market power, so that the relationship between competitiveness and the price transmission elasticity can be easily observed (as in McCorriston et al. 1998, for example).

Adopting the usual open-loop or closed-loop strategies creates one further difficulty for modelling price transmission. Because finding equilibrium strategies in a differential game entails simultaneously solving $N$ interrelated optimal control problems, deriving analytical solutions is difficult. Only for certain simple classes of game, such as the linear-demand quadratic-costs oligopoly, can equilibrium strategies and a price adjustment timepath be found (see Driskill and McCafferty 1989; Dockner 1992). The general linear-quadratic game can be shown to have closed-loop equilibrium strategies which are linear functions of the state variables (see Figuières et al. 2004, p61 for details). Because $p(Q)$ is not necessarily linear, and $A$ depends on $q_i$ as well as $x_i$, the game specified in equation (2.1) above does not belong to this class.

### 2.3.3 Information, beliefs and strategy space assumptions

Consistent with the aims of the chapter, an alternative modelling strategy is used here which simplifies the differential game, allowing parameterisation of market power whilst retaining a reasonably-general specification of industry structure. At the same time, the particular assumptions that permit this imply a highly-specific informational environment for the firm, which is somewhat less intuitive and simple than those corresponding to the open-loop and closed-loop strategy spaces.

The first of these assumptions is that firms are not fully rational in the sense of being able to calculate their rivals’ reaction functions (which is a key assumption underlying closed-loop strategies). Instead, firms ‘conjecture’ a
response of the other firms, expressed as a fixed elasticity, to their own change in output. Thus firm $i$ believes that a proportional change in its output will instantaneously lead to a certain proportional change in industry output, given by $\frac{dQ_i}{dq_i} = \theta$. This concept of conjectures as beliefs held by boundedly rational firms, which represent the firms’ imperfect understanding of the strategic environment they operate in, is similar to that used by Friedman and Mezzetti (2002).

The second assumption is that firms are forced to commit to an output path at $t = 0$, so that their strategies (rates of change of output) are time-dependent rather than state-dependent functions (as in the open-loop strategy case). Firms are locked in to their previously-chosen output plans, and equilibrium strategies are defined as the optimal paths $x_i(t)$ given the exogenously given conjectures, even though these conjectures are incorrect. An alternate interpretation of time-dependent strategies is not that firms are forced to commit but that the price and industry output remain unobservable throughout the adjustment period, and therefore firms have no reason to update their conjectures or to alter their output paths in time periods after $t = 0$.

The conjectures are assumed to be symmetric across firms, and are used here as a proxy for market power since they correspond to each firm’s belief about its control over market output and hence the price. For a monopolist, the firm-level proportional output change will be equal to the industry change ($\theta = 1$). A oligopolist in a symmetric $N$-firm industry who assumes no reaction from rivals would hold belief $\theta = 1/N$, but would hold a conjecture less than (more than) $1/N$ if they believed that competition was more (less) fierce than this benchmark.

While the ability to capture the intensity of competition in a parameter is useful, the two assumptions adopted to achieve this have a strange implication; although each firm is forced to commit, it also believes that its rivals do not face such a constraint and will therefore react to its actions. Alternatively, under the second interpretation of time-dependent strategies, each firm believes that they are the only one who cannot observe the evolution of the industry output and price.
This is justifiable only on the basis of bounded rationality; that is, the firms in this model setup do not fully understand the nature of the game they are part of. The inconsistency of the conjectures with actual behaviour also requires this assumption. Perhaps the best interpretation of the conjecture parameter in this model setup is not as a literal ‘expected response’, but as a representation of a loosely-rationalised perception on the part of a firm with less-than-perfect computational capabilities about its ability to influence the market price. It is plausible that firms form their strategies based on simple rules of thumb such as this, rather than on the basis of a complicated backwards-induction type calculation.

This modelling strategy trades away the use of commonly used and theoretically sound strategy space concepts (open-loop and closed-loop) in favour of retaining the ability to parameterise a wide range of aspects of market structure and conduct in a way that allows for later numerical analysis. That is, in order to examine how asymmetric adjustment cost-induced asymmetry interacts with (perceived) price-making power, and in order to use elasticity parameters to quantify results, it has been necessary to introduce exogenous conjectural variations and adopt an informational structure that is not standard for this type of dynamic game. This choice is deemed appropriate in order to explore the possibilities of a modelling method which is intended to generate quantifiable predictions, thus allowing the relative important of different aspects of market structure and conduct (as well as their direction of influence) to be observed.

2.4 Dynamic model solution

This section considers the simpler optimisation problem that results from the assumptions specified above. The nature of the game is described, before solving for the steady state and adjustment path in order to derive an expression for adjustment of the firm’s quantity through time.
2.4. DYNAMIC MODEL SOLUTION

2.4.1 The firm’s problem: an optimal control problem

The twin assumptions of conjectures describing the firms’ beliefs about rivals’ reactions, and commitment of all firms to output paths chosen at $t = 0$ reduces the differential game to a set of $N$ single-player ‘optimal control’ problems with a unidimensional state variable ($q_i$) and a single control ($x_i$). These problems have the following form for each firm $i$ (the state and control variables are functions of time but the $t$ arguments are omitted for clarity):

$$
\max_{x_i} \int_0^\infty e^{-rt}[p(Q^c(q_i)) q_i - C(q_i) - A(x_i, q_i)] dt
$$

(2.2)

The analysis below proceeds with two new variables in addition to those introduced in Section 2.3. As well as the true industry quantity and price $Q$ and $p$, some equations will be expressed in terms of $Q^c$ and $p^c$, two hypothetical variables representing the firm’s beliefs.

The variable $p^c$, the price level at time $t$ as conjectured by the firm, is directly related via the industry demand function to $Q^c$, the industry quantity as conjectured by the firm ($p^c = p(Q^c)$). The variable $Q^c$ depends on the firm’s quantity choice according to some unknown function $Q^c = Q^c(q_i)$ with elasticity $\frac{dQ^c}{dq_i} \frac{q_i}{Q^c} = \theta$.

To summarise the problem: each firm is (a) endowed with full knowledge of the industry demand curve, (b) forced to commit to an output path and thus play a time-dependent strategy, and (c) holds a simple (but mistaken) belief about their price-making ability as expressed in the function $Q^c(q_i)$. Given these assumptions, each firm will act as though to solve the problem presented in equation (2.2) above. All firms solve the same problem and hence act identically. The true relationship between firm and industry output levels is therefore given by $Q = Nq_i$, which will differ from firm $i$’s belief that $Q = Q^c(q_i)$. 
2.4.2 Optimal quantity strategies

The optimal time-dependent strategy for firm \(i\), given the conjectured responses of other firms, can be found via Pontryagin’s maximum principle.\(^6\) The current value Hamiltonian is:

\[
H(x_i, q_i, \phi_i, t) = p(Q^c(q_i))q_i - C(q_i) - A(x_i, q_i) + \phi_i x_i
\]  

(2.3)

where \(\phi_i\) is the ‘co-state variable’. The maximum condition and adjoint equation are:

\[
\frac{\partial H}{\partial x_i} = 0 \Rightarrow A_{x_i} = \phi_i
\]  

(2.4)

and:

\[
\dot{\phi}_i = r \dot{x}_i - \frac{\partial H}{\partial q_i}
= r \phi_i - p^c(1 - \frac{\theta}{\eta}) + c + A_{q_i}
\]  

(2.5)

respectively (\(c\) represents the constant marginal cost). The latter equation uses the fact that the derivative of \(p^c = p(Q^c)\) with respect to \(q_i\), given that the conjectured quantity has elasticity \(\theta = \frac{dQ^c}{dq_i} \frac{p^c}{Q^c}\), is:

\[
\frac{dp^c}{dq_i} = \frac{dp^c}{dQ^c} \frac{dQ^c}{dq_i} = \frac{dp^c}{dQ^c} \frac{dQ^c}{dq_i} \frac{Q^c}{q_i} = -\frac{\theta p^c}{\eta q_i}
\]  

(2.6)

where \(\eta\) is the absolute value of the elasticity of demand \((\eta = \frac{dQ}{dp} \frac{p}{Q}) = \left| \frac{dQ^c}{dp^c} \frac{p^c}{Q^c} \right|\).

To obtain a differential equation system in \(x_i\) and \(q_i\), the maximum condition is differentiated with respect to \(t\), giving:

\[
A_{x_i, q_i} \dot{q}_i + A_{x_i, x_i} \dot{x}_i = \dot{\phi}_i
\]  

(2.7)

When this expression and the original maximum condition are substituted for \(\dot{\phi}_i\) and \(\phi_i\) in the adjoint equation, the latter becomes:

\[
A_{x_i, q_i} \dot{q}_i + A_{x_i, x_i} \dot{x}_i = r A_{x_i} - p^c(1 - \frac{\theta}{\eta}) + c + A_{q_i}
\]  

(2.8)

---

\(^6\)See Dockner et al. (2000, ch4) for discussion of solution methods for differential games and optimal control problems.
2.4. DYNAMIC MODEL SOLUTION

which can be rearranged to give:

\[ \dot{x}_i = \frac{1}{A_{x_ix_i}} \left( rA_{xi} + A_{qi} - A_{x_iq_i}x_i - p^c(1 - \frac{\theta}{\eta}) + c \right) \quad (2.9) \]

This equation, and the law of motion of the state (\( \dot{q}_i = x_i \)) form the differential equation system (2.10), the solution to which is the time-dependent quantity strategy which firm \( i \) believes is most profitable:

\[ \begin{align*}
\dot{x}_i &= \frac{1}{A_{x_ix_i}} \left( rA_{xi} + A_{qi} - A_{x_iq_i}x_i - p^c \left( 1 - \frac{\theta}{\eta} \right) + c \right) \\
\dot{q}_i &= x_i
\end{align*} \quad (2.10) \]

2.4.3 Steady state condition and interpretation of \( \theta \)

Setting \( \dot{q}_i = \dot{x}_i = 0 \), and bearing in mind the conditions imposed on \( A \) in Section 2.3, the steady state of equation system (2.10) reduces to:

\[ p^c \left( 1 - \frac{\theta}{\eta} \right) = c \quad (2.11) \]

This expression defines the steady state output level \( q_i^* \) and is similar to a static conjectural variations equilibrium with conjectural elasticity \( \theta \) (see McCorriston et al. 1998), although the condition here describes not the actual price-cost markup but the long-run markup that the firm expects to be realised, given their conjecture.

In a conjectural variations model, larger values for the conjecture imply a larger markup of price over cost, justifying the use of the conjecture as a parameter representation of the market power of firms. In this framework, since \( \theta \) represents the firm’s belief about its own degree of price-making power, it is expected that higher \( \theta \) will result in lower steady-state output and a higher price so that higher \( \theta \) can similarly be considered a market power parameter. It can be shown by a similar argument that as \( \eta \) increases the steady state output value increases and the markup of price over cost falls, indicating that the markup is related to the elasticity of demand in the usual way.

This interpretation of \( \theta \) can, in fact, be sustained in this dynamic model; in Appendix A it is shown that even though a simple expression for the true steady state markup is unavailable, the price is higher for larger values of \( \theta \).
There is an important qualification to this interpretation of $\theta$. The solution to the firm’s optimisation problem derived in subsection 2.4.2 is, in fact, only valid over the range $\theta > 0$. Moreover, the limiting case of perfect competition cannot in fact be sensibly accommodated in the model as it is set up in subsection 2.4.1.

To illustrate this, consider the optimal strategy in the model environment if the firm conjectured a zero response – i.e. believed itself to be a price-taker in a perfectly competitive market. By assumption the firm does not observe the evolution of the actual price, and hence at every point in time the firm’s conjectured price equals the initial price, so that the conjectured price-cost margin is fixed for all $t$. The firm’s optimal response would be to alter output at a rate which equates the marginal gain from adjusting output ($p - c$) with the marginal adjustment cost ($A_{x_i}$). If $p < c$, this rate of adjustment would continue until the 0 bound was reached (at a steady state of 0 industry quantity), and if $p > c$ this rate would continue indefinitely (with no steady state).

For both positive and negative shocks from an initial equilibrium position this optimal response offers no sensible measure of the (proportional) adjustment speed, and accordingly only values of $\theta$ greater than 0 are considered from this point.

### 2.4.4 Quantity adjustment path

The equation system (2.10) describes the movement through time of the two variables of interest, $q_i$ and $x_i$. Given the unspecified functional forms the system is of an analytically unsolvable type, and a linear approximation around the steady state of the system is used to solve for the output timepath instead (see Appendix A). The linear approximation is:

$$
\begin{pmatrix}
\dot{x}_i \\
\dot{q}_i
\end{pmatrix} = 
\begin{bmatrix}
r & K \\
1 & 0
\end{bmatrix}
\begin{pmatrix}
x_i - x^*_i \\
q_i - q^*_i
\end{pmatrix}
$$

where $K = \frac{1}{A_{x_i q_i}} \left(r A^*_{x_i q_i} + A^*_{q_i q_i} + \frac{c q_i}{q^*_i} (1 + \frac{\alpha}{\eta - \theta}) \right)$. Subscripts denote partial derivatives, and a * indicates the steady state value of a variable. In the
expression for $K$ parameter $\omega$ is defined as $\omega = \frac{dnP}{dp\eta}$, or the elasticity of the elasticity-of-demand with respect to the price. The value of $\omega$ depends on the form of the demand curve; for example, a constant-elasticity demand curve has $\omega = 0$, log-linear demand implies $\omega = 1$, and linear demand implies $\omega = 1 + \eta$.

In the saddlepath stable solution to this system, firm $i$’s output is shown to converge to the steady state value following a continuous-time version of a ‘partial adjustment’ process (see Appendix A). In such a process the rate of output change is proportional to the difference between the current and steady state output levels:

$$\dot{q}_i(t) = \lambda (q_i^* - q_i(t))$$

(2.13)

Since all firms act identically, the industry output adjusts in an identical fashion. The adjustment coefficient $\lambda$, which is the key adjustment speed parameter examined in Chapter 4, is a constant given by:

$$\lambda = \frac{1}{2} \left( \sqrt{r^2 + 4K} - r \right)$$

(2.14)

The output timepath associated with this process is:

$$q_i(t) = q_i^* + (q_i(0) - q_i^*)e^{-\lambda t}$$

(2.15)

which can be more usefully expressed in terms of proportional changes from the initial (time $t = 0$) quantity level $q_i(0)$:

$$\frac{q_i(t) - q_i(0)}{q_i(0)} = \left( \frac{q_i^* - q_i(0)}{q_i(0)} \right) (1 - e^{-\lambda t})$$

(2.16)

For consistency with the next section, log notation is used for these proportional changes and it is assumed that the quantity changes are ‘small’ in the sense that the change in logarithms is approximately equivalent to the proportional change from the initial quantity level. The term $d\ln q_i(t)$ refers to the proportional change in the firms quantity between time 0 and time $t$, as shorthand for the proportional change $\frac{q_i(t) - q_i(0)}{q_i(0)}$. The term $d\ln q_i^*$ represents the proportional quantity change from the old steady state quantity to
the new steady state quantity, or \( \frac{q_i - q_i(0)}{q_i(0)} \). The quantity adjustment path will thus be expressed from this point as:

\[
d \ln q_i(t) = d \ln q^*_i(1 - e^{-\lambda t})
\]  

Equations (2.11), (2.14) and (2.17) are the key outcomes of the dynamic modelling exercise. The steady state condition in equation (2.11) will be used in the next section to link the change in steady state output to the change in the agricultural input price. The quantity adjustment path in equation (2.17), which describes the proportional output change from time \( t = 0 \) in terms of both the adjustment speed parameter \( \lambda \) from (2.14) and the as-yet-unknown change in the steady state, will be used in Chapter 4 to analyse asymmetry in adjustment processes.

2.5 Static input choice model

In order to begin examining the price adjustment process the size of the change in steady state quantity needs to be related to the size of the disequilibrating input price shock. The extent to which output increases as input costs fall (or decreases as costs rise) depends on a range of factors including the nature of the production technology (particularly the substitutability between inputs in production and hence the relationship between input prices and production costs), the elasticities of supply of inputs, the elasticity of demand for output, and the degree of competition and the price-cost markup in the industry.

This section constructs a system of equations to describe these aspects of the supply chain and to derive an expression for \( d \ln q^*_i \), the steady state change in output. In the final subsection this expression is combined with equation (2.17) and rearranged to show the adjustment path of prices.

2.5.1 Input cost minimisation

Suppose each firm in the downstream industry uses an identical variable-proportions production technology \( f \) which, as noted in Section 2.3, exhibits constant returns to scale. This technology combines an agricultural input
(quantity $q_i^A$) and a marketing input (quantity $q_i^M$) to produce the final good:

$$q_i = f(q_i^A, q_i^M)$$  \hspace{1cm} (2.18)

The input prices $\bar{p}^A$ and $\bar{p}^M$ are assumed to be exogenously determined, and the shock examined in the next subsection is to the price of the agricultural input. In McCorriston et al. (1998) the input supplies were not perfectly elastic, and input prices were therefore determined endogenously.

The inverse demand function is defined in terms of industry quantity $Q$ as:

$$p = p(Q)$$  \hspace{1cm} (2.19)

As argued above, each firm acts identically and industry quantity is equal to $N$ times the quantity of firm $i$; $Q$ can therefore be substituted out of equation (2.19) to give:

$$p = p(Nq_i)$$  \hspace{1cm} (2.20)

For any $q_i$, the firm chooses an input mix to minimise total production costs ($\bar{p}^A q_i^A + \bar{p}^M q_i^M$) subject to the constraint $f(q_i^A, q_i^M) \geq q_i$, yielding cost minimisation conditions:

$$\bar{p}^A = c_{fA}$$  \hspace{1cm} (2.21)

$$\bar{p}^M = c_{fM}$$  \hspace{1cm} (2.22)

where $f_A$ and $f_M$ are the partial derivatives of $f$ with respect to $q_i^A$ and $q_i^M$ respectively, and $c$ is marginal cost (the first derivative of the cost function $C(q_i; \bar{p}^A, \bar{p}^M)$). In the dynamic model, constant returns to scale meant that $c$ was independent of $q_i$. However although $c$ remains constant with respect to $q_i$, it is treated as an endogenous variable in this static model, because it is determined in part by the value of the (exogenously-determined) agricultural input price, which is assumed to change here.

### 2.5.2 Equilibrium displacement model

In equilibrium displacement modelling functional forms are left unspecified and the structural equations are totally differentiated, yielding a set of equa-
tions which relate equilibrium changes in the industry variables. These equations are jointly solved to give proportional changes in endogenous variables, as a function of elasticity parameters and changes in exogenous (shock) variables. The technique allows the modeller to perform a (casual) quantitative analysis without using econometrics, so long as realistic values for the elasticity parameters can be chosen.\footnote{See Piggott (1992) for discussion of the limitations and overall merits of the equilibrium displacement modelling approach.}

In McCorriston \textit{et al.}'s (1998) model, a system similar to the equations above was supplemented with a first order condition derived from a static quantity-setting profit maximisation problem. With as many equations as endogenous variables, the system could be solved for the proportional change in any endogenous variable in terms of the proportional change to a shock variable (the agricultural input price $\bar{p}^A$, for instance).

Since the supply chain model here describes relationships between variables which hold in the steady state, the system above is supplemented with equation (2.11), the steady state condition from the dynamic profit maximisation problem of the previous section.

As explained, the price variable in this condition is the conjectured price $p^c$ – what each of the boundedly-rational firms in this model believe the price will be – rather than the actual price $p$, implying that the addition of this condition does not balance the system of equations. Two additional equations are therefore included which explain how the conjectured variables are determined. The first states that $p^c$ is given by the true (inverse) demand function applied to the conjectured quantity $Q^c$:

$$p^c = p(Q^c) \tag{2.23}$$

The second states that the conjectured quantity is given by function $Q^c(q_i)$, which has a constant elasticity $\theta$:

$$Q^c = Q^c(q_i) \tag{2.24}$$

Equations (2.11), (2.18), and (2.20) to (2.24) constitute a system of 7 equations which determine the values of 7 endogenous variables – $q_i$, $q_i^A$, $q_i^n$, $q_i^p$, $q_i^m$, $q_i^h$, and $q_i^l$.\footnote{See Piggott (1992) for discussion of the limitations and overall merits of the equilibrium displacement modelling approach.}
2.5. STATIC INPUT CHOICE MODEL

$q_i^M$, $p$, $c$, $p^c$, and $Q^c$ – in terms of a variety of elasticity parameters and exogenous variables. In order to derive the relationship between changes in the exogenous agricultural input price ($\bar{p}^A$) and the resulting change in any endogenous variable, the system is totally differentiated and rewritten in terms of changes in logarithms, giving the following (see Appendix A):

\begin{align*}
    d \ln p^c + \left( \frac{\omega \theta}{\eta - \theta} \right) d \ln p &= d \ln c \quad (2.25) \\
    d \ln q_i &= \alpha d \ln q_i^A + \beta d \ln q_i^M \quad (2.26) \\
    d \ln p &= -\frac{1}{\eta} d \ln q_i \quad (2.27) \\
    d \ln \bar{p}^A &= d \ln c - \frac{\beta}{\sigma} (d \ln q_i^A - d \ln q_i^M) \quad (2.28) \\
    0 &= d \ln c + \frac{\alpha}{\sigma} (d \ln q_i^A - d \ln q_i^M) \quad (2.29) \\
    d \ln p^c &= -\frac{1}{\eta} d \ln Q^c \quad (2.30) \\
    d \ln Q^c &= \theta d \ln q_i \quad (2.31)
\end{align*}

The newly introduced parameters are defined as follows: $\alpha$ ($\beta$) is the elasticity of output with respect to the agricultural (marketing) input, $\alpha = f_A \frac{q_i^A}{q_i}$ ($\beta = f_M \frac{q_i^M}{q_i}$). These two parameters can also be interpreted as cost shares of each input, with $\alpha + \beta = 1$.\(^8\) The elasticity of substitution between inputs in production is $\sigma$ ($\sigma = \frac{f_A f_M}{f_A f_M q_i}$). These parameters and the three defined earlier ($\theta$, $\eta$, and $\omega$) are all positive.

The relationship between the percentage change in the agricultural input price $\bar{p}^A$ and the percentage change in the firm’s steady state output level $q_i^*$ is found using Cramer’s rule (see Appendix A):

\begin{equation}
    d \ln q_i^* = \frac{-\alpha \eta}{1 + \frac{\omega \theta}{\eta - \theta}} d \ln \bar{p}^A \quad (2.32)
\end{equation}

\(^8\) Since $\bar{p}^A = c_f A$ and $\bar{p}^M = c_f M$ according to the cost minimisation conditions (2.21) and (2.22), $\alpha$ ($\beta$) can be written $\alpha = \frac{\bar{p}^A q_i^A}{c_q}$ ($\beta = \frac{\bar{p}^M q_i^M}{c_q}$). Constant returns to scale in production imply a constant marginal cost, so that $C(q_i) = c_q$, and thus $\alpha$ and $\beta$ represent cost shares.
The change in the (actual, not conjectured) steady state price \( p^* \) can be similarly found. The ratio of steady state output and input price changes, referred to as the magnitude or elasticity of price transmission and denoted by \( \tau \), is given by:

\[
\tau = \frac{d \ln p^*}{d \ln \bar{p}^A} = \frac{\alpha}{1 + \frac{\omega \theta}{\eta - \theta}}
\] (2.33)

### 2.5.3 Price adjustment path

Equation (2.32) describes the steady state firm-level quantity change in terms of the size of the input price shock and a variety of elasticities. Substituting this into equation (2.17), and using the demand function to substitute the proportional quantity change on the left hand side of the equation for the proportional price change (as in equation (2.27)), gives:

\[
d \ln p(t) = \frac{\alpha}{1 + \frac{\omega \theta}{\eta - \theta}} d \ln \bar{p}^A (1 - e^{-\Lambda t})
\] (2.34)

This equation is the final output of the modelling exercise, forming the basis for the numerical analysis of price transmission and asymmetric price adjustment in Chapter 4. It describes the percentage change in the downstream price since time \( t = 0 \) as a multiplicative function of three elements: the size of the upstream price shock \( d \ln \bar{p}^A \), the expression for the magnitude of price transmission (itself a function of the agricultural input cost share \( \alpha \), the market power parameter \( \theta \), the (absolute value of the) demand elasticity \( \eta \), and the shape of the demand curve \( \omega \)), and a term which increases from 0 to 1 over time at a rate determined by the adjustment speed parameter \( \lambda \).

The adjustment speed parameter is itself influenced by certain aspects of industry structure and behaviour (parameters \( \theta, \eta, \omega \), and the discount rate \( r \)), as shown in equation (2.14). Also relevant to the adjustment speed is the adjustment cost function (which enters through various partial derivatives of \( A \)) and the values of marginal cost and steady-state output.

Equation (2.34) is the basis for the study of asymmetric price transmission in Chapter 4. This is done by substituting different adjustment cost values (intended to represent the asymmetric costs of expanding and contracting
output) and observing the difference in the speed of adjustment. However this cannot be done without first specifying an adjustment cost function $A(x, q)$, as the size of the adjustment speed term $\lambda$ is dependent on this function. Chapter 3 presents several forms of adjustment cost function, with easily interpretable parameters, that allow the expression for $\lambda$ to be simplified.

2.6 Conclusion

The model presented in this chapter is designed to explain both the speed and the long-run magnitude of pass-through of input cost shocks in order that the relationship between adjustment-cost asymmetry and asymmetric price transmission – and the influences of other market characteristics on this relationship – might be explored.

Four key steps characterised the modelling method. First, although the dynamic output choice model was specified as a differential game, the game was reduced to a simpler optimal control problem by assuming a particular form of bounded rationality for the firm. A linearised form of this problem was solved to generate the equation representing the path of adjustment to steady state.

Second, these same assumptions about the informational environment and rationality involved the introduction of a ‘conjecture’ parameter, which represented the perceived price-making ability of the firm. The value of this market power parameter influenced the size of the steady state markup of price over cost.

Third, the steady state condition from the dynamic output choice model formed one equilibrium condition for the static input choice model, enabling the steady state output change to be linked to the shock to the input price.

Fourth, the static model was manipulated in the style of equilibrium displacement modelling, so that the relationships between changes in the input price and the other variables could be expressed in terms of elasticities which can potentially be assigned non-arbitrary values.

Various aspects of these methods restrict the applicability of the model somewhat. The model is representative only of those markets in homoge-
neous goods which exhibit frequent, small shocks to the prices of inputs, and where the price level smoothly adjusts to the quantities chosen by firms. In particular, the use of a linearised version of the differential equation system to determine the firm’s output path, and the equilibrium displacement modelling technique in which the relationships between variables are accurate only for small perturbations to the equilibrium, both require the disequilibrating shock to be small. The assumption of output-path commitment seems to imply a relatively short period of adjustment to shocks, which would also make the assumption of ‘inconsistent conjectures’ without learning easier to justify.

As mentioned in the introduction to the chapter, much discussion around asymmetric pass-through of shocks centres on the concern that supply-chain intermediaries – such as large food processors, or distributors within agricultural or non-agricultural (e.g. gasoline) supply chains – are withholding cost decreases due to collusion or the exercise of market power in some other form. Despite the restrictions mentioned above, the model’s structure still reflects many of the key characteristics of these intermediate markets: fluctuating input prices, fewness of firms, a homogeneous product, quantity-setting strategies with a market-clearing price, and costly quantity adjustments (with the asymmetry of capacity constraints and the potentially-asymmetric costs of sourcing new inputs being plausible drivers of asymmetric adjustment costs).

The restrictions on the applicability of the model are unavoidable aspects of the overall approach, which aims to retain the ability to express characteristics of the firm and industry via easily-understood, quantifiable elasticity parameters. The use of conjectural variations within the model (to express market power as an exogenous parameter), and the restrictive assumptions placed on the firms’ beliefs (which allow the model to be solved with all parameters expressed as elasticities) are in pursuit of the particular goal of examining asymmetry via numerical exercises, to build a picture of the realistic degree of asymmetry expected under asymmetric adjustment costs.

This approach is the innovative aspect of the modelling, although it requires compromises from an ideal modelling approach (as explained in section 2.3) that effectively make the conclusions drawn more ‘suggestive’ than ‘de-
2.6. CONCLUSION

...nitive’. Nonetheless, the modelling exercise can be seen as valuable by virtue of demonstrating a method of dynamic modelling that could potentially offer useful insights into price adjustment and asymmetric adjustment processes.

This chapter has also usefully highlighted the difficulties inherent in using dynamic modelling to further the agenda put forth by Meyer and von Cramon-Taubadel (2004), to engage in deductive, theoretical work to discern "exactly what form this [market power-driven or adjustment cost-driven] asymmetric price transmission could be expected to take". The attempt to examine asymmetry via a dynamic adjustment cost model, in a manner which allowed for numerical analysis and which attempted to incorporate market power as an exogenous parameter, faces obstacles which are not initially apparent but which have been drawn out here.

These obstacles present obvious avenues for improvement in any future work which attempts to use dynamic models to study price adjustment speeds and asymmetry in adjustment. Ideally, firm beliefs and behaviour should be constrained to follow the usual game-theoretic norms, with the firms’ behaviour represented by either open-loop or closed-loop strategies.

While there are existing models which follow these norms (such as Dockner’s 1992 model, discussed in Part II), the challenge in using these to understand asymmetric price transmission is the need to give the models predictive content beyond simply describing the directions of influence of different variables. That is, it is clear that asymmetric adjustment costs will generate asymmetry in price adjustment; the question of interest is how asymmetric the resulting adjustment processes could be expected to be in any particular setting. Only when this is clear will applied researchers be able to understand what patterns of price changes are a natural, efficient, reflection of real costs of output adjustment and which instead point towards the exercise of market power.
Chapter 3

Adjustment cost functions

3.1 Introduction

Chapter 2 derived an expression for path of adjustment path of the price following a shock to an input price, in the context of a particular dynamic oligopoly game. The adjustment speed parameter $\lambda$ in this equation was given by:

$$\lambda = \frac{1}{2} \left( \sqrt{r^2 + 4K} - r \right)$$

with $K$ dependent on the functional form of the adjustment cost function and the steady state values of several other variables according to:

$$K = \frac{1}{A_{x_i}} \left( rA_{x_i}^* + A_{q_i}^* + \frac{c}{q_i^*} \eta (1 + \frac{\omega}{\eta - \theta}) \right)$$

(3.2)

The task of this chapter is to suggest several forms of adjustment cost function which can be used to specify $K$ and $\lambda$ in terms of easily-understood parameter values. These functions must satisfy the restrictions placed on them in the previous chapter: convexity of the adjustment cost function with respect to the rate of change of output ($A_{x_i} > 0$), zero adjustment cost at the steady state ($A(0, q_i^*) = 0$), and zero first-order partial derivatives at the steady state ($A_{x_i}(0, q_i^*) = A_{q_i}(0, q_i^*) = 0$).

Dynamic adjustment cost models typically use a quadratic cost function of the form $A(x_i) = kx_i^2$ (Driskill and McCafferty (1989); Dockner (1992)), which depends only on the rate of output change and not on the output level.
itself; that is, the adjustment cost depends on the absolute rather than the proportional rate of output change.

For the purposes of those models this form suffices, but in order to easily interpret the adjustment speed parameter and assign it a meaningful value it is necessary to specify an adjustment cost function with two particular properties. Firstly, the adjustment cost should be expressed in terms of the proportional rather than the absolute rate of output change and, secondly, it should be related to the size of the firm’s production costs.

Two forms of adjustment cost function are presented here which satisfy the technical requirements mentioned in Chapter 2, and which can be related to the proportional change in output and to the firm’s production costs. The adjustment cost parameters of these functions are useful ways of describing the ‘burden’ of the adjustment cost relative to production costs. In the first (quadratic) form, the parameter maps the proportional rate of change of output \((x_i/q_i)\) into the average ‘premium’ paid over the marginal production cost for the new units produced. In the second (logarithmic) form, the parameter maps the proportional rate of change of output into the premium paid over the marginal production cost for the marginal unit produced.

### 3.2 Quadratic form

Suppose the level of adjustment costs is given by the function:

\[
A^\delta(x_i, q_i) = \delta c q_i \left( \frac{x_i}{q_i} \right)^2
\]

(3.3)

where \(x_i\) is the rate of change of \(q_i\).

The adjustment cost here is the square of the proportional rate of change, scaled up by the firm’s total cost \(c q_i\) and a positive parameter \(\delta\).\(^1\)

\(^1\) \(A^\delta\) can equally be expressed as \(A^\delta(x_i, q_i) = \delta c \frac{x_i^2}{q_i}\), but the first formulation makes clear that the cost is fundamentally a quadratic function of the proportional rate of output change, with \(\delta c q_i\) merely scaling this quadratic cost so that the magnitude can be expressed relative to production costs.
3.2. QUADRATIC FORM

3.2.1 Properties

The function $A^\delta$ satisfies the restrictions assumed for the dynamic model and employed in the working above – $A^\delta(0, q_i) = A^\delta_q(0, q_i) = A^\delta_x(0, q_i) = 0$, and $A^\delta_{xx}(x_i, q_i) > 0$. In Chapter 2 the parameter values for which saddlepath stability holds in the solution to the linearised differential equation system were found assuming $A^*_{xq} = A^*_{qq} = 0$ (see Appendix A), which is also satisfied by this function.

3.2.2 Interpretation

The adjustment cost parameter $\delta$ can be interpreted as follows.

First, suppose that the proportional rate of change of output is $\frac{x_i}{q_i}$ and, to provide a concrete example, suppose that $x_i$ is positive which means that $x_i$ new units are brought into production over a single period of time. The production cost of each new unit is $c$ and the total adjustment cost incurred, divided equally over the new units produced, is $A^\delta(x_i, q_i)/x_i$. This adjustment cost per new unit produced, expressed as a proportion of the unit production cost, is:

$$\frac{A^\delta(x_i, q_i)}{cx_i} = \delta \left( \frac{x_i}{q_i} \right) \quad (3.4)$$

Thus the most straightforward interpretation of $\delta$ is as a multiplicative scalar which maps the percentage output change into the percentage ‘premium’ effectively paid to produce each additional unit.

For instance, suppose output were increased by 10 percent over a period and, attributing the adjustment cost to the new units produced, this added 20 percent to the cost of production of those new units on average. This level of ‘adjustment cost burden’ could be modelled with the function above and the parameter choice $\delta = 2$. A nearly identical argument holds for the case where $x_i$ is negative, so that the adjustment cost incurred per unit withdrawn from production, relative to marginal production cost, is therefore $\delta \left| \frac{x_i}{q_i} \right|$. The convexity of adjustment costs with respect to the proportional rate of change of output is fixed in this formulation: a doubling of the proportional rate of change doubles the ‘premium’. A two-parameter version of
this function which allowed for greater or lesser diseconomies of scale in the adjustment of output could be easily specified, although the one-parameter version here is seen as the simplest option for use in the numerical exercises of Chapter 4.

3.2.3 Adjustment speed parameter

The function $A^\delta$ has the steady state properties $A^\delta_{x_i,q_i} = 0$ and $A^\delta_{q_i,q_i} = 0$. The second partial derivative of $A^\delta$ with respect to $x_i$ is $A^\delta_{x_i,x_i}(x_i,q_i) = 2\delta c/q_i$, which at the steady state is $A^\delta_{x_i,x_i} = 2\delta c/q_i^*$. Substituting these three expressions into equation (3.1) yields:

$$\lambda = \frac{1}{2} \left( \sqrt{r^2 + \frac{2\theta}{\delta\eta} (1 + \frac{\omega}{\eta - \theta}) - r} \right)$$

(3.5)

As well as $\delta$ being easily interpretable, another useful property of $A^\delta$ is that the nominal variables in the second partial derivative ($c$ and $q_i$) cancel the nominal variables in the expression for $K$, meaning the speed of adjustment parameter is not dependent on values of variables which have arbitrary units.

3.3 Logarithmic form

Suppose the level of adjustment costs is given by the function:

$$A^\ln(x_i,q_i) = \delta c \left( |x_i| - (q_i - |x_i|) \ln \left( \frac{q_i}{q_i - |x_i|} \right) \right)$$

(3.6)

This functional form is only defined for $|x_i| \leq q_i$; thus the range of this function is restricted somewhat. The range of the function includes $|x_i| = q_i$, in which case the value of $A^\ln$ is not infinite, but is equal to $\delta c q_i$, the limit as $|x_i| \rightarrow q_i$ (the proof uses l’Hopital’s rule).

A functionally-equivalent version of $A^\ln$, which more clearly shows that the adjustment cost is a function of the proportional rate of change of output,
3.3. LOGARITHMIC FORM

is given by:

\[ A^{ln}(x_i, q_i) = \delta c q_i \left| \frac{x_i}{q_i} \right| \left( 1 - \ln \left( 1 - \left| \frac{x_i}{q_i} \right| \right) \right)^{-1} \]  (3.7)

3.3.1 Properties

The partial derivatives of \( A^{ln} \) are presented here; as with the function \( A^s \) in the previous section, these partial derivatives and their steady state values satisfy all the restrictions assumed in the previous chapter.

The first-order partial derivative with respect to \( x_i \), and its steady state value, are:

\[
A^{ln}_{x_i} = \delta c \ln \left( \frac{q_i}{q_i - |x_i|} \right) \\
A^{ln}_{x_i}(0, q_i^*) = 0
\]  (3.8)

The equivalent expressions with respect to \( q_i \) are:

\[
A^{ln}_{q_i} = \delta c \left( \frac{x_i}{q_i} - \ln \left( \frac{q_i}{q_i - |x_i|} \right) \right) \\
A^{ln}_{q_i}(0, q_i^*) = 0
\]  (3.9)

The second-order partial derivatives and their steady state values are:

\[
A^{ln}_{x_i,q_i} = -\delta c \left| \frac{x_i}{q_i} \right| \frac{1}{q_i - |x_i|} \\
A^{ln}_{x_i,q_i}(0, q_i^*) = 0
\]  (3.10)

\[
A^{ln}_{q_i,q_i} = \delta c \left| \frac{x_i}{q_i} \right|^2 \frac{1}{q_i - |x_i|} \\
A^{ln}_{q_i,q_i}(0, q_i^*) = 0
\]  (3.11)

\[
A^{ln}_{x_i,x_i} = \frac{\delta c}{q_i - |x_i|} \\
A^{ln}_{x_i,x_i}(0, q_i^*) = \frac{\delta c}{q_i^*}
\]  (3.12)

It can be seen from equations (3.8), (3.9) and (3.12) that the properties of the adjustment cost assumed in Chapter 2 are satisfied for function \( A^{ln} \).
CHAPTER 3. ADJUSTMENT COST FUNCTIONS

3.3.2 Interpretation

To interpret the adjustment cost parameter associated with the function $A^\delta$ above, it was supposed that the adjustment cost incurred during any period of time was spread evenly, or averaged, over the new units brought into production (or over the number of units previously produced for which production was cancelled, in the case of negative $x_i$). A natural way to express the size of adjustment costs is as a ‘premium’ paid over the ordinary production cost of the new units produced (or as a ‘penalty’ paid as a proportion of the cost savings from reducing the quantity produced, for negative $x_i$).

Convexity of the adjustment cost with respect to the proportional rate of change of output implies that this premium is larger with higher rates of change. Adjustment cost function $A^\delta$ is the particular function in which this premium is directly proportional to the rate of change of output, with $\delta$ the multiplicative scalar that maps the rate of change $x_i/q_i$ into the size of the premium.

A similar interpretation of parameter $\delta$ in the logarithmic form is available. However instead of mapping $x_i/q_i$ into the premium paid for new units on average, the parameter $\delta$ in $A^{\ln}$ maps $x_i/q_i$ into the premium paid for the marginal new unit produced. This premium is the incremental addition to total adjustment costs from the last new unit produced, not the average adjustment cost from all new units produced.

To see this, consider the addition to the firm’s costs from the decision to increase output by one unit for some given values of $x_i$ and $q_i$. This decision will increase the values of both $x_i$ and $q_i$ by one (since to increase output the rate of change of output must be increased), so the change in adjustment costs depends on the changes in function $A^{\ln}$ with respect to both $x_i$ and $q_i$. By this logic, the ‘marginal adjustment cost’ is given by the sum of the partial derivatives of $A$ with respect to both variables:

$$\frac{dA^{\ln}}{dq_i} = \frac{\partial A^{\ln}}{\partial x_i} \frac{dx_i}{dq_i} + \frac{\partial A^{\ln}}{\partial q_i} \frac{dq_i}{dq_i}$$

$$= \frac{\partial A^{\ln}}{\partial x_i} + \frac{\partial A^{\ln}}{\partial q_i}$$

$$= A_{x_i} + A_{q_i} \quad (3.13)$$
3.3. LOGARITHMIC FORM

Using the expressions for the partial derivatives of $A^\ln$ given above, the marginal adjustment cost expressed as a premium over the marginal production cost is:

$$\frac{dA^\ln}{dq_i}/c = \delta \frac{x_i}{q_i}$$

Thus $\delta$ is interpreted as a multiplicative scalar which maps the proportional rate of change of output into the percentage premium (over marginal production cost) paid for the marginal new unit brought into production (with a similar interpretation for negative $x_i$).

3.3.3 Derivation of $A^\ln$

The logarithmic form $A^\ln$ is in fact the only form of adjustment cost function where the marginal adjustment cost (expressed as a premium over marginal production cost) is directly proportional to the percentage rate of change of output.\(^2\) This section demonstrates this claim by presenting a derivation of $A^\ln$.

The requirement that the ratio of marginal adjustment cost to marginal production cost equates to the product of $\delta$ and $x_i/q_i$ defines a partial differential equation in the unknown function $A(x_i, q_i)$:

$$A_{x_i} + A_{q_i} = \delta c \left( \frac{x_i}{q_i} \right)$$  \hspace{1cm} (3.14)

For any function $w$ with arguments $x$ and $y$, a partial differential equation of the form:

$$w_x + w_y = f(x)g(y)$$  \hspace{1cm} (3.15)

has a general solution given by:

$$w(x, y) = \int_{x_0}^{x} f(t) g(y - x + t) \, dt + \Phi(y - x)$$  \hspace{1cm} (3.16)

where $\Phi(y - x)$ is an arbitrary function of the argument $y - x$, and $x_0$ is also arbitrary.

\(^2\)With the exception of functions that differ from $A^\ln$ only by a constant, of course.
For equation (3.14) this defines a solution:

\[ A(x_i, q_i) = \int_{s=x_0}^{s=x} \delta c s \left( \frac{1}{q_i - x_i + s} \right) ds + \Phi(q_i - x_i) \]

\[ = \delta c \int_{s=x_0}^{s=x} \frac{s}{q_i - x_i + s} ds + \Phi(q_i - x_i) \quad (3.17) \]

The integral of a function \( s/(a + s) \), where \( a \) is a constant and \( s \) is the integrating variable, is given by:

\[ \int \frac{s}{a + s} ds = s - a \ln(a + s) \quad (3.18) \]

In this case, this defines \( A(x_i, q_i) \) as:

\[ A(x_i, q_i) = \delta c \left[ s - (q_i - x_i) \ln(q_i - x_i + s) \right]_{s=x_0}^{s=x} + \Phi(q_i - x_i) \quad (3.19) \]

and choosing \( x_0 = 0 \) gives:

\[ A(x_i, q_i) = \delta c \left( x_i - (q_i - x_i) \ln \left( \frac{q_i}{q_i - x_i} \right) \right) + \Phi(q_i - x_i) \quad (3.20) \]

To find \( \Phi(q_i - x_i) \), \( x_i = 0 \) is substituted, and given that \( A(0, q_i) \) is zero by assumption, \( \Phi(q_i - x_i) \) must be equal to the zero function: \( \Phi(q_i - x_i) = 0 \) for all \( q_i, x_i \). Therefore the adjustment cost function which satisfies the property outlined above, and which is denoted \( A^{ln} \), is given by:

\[ A^{ln}(x_i, q_i) = \delta c \left( x_i - (q_i - x_i) \ln \left( \frac{q_i}{q_i - x_i} \right) \right) \quad (3.21) \]

### 3.3.4 Adjustment speed parameter

Substituting expressions (3.10) to (3.12) into equation (3.1) yields:

\[ \lambda = \frac{1}{2} \left( \sqrt{r^2 + \frac{4\theta}{\delta \eta} \left( 1 + \frac{\omega}{\eta - \theta} \right)} - r \right) \quad (3.22) \]

As with \( A^\delta \), the partial derivative of \( A^{ln} \) evaluated at steady state has the useful property of cancelling the nominal variables in the adjustment speed parameter expression.
3.4 Conclusion

Two functional forms for an adjustment cost have been suggested, which satisfy the requirements of the model in Chapter 2 and which each have a single easily-interpreted parameter that can be assigned realistic values.

The quadratic adjustment cost function expresses the burden of adjustment costs by considering a hypothetical premium over production costs, incurred when the output level is adjusted. This premium is calculated by averaging adjustment costs over the units either newly produced or withdrawn from production. The parameter in the function $A^\delta$ scales the proportional rate of change of output into the value of this premium.

A similar interpretation of $A^{\ln}$ exists; the parameter in this function maps the proportional rate of change of output into the premium on the marginal newly produced (or withdrawn) unit. That is, with function $A^{\ln}$ the ratio of the marginal adjustment cost to marginal production cost (or the premium on the marginal unit of production) is given by the product of the parameter and the proportional rate of change of output.

The parameter $\delta$ in this formulation is perhaps less representative of the impact of output adjustment on the firm’s total profits than the $\delta$ used in $A^\delta$. However it may be a better reflection of the cost of adjustment at the margin, i.e. that factor which will determine the firm’s optimal rate of output change when balancing adjustment costs against the gains from moving closer to the profit-maximising steady-state level of output.

These forms offer a useful parameterisation of the burden of adjustment costs, expressed relative to relevant variables such as the production cost and
the proportional rate of change of output. As such they are interesting in
their own right, as well as holding convenient properties for allowing further
analysis in the next chapter of the model developed in Chapter 2.
Chapter 4

Analysis: adjustment speeds and asymmetric price transmission

4.1 Introduction

The outcome of the modelling exercise in Chapter 2 was a ‘price adjustment path’: an expression describing the change in the output price over time as a function of an input price shock as well as a set of parameters describing production and adjustment costs, market demand, and the competitiveness of conduct of firms. As described in the introduction to that chapter, the motivation for this exercise was to better understand how asymmetry in the costs of adjusting output levels upwards and downwards fed into asymmetry in pricing practices. In particular, it would be useful to understand the quantitative relationship between the degree of asymmetry in adjustment costs and the extent of asymmetry in price transmission, and how this depends on other factors that may plausibly influence the speed and asymmetry of price changes, such as discount rates and market power.

In this chapter the price adjustment expression, equation (2.34) from Chapter 2, is analysed for this purpose. Section 4.2 looks at the determinants of the adjustment speed within the model and ‘sizes’ the influences on the adjustment speed from realistic variations in parameter values.

Section 4.3 looks at the patterns of asymmetry predicted by the model,
after first considering how asymmetry could best be measured. This section argues an important point that has not been raised in the existing literature, regarding the standard method for quantifying asymmetry (by calculating the difference between positive and negative responses to a simulated shock). While this approach is acceptable when testing for asymmetry in a single market, it is argued here to be inappropriate for comparisons across markets. Asymmetry measured in this way will be correlated with the long-run elasticity of price transmission, which can vary substantially across markets, irrespective of the true asymmetry of the adjustment process.

This point is particularly significant in view of the research agenda promoted by Meyer and von Cramon-Taubadel in their survey. These authors propose that theoretical work to better understand the form asymmetry could be expected to take in particular environments should be complemented by greater use of cross-sectional studies, to "attempt to exploit differences in factors that might cause asymmetric price transmission – for example market power – across products and/or countries" (2004, p605).

Section 4.3 presents several original measures of asymmetry that avoid the problems with the standard measure, and would thus be suitable for comparisons of the extent of asymmetry across different markets. Finally, the asymmetry in price responses generated by the model of Chapter 2 is examined using both the standard measure and suggested measures of asymmetry. Preliminary conclusions are drawn regarding the degree of asymmetry expected for any particular asymmetry in output adjustment costs, and regarding the question of whether aspects of market structure and conduct have any significant influence on this relationship.

4.2 Market structure and conduct, and the speed and magnitude of price transmission

Although the main focus of this article is asymmetry in transmission processes, the price adjustment expression can also be used to examine the determinants
4.2. THE SPEED AND MAGNITUDE OF TRANSMISSION

of the speed and magnitude of price transmission. The speed of adjustment is captured in the parameter $\lambda$, given by:

$$\lambda = \frac{1}{2} \left( \sqrt{r^2 + 4K} - r \right)$$

(4.1)

with $K$ given by:

$$K = \frac{1}{A_{x,q_i}} \left( r A^*_{x,q_i} + A^*_{q_i,q_i} + \frac{c}{q_i^e} \theta (1 + \frac{\omega}{\eta - \theta}) \right)$$

(4.2)

These parameters entered the price adjustment path expression derived in Chapter 2, which is reproduced here:

$$d \ln p(t) = \frac{\alpha}{1 + \frac{\omega}{\eta - \theta}} d \ln \bar{p}^A (1 - e^{-\lambda t})$$

(4.3)

The one-parameter quadratic form of adjustment cost function presented in Chapter 3 is used for the analysis here, on the basis that the interpretation of the adjustment cost parameter $\delta$ is most straightforward. As explained, $\delta$ multiplies the rate of output change into the percentage ‘premium’ over marginal production cost that a firm pays for the production of extra units (or the ‘penalty’ paid for each unit withdrawn from production), when the adjustment costs incurred are averaged over these new (or withdrawn) units of output. Use of the quadratic form effectively imposes a particular degree of diseconomies of scale in adjustment: a doubling of the rate of change doubles the premium-per-unit paid by virtue of adjustment costs. For instance, with $\delta = 2$ a 1 percent increase in output results in a 2 percent premium over marginal production cost for the new units produced, while a 10 percent increase in output results in a 20 percent premium.

While this interpretation is conceptually reasonably straightforward, it appears at first glance that the selection of realistic parameter values will still pose problems. This is because $\delta$ is essentially a ‘timeless’ parameter – it frames the size of adjustment costs as a premium over production cost without considering the period over which output and the proportional rate of change of output are to be measured. However because the convexity of the adjustment cost function is fixed, a particular choice of $\delta$ can in fact be seen to
to be consistent with the same adjustment cost ‘burden’, whatever accounting period is used to measure output, production costs, and adjustment costs.

To illustrate, suppose the length of a production period was one quarter of a year, and the quarterly cost of production was $1 million. A doubling of production from (say) the third quarter to the fourth quarter, which doubled fourth-quarter production costs to $2 million and generated an additional $1 million adjustment cost, can be seen to be represented by the same $ whether the measurement period is a quarter or a year. If the reference period were quarterly, then a doubling of output and a one-hundred percent cost premium would be represented by $\delta = 1$. If the reference period were the financial year, then both the proportional rate of change of output and the premium paid over ordinary (annual) production costs would be 25 percent; hence the same $\delta$ would apply.

While the choice of $\delta$ does not require specification of a particular measurement period, there are nevertheless no obvious empirical sources from which to select particular values for this parameter.\(^1\) Accordingly the numerical exercises that follow consider only a limited and somewhat arbitrary range of $\delta$, based on the values appearing ‘reasonable’. The faster price responses are based on $\delta = 1$, and the slower responses are based on various multiples of the lower value to a maximum of 5. To test the sensitivity of results to the magnitude of values used, section 4.3.3 examines the impact of using different lower values (the largest being $\delta = 10$). The discussion also focuses not on the absolute value of $\lambda$ (which would require specification of a time period to be meaningful), but on the direction of change of $\lambda$ with other parameters, and on the size of that change expressed in terms of the change in adjustment costs that would have the equivalent effect.

The adjustment speed under the one-parameter version of adjustment costs is given by:

\begin{equation}
\lambda = \frac{1}{2} \left( \sqrt{r^2 + \frac{2\theta}{\delta \eta} (1 + \frac{\omega}{\eta - \theta}) - r} \right)
\end{equation}

\(^1\)Issues in estimating adjustment costs and the variation in results within the existing empirical literature are summarised in Hamermesh and Pfann (1996).
4.2. **The Speed and Magnitude of Transmission**

The long-run magnitude (or elasticity) of price transmission is given by the ratio of output and input price changes, or:

\[
\tau = \frac{d \ln p^*}{d \ln \bar{p}^A} = \frac{\alpha}{1 + \frac{\omega}{\eta - \theta}}
\]  

(4.5)

The direction of influence of each parameter on the speed and magnitude of price adjustment, as described in equations (4.4) and (4.5), is shown in the table below (the partial derivatives are given in Appendix B):

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Magnitude</th>
<th>Speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta )</td>
<td>– (no effect)</td>
<td>( \downarrow ) (slower)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>( \uparrow ) (larger)</td>
<td>– (no effect)</td>
</tr>
<tr>
<td>( r )</td>
<td>– (no effect)</td>
<td>( \downarrow ) (slower)</td>
</tr>
<tr>
<td>( \theta )</td>
<td>( \downarrow ) (smaller)</td>
<td>( \uparrow ) (faster)</td>
</tr>
<tr>
<td>( \eta )</td>
<td>( \uparrow ) (larger)</td>
<td>( \downarrow ) (slower)</td>
</tr>
<tr>
<td>( \omega )</td>
<td>( \downarrow ) (smaller)</td>
<td>( \uparrow ) (faster)</td>
</tr>
</tbody>
</table>

Recall that \( \alpha \) represents the cost share of the fluctuating-price input, while \( r \) is the discount rate, \( \theta \) is the market power parameter and \( \eta \) the absolute value of the demand elasticity. Parameter \( \omega \) is the elasticity of \( \eta \) with respect to the price, so that \( \omega \) is higher for flatter demand curves (\( \omega = 0 \) for a constant-elasticity demand curve, \( \omega = 1 \) for log-linear demand, and \( \omega = 1 + \eta \) for linear demand).

The relationships between parameter values and the magnitude of price transmission match those of the model of McCorriston *et al.* (1998), on which the static input choice framework in Chapter 2 was based. Of note is that the result that market power has a negative impact on the magnitude of pass-through carries over to this model, despite the parameterisation of market power being somewhat different. In McCorriston *et al.*’s (1998) model market power was represented by a static conjectural variations parameter, a different concept to the ‘dynamic conjecture’ used in Chapter 2 which captured the firm’s (mistaken) belief about its ability to influence the market price in each period of a multi-period game.
As shown in the right-hand column of the table, a greater degree of market power also tends to increase the transmission speed. To gauge the quantitative significance of this influence, the relationship between the adjustment cost parameter $\delta$ and the adjustment speed $\lambda$ is plotted in Figure 4.1 below, for various values of $\theta$. The ‘base case’ values of other parameters are as shown.

Figure 4.1: Value of adjustment speed parameter for various levels of market power

The downward sloping lines in Figure 4.1 illustrate the negative influence of the adjustment cost parameter on the adjustment speed. The vertical spread indicates that the effect of market power on the adjustment speed is large: the difference in adjustment speeds between markets with $\theta = 0.25$ and $\theta = 0.5$ is approximately equivalent to a 50 percent reduction in adjustment costs, and the adjustment speed parameter is approximately 3 times larger in the least competitive situation than in the most competitive.
Recalling that $\theta$ represents a conjectured proportional response of industry quantity to a 1 percent change in the firm’s quantity, it follows that a monopolist (or a firm in a collusive arrangement) who expects a one-to-one relationship between these quantities will hold a conjecture of $\theta = 1$. A value of $\theta = 1/N$ represents the level of competition that holds when a player in an N-firm industry expects no dynamic response (or when an industry with fewer firms is slightly more competitive than the ‘no response’ benchmark, for example). The difference between $\theta = 0.25$ and $\theta = 0.5$, for instance, corresponds to the difference between a 4-firm and a 2-firm oligopoly where the players expect no dynamic responses from rivals and act accordingly.

That is, $\theta$ is being interpreted by reference to other forms of dynamic model rather than to actual beliefs. In particular, $\theta$ can be related to the degree of competition that arises under open-loop or time-dependent strategies. A value of $\theta = 1/N$, in this interpretation, refers to the competitiveness of outcomes of an N-firm open-loop oligopoly game; i.e. to the outcome that arises when all firms understand that all others are forced to commit to output paths as well, so that responses are impossible. Recall also that the limiting case of $\theta = 0$ cannot be sensibly accommodated in the model, and is not being considered here.

The output response of each firm in the optimisation problem of the model essentially balances two ‘costs’ the firm faces. The first is the losses from disequilibrium – from producing a sub-optimal quantity and hence facing a market price that is either too low (or too high) to maximise profits given the newly-elevated (or reduced) level of costs. The second is the adjustment cost incurred in altering that quantity to push the market price closer to the optimal level.

When the firm expects a greater response of industry quantity (and price) to its own change in output, this raises the extent to which the firm expects disequilibrium losses to be reduced by its own output adjustment. Thus for any level of adjustment costs, firms holding more ‘monopolistic’ conjectures will adjust output more quickly than firms with more competitive conjectures, explaining the relationship observed in Figure 4.1.

Another interesting result from Table 4.1 is that industries with more
elastic demand are expected to exhibit slower adjustment to equilibrium. As shown in Figure 4.2, this effect is also sizeable: the difference in adjustment speed between an industry with elastic demand ($\eta = 1.5$) and one with unit elasticity ($\eta = 1$) is approximately equivalent to a doubling of adjustment costs. The difference between highly inelastic ($\eta = 0.5$) and inelastic ($\eta = 0.75$) industries is similar.

Figure 4.2: Value of adjustment speed parameter for various demand elasticities

![Graph showing adjustment speed parameter for different demand elasticities](image)

The other parameters have a less pronounced influence on the adjustment speed. Higher discount rates $r$ (corresponding to heavier discounting of future profits) cause slower adjustment. Intuitively, heavier discounting implies that postponing adjustment cost expenditure is a rational response, with consequently slower adjustment. The curvature of the demand curve is also related to the adjustment speed, with faster adjustment when demand is
4.2. **THE SPEED AND MAGNITUDE OF TRANSMISSION**

closer to linear.

Table 4.2 summarises the size of the influence on the price adjustment speed of changes in these four parameters away from the ‘base case’ values of $\delta = 1$, $\theta = 0.25$, $\eta = 0.75$, $r = 0.1$, and $\omega = 1+\eta$. The right-hand side column shows the relative size of these effects using a common metric: the change in the adjustment cost parameter that would induce an equivalent change in the adjustment speed to the variation in the parameter in question.

| Parameter variation | Adjustment speed ($|\lambda|$) | Percent change from base case | Equivalent percent change in $\delta$ |
|---------------------|-------------------------------|------------------------------|--------------------------------------|
| Base case           | 0.51                          |                              |                                      |
| Market power        |                               |                              |                                      |
| $\theta = 0.1$      | 0.25                          | -52                          | +269                                 |
| $\theta = 0.5$      | 1.18                          | +130                         | -79                                  |
| Demand elasticity   |                               |                              |                                      |
| $\eta = 0.5$        | 0.74                          | +45                          | -50                                  |
| $\eta = 1$          | 0.41                          | -20                          | +50                                  |
| $\eta = 1.5$        | 0.31                          | -40                          | +150                                 |
| Discount rate       |                               |                              |                                      |
| $r = 0.01$          | 0.55                          | +8                           | -14                                  |
| $r = 0.5$           | 0.36                          | -29                          | +87                                  |
| Demand curve        |                               |                              |                                      |
| $\omega = 0$ (C.E.D) | 0.36                         | -29                          | +88                                  |
| $\omega = 1$ (L.L)  | 0.45                          | -11                          | +25                                  |

In considering the range of variation in the discount rate $r$, a similar issue arises to that discussed in selecting values for $\delta$. The choice of a realistic value for $r$ depends on the length of the reference period: a 10 percent value may be a suitable borrowing rate or internal rate of return on investment over a year-long horizon, but if the measurement period is shorter then a lower value should be used. However since $\lambda$ itself is also without meaning in the absence of a specified measurement period, the particular choices of $r$ are not greatly relevant; what is of interest is the proportional change in $\lambda$ (or the change in $\delta$ that would have an equivalent effect), as $r$ is changed
by a certain proportion. Thus Table 4.2 shows that (for instance) a 500 percent increase in $r$ – whatever the reference period is – slows adjustment by around one-third, which is similar to the effect that an approximate doubling of adjustment costs would have. Similarly, in considering below how the degree of asymmetry varies with parameter values below, it is the relationship between proportional changes in the discount rate and proportional changes in measured asymmetry that is of interest in understanding the quantitative influence of the discount rate.

4.3 Asymmetric price transmission

4.3.1 Introduction

This section examines the patterns of asymmetric adjustment cost-induced pricing asymmetry, and the supporting roles of other variables, as reflected in the price adjustment path of Chapter 2. The purpose is to provide estimates of what degree of asymmetry in price transmission would be consistent with a certain asymmetry in adjustment costs (at least for the type of market modelled, a homogeneous-good quantity-setting oligopoly). Ideally, estimates such as these might allow empirical researchers to conclude that asymmetry observed in a particular market was either consistent with, or in excess of, that which might be expected given realistic estimates of asymmetry in adjustment costs.

The basic method is to use equations (4.3) and (4.4) to generate price responses to 1 percent shocks in the agricultural input price, and then to observe the degree of asymmetry in price responses when adjustment costs are asymmetric. To generate asymmetry, a higher value of $\delta$ is assumed to apply to output expansions, which in this model follow reductions in input costs, than to output contractions which follow increases in input costs. The parameters $\delta^+$ and $\delta^-$ refer to adjustment costs applying to positive and negative input price changes respectively, not those applying to positive and negative output changes. A ratio $\delta^+/\delta^- < 1$ is assumed, which corresponds to positive asymmetry (faster pass-through of positive cost shocks).
4.3. ASYMMETRIC PRICE TRANSMISSION

Figure 4.3 plots a single adjustment path in response to a positive shock (this is labelled $d\ln p^+(t)$ and shown by a dashed line) and two paths representing responses to negative shocks (the downstream price changes are negative so the absolute values $|d\ln p^-(t)|$ are shown, as dotted lines). The long-run price changes are equivalent, because the adjustment cost parameter has no influence on the magnitude of price transmission. Also shown is a basic indicator of the degree of asymmetry in price responses: the difference between positive and negative price changes at each time $t$ (calculated as $|d\ln p^+(t)| - |d\ln p^-(t)|$ and shown as a solid line).

Figure 4.3: Downstream price responses and asymmetry in price responses

The discrepancy between positive and negative output price responses to identically sized shocks is obviously largest when adjustment costs are the most asymmetric (when $\delta^- = 5\delta^+$). To examine in more depth how
the adjustment cost ratio $\delta^+ / \delta^-$ determines the degree of asymmetry and how the other parameters influence this relationship a range of asymmetry measures are introduced. Before exploring in section 4.3.3 in more detail how this discrepancy varies with adjustment cost asymmetry (and other market characteristics), the important question of how best to quantify the extent of this asymmetry is considered in section 4.3.2.

### 4.3.2 Asymmetry measures

#### Standard measure of asymmetry

The most straightforward representation of asymmetry in price adjustment speeds generated by the model of Chapter 2 is the measure $\lambda^+ / \lambda^-$, the ratio of the adjustment speed parameters. In the context of this model $\lambda$ represents the amount by which the price adjusts in a period as a (constant) proportion of the difference between the current price and the steady-state price; this is referred to as a ‘partial adjustment’ process, and the ratio of adjustment parameters for positive and negative price changes is a useful summary measure of asymmetry.

However this representation is model-specific, with no reason to expect adjustment paths to generally match the type of ‘partial adjustment process’ that emerged from this model. Estimating $\lambda$ from data effectively involves ‘shoehorning’ the actual adjustment process into a particular pattern, when ideally asymmetry should be measured without imposing restrictions on the data.

While the linear error correction models commonly used in econometric studies of asymmetric adjustment have a partial adjustment term, they typically allow output prices to respond directly to input price shocks as well as through the error correction term (see Meyer and von Cramon-Taubadel’s 2004 survey). This means the overall adjustment process does not strictly follow a partial adjustment process, and while the ratio of error correction terms is an acceptable representation of the relative speeds of disequilibrium adjustment it is not the correct measure of the overall asymmetry in adjustment paths (Johnson 2002, p39 discusses this point).
A more common representation of asymmetric response is to generate cumulative downstream price responses to identically-sized positive and negative upstream shocks, and to plot or calculate the difference between these with appropriate confidence intervals (see Borenstein et al. 1997, Balke et al. 1998, and Peltzman 2000 for examples). The representation of asymmetry in Figure 4.3 is an example of the usual form of presentation of empirical results. The size and statistical significance of the difference between positive and negative responses at various time horizons indicates whether asymmetry is detectable in the immediate impact of the shock, in the long-run effects, or in both. Some studies have also conducted basic welfare analysis based on this measure, by calculating how much more consumers pay by virtue of asymmetry over the adjustment period (see the studies by Borenstein et al. 1997 and Douglas 2010 on gasoline price adjustment).

This standard ‘difference measure’ of asymmetry at time $t$ following the shock is given by:

$$D(t) = |d \ln p^+(t)| - |d \ln p^-(t)|$$  \hspace{1cm} (4.6)

It potentially ranges between 0 for adjustment which is symmetric at all horizons, and $|d \ln p^+|$ when upwards adjustment is complete before downwards adjustment has begun. The measure labelled here as $D(t)$ is equivalent to the asymmetry measure defined by Potter (1995), as the difference between non-linear impulse response functions.

There has been little discussion in the existing empirical literature about whether there might be a more suitable summary measure of asymmetry than $D(t)$. Meyer and von Cramon-Taubadel’s survey, for example, while promoting the idea of cross-sectional regressions to uncover correlations between asymmetry and potential causal factors (2004, p605) never discusses how the extent of asymmetry might be meaningfully compared across markets. A survey of econometric methods by Frey and Manera (2007) distinguishes eight different categories of asymmetry and discusses how the presence of each might be tested, without considering how the overall asymmetry of the adjustment processes might be summarised and quantified.

This is understandable, since the majority of studies focus on a single
market or supply chain, and the test of whether \( D(t) \) is statistically different from 0 is the most straightforward way of showing asymmetry. As argued below, however, this ‘difference measure’ might not be the most accurate representation of the true asymmetry of adjustment processes since it will depend greatly on the long-run extent of price pass-through, which naturally varies across markets. As such any search for factors correlated with asymmetry will also likely be unsuccessful when based around this measure.

**Suggested alternative measures**

This subsection presents a number of alternative methods of representing asymmetry in adjustment processes via a single measure, and it is suggested they could be usefully employed in empirical work. The discussion here is prompted by the potential deficiencies of the ‘difference measure’ described above, which is standardly used when testing for asymmetry in a single market. There appears to have been little consideration in the literature of alternative (or the most appropriate) methods for measuring the degree of asymmetry, and the summary measures suggested below (while intuitive) have not featured in existing work to date.

A summary measure of the degree of asymmetry is not necessarily helpful in gaining an in-depth understanding of price responses in a particular market, since price adjustments may be asymmetric in several respects: in their timing, in their (long-run) amount, or in the pattern of adjustment to long-run equilibrium.\(^2\) To take one simple example, Karrenbrock (1991) finds symmetry in retail gasoline prices in the number of lag months over which positive and negative oil price changes are passed through, as well as in the long-run magnitude of pass-through, yet also finds that the pattern of adjustments in the initial and lagged months differ for positive and negative price changes; a summary measure of asymmetry would obviously collapse this detail down.

Yet a single measure may be useful to compare industries (e.g. through

\(^2\)This is the simplest characterisation of types of asymmetry. Frey and Manera’s (2007) survey of empirical studies and techniques identifies a list of eight categories of asymmetry that are potentially identifiable from econometric results.
cross-sectional regressions) where asymmetry manifests in different respects. And the welfare consequences of asymmetry are clearly related to both the relative lengths of time for complete upwards and downwards adjustment, and the size and timing of intermediate adjustments; to the extent a summary measure expresses asymmetry in terms relevant for welfare analysis, it may be more useful than a description of asymmetry along multiple dimensions.

Two complications arise when using the sort of ‘difference measure’ outlined above to compare the asymmetry present in different markets, or to search for potential causes of asymmetry by regressing the measured degree of asymmetry on various explanatory variables (as in Peltzman 2000). The first (minor) complication is the selection of an appropriate lag length at which to use the difference between upwards and downwards pass-through as the measure of asymmetry, given that the rate of pass-through might differ by industry. Using the wrong ‘snapshot’ of upwards and downwards price adjustments might underestimate the true extent of asymmetry for some markets, particularly if the cross-section contains a wide variety of industries with potentially quite different adjustment horizons (compare the ease of altering retail petrol prices with the lags in increasing agricultural production, for instance).

An obvious solution to this problem is to use a summary ‘difference measure’ of asymmetry, such as the maximum difference between positive and negative price changes over the period of adjustment to the shocks:

\[ D = \max_t \left\{ \left| d \ln p^+(t) \right| - \left| d \ln p^-(t) \right| \right\} \quad (4.7) \]

However a second (and more significant) problem with these difference measures is that long-run magnitudes of price transmission vary according to different characteristics of the industry, such as the cost share of the input which changed price, the elasticity of demand, the substitutability between inputs in production, and a range of other factors including the competitiveness of the industry. The last of these is clearly seen from basic models of price determination in monopolistic, oligopolistic and competitive markets: the more competitive the market, the greater the proportion of a cost shock that is passed through. The combination of all such industry charac-
teristics in determining the price transmission elasticity is the subject of the static price transmission literature (see McCorriston et al. 1998 and 2001 in particular).

This means a price transmission process which is mildly asymmetric but with a large magnitude of transmission (reflecting a high input cost share, say) will score highly on a non-proportional measure such as \( D(t) \) or \( D \), but if the price transmission magnitude is small the value of \( D \) will be low no matter how dissimilar are the positive and negative adjustment paths.

Any measure of differences between upwards and downwards price movements which does not correct for these factors is thus at risk of losing the true inter-industry variation in the asymmetry of adjustment processes within the ‘noise’ of inter-industry variation in long-run elasticities of price transmission. For instance, in neither of the only two studies to compare asymmetry across multiple industries – Peltzman (2000) and Gwin (2007), which largely replicated Peltzman’s methodology – was the difference measure of asymmetry adjusted to take into account inter-industry differences in long-run pass-through (although Peltzman’s regressions in search for correlates of asymmetry at least included the input cost share as an explanatory variable).

Yet the difference between input cost shares of, say, 0.3 and 0.5 (the inter-quartile range in Peltzman’s producer sample) could be expected to exert significant influence on the absolute difference between positive and negative price responses, in a way entirely unrelated to the true asymmetry of price responses.

It is suggested here that one straightforward way of overcoming this problem is to use a proportional version of the summary difference measure introduced above:

\[
PD = \max_t \frac{|d\ln p^+(t)| - |d\ln p^-(t)|}{|d\ln p^*|} \quad (4.8)
\]

The ‘proportional difference’ measure \( PD \) ranges between 0 for upwards and downwards movements which are identical in timing, and 1 when upwards adjustment is complete before downwards adjustment has begun. An advantage of \( PD \) is that it is invariant in the magnitude of price transmission \( |d\ln p^*| \) while \( D(t) \) and \( D \) are influenced by this magnitude, making \( PD \)
more suitable for comparisons across markets.

There is a further reason why the choice of measure may be important for understanding the causes of asymmetry. As shown in McCorriston et al. (1998), as well as in Chapter 2, the degree of market power exerts a sizeable negative influence on the price transmission elasticity. Market power is of particular interest as a cause of asymmetry, both because (local or temporary) market power features in several theoretical explanations centering on consumer search or tacit collusion, and because market power-driven asymmetry implies net welfare losses relative to the symmetric adjustment situation. If a non-proportional metric such as \( D \) or \( D(t) \) is used to measure asymmetry, mildly asymmetric pricing practices in competitive markets where the price transmission elasticity is close to \( \alpha \) (the input cost share) may nonetheless result in these markets being mistakenly classed as ‘highly asymmetric’. Equally, highly asymmetric pricing behaviour in a non-competitive market with a long-run price transmission elasticity much below \( \alpha \) may appear insignificant. If less competitive markets are more likely to display asymmetric price transmission, as is commonly suggested, but the measure used to gauge the extent of asymmetry downplays asymmetric adjustment when only a small proportion of the cost shock is passed through (as in uncompetitive markets), then attempts to prove a general correlation between market power and asymmetry across multiple industries will be frustrated. This may explain why Peltzman (2000) was unable to find any general correlation between industry concentration and asymmetry.

A disadvantage associated with both the difference and proportional difference measures above is that they are insensitive to the relative lengths of positive and negative adjustment periods, an aspect of asymmetry which is highly relevant to understanding welfare consequences. For example, consider two processes: one consisting of immediate full adjustment upwards with a 1 period lag before full adjustment downwards, and a similar process with a 2 period downwards lag. These would appear equally asymmetric as measured by \( D \) and \( PD \) although the speed of adjustment is clearly more asymmetric in the latter case; the welfare and distributional impacts would be correspondingly more significant in the latter case too.
Two further measures are suggested here which avoid this problem. The first measures asymmetry according to a ratio of ‘half-lives’:

$$HLR = \frac{t_H^+}{t_H^-} \quad (4.9)$$

In this measure $t_H^+$ and $t_H^-$ represent the times at which half of the upwards and downwards adjustment (respectively) is complete. A value for $HLR$ greater (less) than one represents positive (negative) asymmetry, and $HLR = 2$ for example can be interpreted as indicating that downwards price adjustment takes twice as long as upwards adjustment. A similar measure with a proportion other than 50 per cent (of adjustment) might be more appropriate when calculating the ratio of price adjustment lags for positive and negative shocks from empirical studies if, for example, the market being studied tends to exhibit quite substantial (i.e. more than 50 per cent) immediate pass-through of both positive and negative shocks but with asymmetric adjustment after that.

A second alternative utilises the area under the two adjustment paths, measured from the time of the shock to the point where upwards and downwards adjustment are both complete. The ratio of these two areas, $AR$, will be greater (less) than one when there is positive (negative) asymmetry. This ‘area ratio’ measure is given by:

$$AR = \frac{\int_{t=0}^{t=t^*} |d \ln p^+(t)| \, dt}{\int_{t=0}^{t=t^*} |d \ln p^-(t)| \, dt} \quad (4.10)$$

The time $t^*$ is the length of the longer of the two adjustment periods (i.e. the time taken for a negative cost shock to be fully passed through, in the case of ‘positive’ asymmetry). While empirical studies can use the number of statistically significant lag lengths as an indicator of the adjustment time, for the purposes of calculating $AR$ in the next subsection the adjustment length was approximated by the time taken for the 90 percent of the long-run price change to have occurred.
4.3. ASYMMETRIC PRICE TRANSMISSION

The measure $AR$ also has an interpretation in terms of industry revenue. It corresponds approximately to the ratio of the percentage change in industry revenue caused by the positive shock (relative to the baseline of no shock) to the percentage change in industry revenue caused by the negative shock, as measured over the length of the adjustment period.

To see the relationship between $AR$ and the ratio of revenue changes, note that the percentage change in revenue $R$ (where $R = pQ$) is the sum of percentage changes in price and quantity sold, or $d \ln R = d \ln p + d \ln Q = d \ln p(1 - \eta)$. The ratio of percentage changes in revenue received over the adjustment period $\frac{\int d\ln R^+(t) \, dt}{\int d\ln R^-(t) \, dt}$ therefore simplifies via the cancellation of $1 - \eta$ terms to the ratio of percentage changes in price, which is what $AR$ measures. The approximation lies in assuming the elasticity of demand is constant before and after the shocks.

The half-life and area ratio measures have the desirable properties of a higher asymmetry score when the lag between upwards and downwards adjustment is greater, and of being unaffected by changes in the price transmission magnitude. Although the standard difference measure seems to be in use in econometric work and is appropriate for detecting asymmetry in a single industry, the proportional difference, half-life and area measures may capture the asymmetry of the adjustment process more accurately and thus be more useful when searching for determinants or correlates of asymmetry over a sample of industries.

The four measures $D$, $PD$, $HLR$, and $AR$ are used below to examine the relationships between the parameters of the model and the degree of asymmetry generated by asymmetric adjustment costs.

Since the adjustment process as specified by equation (4.3) is smooth – responses to both positive and negative shocks are gradual and differ only in their speed, not their nature – the various measures give similar results. The presentation of multiple measures is aimed less at exploring the properties of the particular adjustment process generated by the model, and more at highlighting the possibilities for measurement of asymmetry where behaviour is not so straightforward, such as ‘rockets and feathers’-type situations where upwards adjustment is lumpy and downwards adjustment gradual.
4.3.3 Model predictions

Figures 4.4 through 4.7 below show how the asymmetry generated by the model varies according to the asymmetry in adjustment cost parameters, as shown on the horizontal axes. Each of the figures varies one of the other parameters of the model from its base case value, in order to see how sensitive the asymmetric adjustment cost – asymmetric price transmission relationship is to these characteristics of the market.

To illustrate the interpretation of the graphs, consider for instance the ‘proportional difference’ (\(PD\)) panel in Figure 4.4. The solid line shows the prediction of the model for the degree of asymmetry using the \(PD\) measure when the adjustment cost parameter applying to reductions in output is equal to 1, and when the adjustment cost parameter applying to expansion of output varies between 1 and 5. Recall a parameter value of 1 in the quadratic adjustment cost function introduced in Chapter 3 implies that a 10 per cent increase in output incurs adjustment costs equivalent to a 10 per cent premium over marginal production cost for the additional units produced (a 100 per cent increase results in an effective 100 per cent premium, etc.). With an adjustment cost ratio of 2 – meaning the adjustment cost for expansion of output is twice as large as for contraction of output – the model predicts the maximum difference between positive and negative price responses will be around 15 per cent of the long-run price response (\(PD \approx 0.15\)). When the adjustment cost ratio is 5, the maximum difference will be around 30 per cent of the long-run price transmission elasticity.

The ‘area ratio’ measure similarly tells us that the percentage change in industry revenue over the entire adjustment period will be around 17 per cent higher for positive cost shocks than for negative cost shocks (in absolute terms) when adjustment costs for the latter are twice as large, and about 35 per cent higher when adjustment costs are five times as large. As measured by the ‘half-life ratio’, adjustment to negative cost shocks will take approximately 40 per cent longer when the adjustment cost ratio equals 2, and 250 per cent longer when the adjustment cost ratio equals 5.

Thus a basic finding, evident from any of the graphs, is that the degree
of asymmetry in price adjustment is not proportional to the degree of asymmetry in adjustment costs. A doubling of the adjustment cost ratio results in a less-than doubling of the degree of asymmetry in price transmission.

It is unclear \textit{a priori} whether the absolute size of adjustment costs is related to the extent of adjustment cost asymmetry. This relationship is investigated by plotting the degree of asymmetry against the adjustment cost ratio $\delta^-/\delta^+$ for various adjustment cost sizes, in Figure 4.4. All measures indicate a small positive effect of adjustment cost size on asymmetry, although without clearer evidence of what parameter values would make the adjustment cost representation introduced in Chapter 3 a good approximation to reality the significance of this effect is not obvious.

Although the previous section showed that the demand elasticity had a large impact upon the adjustment speed, Figure 4.5 demonstrates that it has little influence on the degree of asymmetry of the adjustment process. The spread between the inelastic ($\eta = 0.5$) and elastic demand ($\eta = 1.5$) cases is small for all measures, with the exception of $D$ where the measure is largely driven by the negative relationship between the demand elasticity and the price transmission magnitude.

Given the theoretical focus on market power as a cause of asymmetry, an important question is whether this variable also has a role in either increasing or decreasing the pricing asymmetry resulting from asymmetric adjustment costs. Figure 4.6 shows only a negligible influence of the firm’s conjecture on the asymmetry of response to positive and negative shocks, according to $PD$, $AR$, and $HLR$. The negative effect is largest for measure $D$ but, as explained above, this is a result of the influence of market power on the magnitude of price transmission.

Plots of price transmission asymmetry for varying $\alpha$ and $\omega$ are not shown; only the difference measure $D$ varies with $\alpha$, reflecting the influence of this parameter on the magnitude of price transmission. The curvature of demand $\omega$ also appears to exert little difference on the degree of asymmetry and so the plots are not shown.

The insignificant influence of these parameters on the asymmetry predicted by the model is itself an interesting finding. It indicates that the
Figure 4.4: Asymmetry in price transmission for various adjustment cost levels

**Horizontal axis (all graphs):** Adjustment cost ratio \( \left( \frac{\delta^-}{\delta^+} \right) \)

- \( \delta^- = 1 \)
- \( \delta^- = 5 \)
- \( \delta^- = 10 \)

**Parameters used:**

- \( \alpha = 0.5 \)
- \( \theta = 0.25 \)
- \( \eta = 0.75 \)
- \( \omega = 1 + \eta \)
- \( r = 0.1 \)
Figure 4.5: Asymmetry in price transmission for various demand elasticities

Horizontal axis (all graphs): Adjustment cost ratio \( \frac{\delta^-}{\delta^+} \)

- \( \eta = 0.5 \)
- \( \eta = 0.75 \)
- \( \eta = 1 \)
- \( \eta = 1.5 \)

Symbols:
- \( \alpha = 0.5 \)
- \( \theta = 0.25 \)
- \( \omega = 1 + \eta \)
- \( r = 0.1 \)
- \( \delta^+ = 1 \)
Figure 4.6: Asymmetry in price transmission for various market power levels
4.3. ASYMMETRIC PRICE TRANSMISSION

primary driver of the degree of asymmetry when adjustment costs are asymmetric is the adjustment cost asymmetry itself, with very little ‘noise’ introduced to this relationship by other aspects of the industry (at least by those that feature within the model setup of Chapter 2). If a straightforward relationship between adjustment cost asymmetry and price transmission asymmetry can be found to be a common theme of other models (i.e. to be more than an artefact of the modelling assumptions here), it may help applied researchers searching for causes of asymmetry observed in a particular market to assess whether the extent of this asymmetry is consistent with an asymmetric-costs-of-adjustment explanation. In this case the only information required to make such an assessment would be evidence of the asymmetry of output adjustment costs; information about the absolute size of these costs and about other variables relating to the structure and conduct of the market would appear not to be greatly relevant.

The first panels of Figures 4.5 and 4.6 also highlight the point raised in the previous subsection, that factors which influence the elasticity of price transmission but are unrelated to the asymmetry of adjustment – namely the demand elasticity and industry competitiveness (and obviously the input cost share) – will bias the ‘difference’ measure of asymmetry, making it unreliable as an indicator of the true asymmetry of the adjustment process.

One interesting finding is that the discount rate \( r \) appears to be important in determining the degree of asymmetry, at least in the model under examination. In Figure 4.7 even the difference measure \( D \) is larger for higher discount rates, despite the price transmission magnitude being invariant in \( r \). The other measures also record asymmetry around 50 per cent larger in the highest discount rate situation than in the low discount rate case.

It is intuitive that both higher adjustment costs and a higher discount rate would lead firms to delay output adjustments. However the finding that asymmetry is also more severe under a higher discount rate points to some multiplicative effect within the model, whereby the effect of higher adjustment costs (of output expansion) on the adjustment speed is more pronounced when the discount rate is higher.

This prediction suggests that rapidly-changing or high-risk industries
which face high borrowing rates, or where firms respond to uncertainty about future prospects by heavily weighting present-day profits in their decision-making, would (*ceteris paribus*) tend to exhibit greater asymmetry in adjustment to cost shocks. Bearing in mind the limitations of this particular model, and uncertainty about what range of discount rates (and adjustment cost asymmetries) would be empirically relevant, this finding should perhaps not be overly stressed. It points towards an interesting topic for further modelling, however.

Figure 4.7: Asymmetry in price transmission for various discount rates

![Graphs showing asymmetry in price transmission for various discount rates.](image)

Horizontal axis (all graphs): *Adjustment cost ratio* ($\delta^-/\delta^+$)

- $r = 0.01$
- $r = 0.1$
- $r = 0.5$

- $\alpha = 0.5$
- $\theta = 0.25$
- $\eta = 0.75$
- $\omega = 1 + \eta$
- $\delta^+ = 1$
4.4 Conclusion

The first point standing out from the discussion above is that a simple measure of the difference between positive and negative responses is not a reliable indicator of the underlying asymmetry of price adjustment. Long-run magnitudes of price transmission can vary greatly across markets (depending on input cost shares, market power, and the demand elasticity), and a difference measure effectively suppresses or magnifies the true asymmetry of pricing practices by this magnitude. Use of a difference measure would thus make it more difficult to find patterns in the incidence and extent of asymmetry. The difference between positive and negative responses is useful for testing for asymmetry in a single market, but measures such as the ‘proportional difference’ ($PD$), ‘area ratio’ ($AR$), and ‘half-life ratio’ ($HLR$) provide a more ‘universal’ measure of the degree of this asymmetry.

There has apparently been no recognition of this point in the existing asymmetric price transmission literature. Since most studies have considered only a single market this is to be expected. However exercises such as Peltzman’s (2000) search for correlates of asymmetry over a range of markets (and indeed any cross-market comparison at all) will – according to the arguments presented here – be compromised by the use of a ‘difference’ measure, which is highly correlated with the long-run elasticity of price transmission. In fact, the implication of this argument is even stronger for research attempting to prove a general tendency towards greater asymmetry in less competitive markets (as much speculation and some theory suggests). Since less competitive markets will likely exhibit less long-run pass-through of cost shocks, the difference between positive and negative price responses will appear relatively small regardless of the underlying asymmetry of adjustment processes, thwarting any attempt to find a general correlation. The major contribution of this chapter has been to lay out these points, and to offer several alternative summary measures of asymmetry which avoid this problem and would be suitable for empirical work.

Given the acknowledged limitations of the model of Chapter 2, the predictions of this model regarding the determinants of the speed and asymmetry
of price transmission are regarded as a somewhat lesser contribution of this chapter. Nonetheless the model offers several interesting findings which, were they to prove robust to various model specifications, could be useful for applied research into the drivers of asymmetry in particular markets.

One basic finding is about the predicted degree of asymmetry for any given asymmetry in adjustment costs. When output expansion costs are twice as large as for output contraction, asymmetry is approximately 15-20 per cent under the $PD$ and $AR$ measures, and around 50 per cent under the $HLR$ measure. With highly asymmetric adjustment costs (a ratio of 5), asymmetry under $PD$ and $AR$ is in the realm of 30-40 per cent, and under $HLR$ is around 200-300 per cent. The degree of price adjustment asymmetry is thus not proportional to the asymmetry in adjustment costs.

A second finding is that these figures are consistent with a large range of parameter values; that is, the characteristics of the market and firm seem to make little difference to the basic relationship between adjustment cost-asymmetry and price response-asymmetry. The one exception is the discount rate, the only parameter that seemed to exert any substantial influence on the degree of pricing asymmetry in this model. Higher discount rates resulted in more significant asymmetry for any given asymmetry in adjustment costs.

Regarding the speed of adjustment, both the elasticity of demand and the degree of market power were shown to have a substantial influence. Less competitive markets are found to exhibit faster adjustment to shocks, and more elastic demand results in slower adjustment; the difference between demand elasticities of $-1$ and $-1.5$, for instance, is equivalent to an approximate doubling of adjustment costs.

The broad direction of enquiry regarding asymmetric price transmission is to better integrate theoretical and empirical work so as to improve understanding of both the general causes of asymmetry, and how the drivers of asymmetry in particular markets may be pinpointed. Most important in a policy sense is distinguishing between pricing practices which reflect real adjustment costs, and those which constitute a welfare-diminishing exercise of market power. Economic theory can contribute to these goals by examining the patterns of asymmetry that result from particular types of interactions.
between firms, so as to draw out a set of ‘stylised facts’ about asymmetry which can form the basis of tests of the underlying causes operating in any given market.

The modelling and analysis in Chapters 2 to 4 should be considered in this context as an attempt to explore the implications of asymmetry in convex adjustment costs (such as the costs of ‘hiring and firing’ inputs). The behaviour modelled was quite specific: that of quantity-setting producers of a homogeneous good, operating under a simple and intuitive (but false) understanding of their strategic environment which captured their beliefs about ‘price-making’ power in a single variable. Despite the artificiality of this modelling environment, the overall approach employed in these chapters may find useful application elsewhere. In particular, the novel modelling approach designed to allow the adjustment process to be described in terms of quantifiable elasticity parameters could be explored further.

The material in Chapter 3 was complementary to this approach. The contribution of that chapter was to present two original adjustment cost functions which offer a useful and intuitive way of describing the ‘burden’ of adjustment costs through easily-understood parameters, themselves a form of elasticity. These parameters map the proportional rate of change of output into the ‘premium’ paid over marginal production cost (for either for the average or marginal new unit produced), a formulation that could potentially find application in other theoretical models or in empirical work.
Part II

Market power and the speed of price adjustment
Chapter 5

Conjectural variations, market power and the speed of price adjustment

5.1 Introduction

A long-running line of enquiry in industrial economics concerns the relationships between concentration, market power, and the speed at which markets adjust to shocks. The general motivation for examining this question is to better understand how microeconomic factors, such as the competitiveness of markets and the nature of the strategic interactions between firms, influence the processes underlying inflation.

Beginning with Domberger (1979), a number of papers in the 1980s and 1990s tested the link between concentration and the speed of adjustment, particularly amongst manufacturing industries (Dixon 1983; Bedrossian and Moschos 1988), and in banking (Hannan and Berger 1991; Neumark and Sharpe 1992; Jackson 1997). Results from these studies were contradictory, with some authors finding that more concentrated industries exhibited faster adjustment (Domberger 1979; Hannan and Berger 1991), and others finding that fewness of firms or high concentration was associated with greater price rigidity (Dixon 1983; Bedrossian and Moschos 1988).

Some results were more nuanced: Jackson (1997), for instance, found non-monotonicity in the concentration – adjustment speed relationship (with
least and most concentrated markets adjusting fastest). The banking studies of Hannan and Berger (1991) and Neumark and Sharpe (1992) also tested for asymmetry in the speed of adjustment of deposit rates to movements in interbank lending rates, with both studies showing the presence of asymmetry in more concentrated markets.

During this period there existed (and still exists) little theoretical basis for expecting a strong relationship between concentration and the speed of adjustment in one direction or the other, with several conflicting hypotheses in existence. One idea, building on an argument of Stigler (1964), was that because the effect of one firm failing to pass on a cost increase would be more strongly felt by other firms in highly concentrated industries, these industries would display more rapid and ‘orderly’ pass-through of cost shocks as each firm sought to avoid being observed as out of line with others. In this theory, fewness of firms is associated with less costly observation and communication (and easier detection of cheating), and these same factors that sustain uncompetitive profits also cause price adjustment to move quickly and in parallel fashion.

Domberger (1983) offered a second reason, relying on the profitability of concentrated industries, for expecting this outcome. In an environment of uncertainty about the size of rivals’ price adjustments, one ‘adjustment cost’ is the risk of a price increase exceeding those of other firms and resulting in lost sales. A firm with a significant profit margin to act as a buffer perceives this risk as lower (less costly), and is thus more likely to act as a price leader; higher profitability consequently results in faster adjustment.¹ However as argued by Bedrossian and Moschos (1988), this ‘leadership effect’ may exist only when high concentration results from the dominance of a single firm, and may be subordinate to an ‘industry profitability effect’ when the market is more evenly split. In the logic of the latter, the profitability of oligopolists in more concentrated markets facilitates competition between them as they seek to use delays in raising prices as a tactic to gain market share, while the thin margins in less concentrated markets force quick adjustment of prices.

¹The argument connecting firm profitability and price leadership is originally from Holthausen (1978).
when costs change. Increased concentration, the argument goes, can result in slower upwards adjustment of prices when there is no dominant firm.\footnote{According to this argument, which seems to be similar to a ‘kinked demand curve’ story, the ‘industry profitability effect’ would only result in slower upwards adjustment of prices amongst more profitable firms. Downwards adjustment would actually occur more rapidly, as firms use price cuts as a competitive tactic; that is, price adjustment would be asymmetric in profitable (more concentrated) industries and symmetric in unprofitable (less concentrated) industries (Bedrossian and Moschos 1988, p462-3). Note, however, that this argument seems to be counterintuitively premised on more intense competition existing amongst firms in more profitable industries. The direction of asymmetry predicted is also at odds with that found by Hannan and Berger (1991) and Neumark and Sharpe (1992) for consumer deposits, and by other authors across a variety of product markets.}

Against the idea that concentration ought to result in faster price adjustment, Ginsburgh and Michel (1988) note several aspects of oligopolistic pricing practices that may support the opposite conclusion: the association of concentrated industries with large irreversible investments which encourage firms to peg prices on long-run objectives, rather than on short-run cost fluctuations; and the tendency for oligopolistic prices to change in discrete steps, while prices in more competitive markets tend to vary continuously with costs.

Alongside this collection of hypotheses, another strand of the literature has explored the influence of concentration and market power on adjustment speeds via dynamic models. These models are extensions of Cournot and Bertrand where each firm is forced to balance the adjustment costs of altering its quantity or price against the losses from choosing a quantity or price that is sub-optimal; from the first-order conditions of this optimisation problem the speed of adjustment can be derived as a function of the modelling parameters.

In Ginsburgh and Michel’s (1988) model, firms play a single-shot Cournot game in each period of a multi-period game, with previous-period output entering the firm’s decision making via an adjustment cost function. The industry output in each period can be expressed as a linear function of previous period output, with the coefficient (between 0 and 1) representing the sluggishness of adjustment. Depending on the convexity of adjustment costs, a number of scenarios arise for the relationship between the number of firms and price inertia, including non-monotonicity (faster adjustment in the least and most concentrated markets).
Two subsequent papers by Worthington (1989) and Martin (1993) provide the first attempts to incorporate market power – via a ‘conjectural variation’ parameter – into this type of dynamic model, in order to disentangle the potentially differing effects of concentration and market power on the rate of adjustment. The conjectural variation represents a firm’s expectation about the reaction of industry output or price to a change in the firm’s own output or price (in other words, the parameter measures the firm’s belief about its price-making ability). In Worthington’s quantity model and in Martin’s price model these exogenously determined ‘conjectures’ are placed into each firm’s instantaneous profit function within a repeated game and, as in Ginsburgh and Michel (1988), a sluggishness coefficient is derived from first-order conditions. This coefficient is seen to vary both with the number of firms and with the value of the firm’s conjecture, and in this way the distinct influences of concentration and market power on adjustment speeds can be observed.

This modelling approach appears attractive for several reasons. Because separate parameters for concentration and market power are incorporated into the models, the effects of ‘fewness of firms’ and ‘competitiveness of conduct’ – which are somewhat conflated in the hypotheses described above – can be distinguished. Because conjectural variations parameters are frequently used to proxy market power in static models, and are widely used in econometric estimation to measure competitiveness of conduct, their interpretation within these dynamic models also seems straightforward. Finally, because conjectures literally represent expected reactions of other firms, their use appears to have potential to incorporate into a mathematical framework some of the notions of ‘expected response’ that are important in the hypotheses described above (price conjectures could be used to model the industry profitability effect suggested by Bedrossian and Moschos (1988), for example).3

However there are conceptual problems with the use of conjectural variations which, although they have been extensively debated and ultimately

---

3 Although note that these initial modelling efforts have not incorporated many of the other crucial features in the hypotheses described above, such as asymmetric information about prices, uncertainty about timing of price movements, and inter-firm profitability differences are absent.
do not invalidate the use of the technique, significantly alter the correct interpretation of conjectures. This chapter presents an original argument that the way in which conjectures have been introduced as a parameterisation of market power in Worthington (1989) and Martin (1993) suffers from these game-theoretic problems, with the implication that the method used by these authors to study the impact of market power on adjustment speeds within the setting of a dynamic model is unreliable. Briefly, only when a conjecture parameter is used in a static model as a representation of the outcome of a dynamic ‘supergame’, or when conjectures held by firms within dynamic games are constrained to be consistent with (or at least updated with) reactions that actually occur, can their value have any sensible meaning. Section 5.2 of this chapter describes the basic structure of the Worthington (1989) and Martin (1993) models and argues this point, after first introducing the concept of conjectural variations and briefly sketching the debates and consensus over their appropriate use in modelling.

This point has apparently not been made elsewhere, despite much interest in the relationship between competitiveness of markets and price rigidity. Martin’s paper in particular has been widely cited in support of the view that more intense competition accelerates the rate of adjustment of prices; see the European Central Bank working paper by Przybyla and Roma (2005), research papers from various European national banks including Vermeulen et al. (2007), Álvarez and Hernando (2006), and Fabiani et al. (2005), as well as the Bank of England working paper by Small and Yates (1999). As such, it is important to examine the methodological underpinnings of the modelling techniques used, and the critique laid out below provides a valuable contribution to the literature on dynamic modelling of price adjustment speeds.

The second contribution of the chapter is somewhat more constructive, offering a solution to the problem raised by the critique of Worthington and Martin’s approach to parameterising market power. The contribution here is to suggest and demonstrate how the goal of examining the relationship between market power and adjustment speeds via dynamic models can still be attained, using a conceptually-sound interpretation of conjectural variations
as a market power parameter. The logic is as follows: because the conjecture that emerges as a static ‘shorthand’ for the steady state equilibrium of each type of dynamic model is a function of the parameters of the dynamic model, as is the adjustment speed, the variation of market power and adjustment speeds as underlying parameter values change can be observed, giving a picture of the nature of the relationship as it holds within each class of model.

By way of example, the dynamic duopoly model from Dockner (1992) is presented, and it is shown that in the context of this model higher adjustment costs and discount rates tend to increase the steady state profit margin while also reducing the speed of adjustment to shocks. While no new modelling is required to derive these results (they flow straightforwardly from Dockner’s model), the novel aspect of the chapter is in demonstrating a method by which dynamic models such as this – despite being designed to examine quite different questions – might be used to understand the market power-adjustment speed relationship. Using Dockner’s model to demonstrate this approach also allows comparisons with the results of Part I and of Worthington (1989), since in each case the modelling environment is nearly identical (quantity-setting firms with convex adjustment costs as the ‘friction’). The models differ only in the endogeneity or exogeneity of conjectures and in the time horizon (two-stage or infinite-horizon).

Dockner’s paper is part of a literature that seeks to understand how the informational environment and the structure of dynamic interactions between firms affects their conduct – their optimal responses to rivals and the aggressivity of competition – and, ultimately, affects market outcomes, particularly the mark-up of price over cost in steady-state equilibrium. Dockner (1992) and Cabral (1995) have pointed out that the steady-state markup can be represented by a particular static conjectural variations equilibrium, thereby giving static conjectural variations a solid conceptual grounding. However the concern of these authors has not been with the speed of adjustment to shocks, and as yet it has not been considered that the insights of this liter-

\footnote{See also Fershtman and Kamier (1987), Maskin and Tirole (1987), Worthington (1990), Cabral (1995), and Friedman and Mezzetti (2002).}
This chapter thus brings together two previously-unconnected strands of the literature: the work relating dynamic models to static conjectural variations, and the enquiry into how market power influences the speed of adjustment to shocks. The idea that insights into the latter relationship might be realised from dynamic models, by virtue of the joint reliance of endogenous market power and adjustment speed parameters on underlying structural parameters, has not been suggested or demonstrated before. It offers a promising approach to incorporating the widely-used measure of competitiveness of conduct – the conjectural variations parameter – into models explaining the speed of adjustment, free from the conceptual problems arising from the previous attempts to do so through exogenous ‘dynamic conjectures’.

The chapter is structured as follows. Section 5.2 discusses the conjectural variations concept, talks through the basic setup of the Worthington (1989) and Martin (1993) models, and presents the argument that the application of conjectural variations there is without clear grounding. Section 5.3 reviews Dockner’s (1992) model, showing the equivalence of the dynamic model steady state with a static conjectural variations equilibrium. Section 5.4 derives an adjustment speed measure from Dockner’s model, then examines how this relates to the static conjectural variations measure of market power, with both being driven by the underlying parameters of the model. Section 5.5 concludes the chapter.

5.2 Conjectural variations analysis and dynamic models

5.2.1 Introduction

The conjectural variations model was first introduced by Bowley (1924) as a means of generalising the static Cournot model so that its particular assumption – that each firm behaves as if its rival will not react to its own move – became just one of a range of conjectures that a firm might potentially
In the model, each firm was presumed to hold an expectation about how its rival would respond to a change in output, which then influenced its perceived optimal strategy; that is, changed the slope of that firm’s reaction function. An expectation of no response would generate the reaction functions and equilibrium of the Cournot model, but other beliefs would create other equilibria; hence without specifying the conjectures the outcome of the model was indeterminate.

The basic model in a quantity-setting context is as follows. Each firm maximises a profit function as in Cournot, but with the belief that its rivals will react to its output choice with a variation in their own output levels, as expressed in terms of conjectures $v_{ij}$ held by firm $i$ for each rival firm $j$. Each $v_{ij}$ is defined as $v_{ij} = dq_j/dq_i$, and a conjectured response of all rivals to firm $i$ can be defined as $\chi_i = \sum_{j \neq i} dq_j/dq_i$.

The maximisation problem is:

$$\max_{q_i} \pi_i (q_i) = p(Q)q_i - C_i(q_i)$$  

(5.1)

and the resulting first order condition is:

$$0 = p'(Q)\frac{dQ}{dq_i}q_i + p(Q) - C'_i(q_i)$$  

(5.2)

$$= p'(Q) (1 + \chi_i) q_i + p(Q) - C'_i(q_i)$$

This differs from the corresponding condition in the Cournot model only by the term $\chi_i$. To express this first-order condition in the form of a mark-up of price over cost, the second line of (5.2) is rearranged to give:

$$p(Q) \left( \frac{p'(Q)}{p(Q)} (1 + \chi_i) q_i + 1 \right) = C'_i(q_i)$$  

(5.3)

The name ‘conjectural variations’ comes from Frisch (1951 (1933)). For a history of the early debate over the technique, see Giocoli (2003). Aside from the particular assumption mentioned, the other problematic aspect of Cournot was the confusion between dynamics and statics that allowed a firm’s reaction functions to depend on the other firm’s output (which should be unknown in a simultaneous move situation). However this raised fewer objections and was carried forward into the conjectural variations model which was also based on a static setup (Giocoli 2003, p177).
5.2. CONJECTURAL VARIATIONS

The first term inside the left-hand-side bracket is multiplied and divided by industry quantity $Q$ to give:

$$p(Q) \left( \frac{p'(Q)Q}{p(Q)} \left( 1 + \chi_i \right) \frac{q_i}{Q} + 1 \right) = C'_i(q_i) \quad (5.4)$$

Using $\eta$ to represent the industry demand elasticity, and $S_i$ to represent firm $i$’s share of industry quantity, the above expression can be represented as:

$$p(Q) \left( 1 + \frac{S_i}{\eta} \left( 1 + \chi_i \right) \right) = C'_i(q_i) \quad (5.5)$$

Equation (5.5) is a static conjectural variations model equilibrium, and unlike the standard models of price determination under monopoly, perfect competition, or Cournot it specifies an indeterminate outcome, even if firms are assumed to have identical marginal cost functions and if the demand function is known. The exogenous ‘conjectures’ ultimately determine the equilibrium price and quantity, and since there is no basis on which to know firms’ conjectures about their rivals’ responses for any particular market – even putting aside the objection that the concept of a response within a simultaneous one-shot game is fundamentally flawed – the conjectural variations model itself does not constitute a useful theory to understand market outcomes.

However a conjectural variations equilibrium can be seen to ‘nest’ a range of equilibria of different models of price determination: monopoly, Cournot, and perfect competition for example, making it valuable for econometric work. For a monopolist, for instance, marginal revenue is given by $p'(Q)Q + p(Q)$ and the profit-maximising first-order condition is thus:

$$p'(Q)Q + p(Q) = C''(Q) \quad (5.6)$$

which can be rearranged to:

$$p(Q) \left( \frac{p'(Q)Q}{p(Q)} + 1 \right) = C'_i(q_i) \quad (5.7)$$

$$p(Q) \left( 1 + \frac{1}{\eta} \right) = C'_i(q_i)$$
That is, the monopoly outcome is equal to a conjectural variations equilibrium with $\chi_i = 1$ (and $S_i = 1$). The competitive outcome with $p(Q) = C'_i(q_i)$ is similarly the same as a conjectural variations equilibrium with $\chi_i = -1$.

In Cournot competition, the first-order condition under which marginal revenue equals marginal cost is:

$$p' (Q) . q_i + p (Q) = C'_i(q_i) \quad (5.8)$$

$$p (Q) \left( \frac{p' (Q) Q . q_i}{p (Q)} + 1 \right) = C'_i(q_i)$$

$$p (Q) \left( 1 + \frac{S_i}{\eta} \right) = C'_i(q_i)$$

This is identical to a conjectural variations equilibrium with $\chi_i = 0$ (for any $S_i$). The Cournot outcome as a special case of a conjectural variations equilibrium can be even more simply understood by recalling that the setup of the Cournot model requires a particular assumption – no reaction from rivals to a change in own-output – which itself represents a literal conjecture (of zero response).

The property of the conjectural variations model as nesting other models with varying degrees of competition makes it a convenient form for econometric estimation. Parameter $\chi_i$ in equation (5.5) effectively captures all aspects of the market and the interactions between firms which determine the price-cost mark-up, apart from the elasticity of demand, and the number of firms (as reflected in $S_i$).

For econometric applications a further-simplified form of equation (5.5) is typically used:

$$p (Q) \left( 1 + \frac{\theta_i}{\eta} \right) = C'_i(q_i) \quad (5.9)$$

Here the product of $1 + \chi_i$ and $S_i$ is denoted $\theta_i$, the conjectural elasticity (so-called because $(1 + \chi_i) . S_i = (dQ/dq_i) . (q_i/Q)$, meaning $\theta_i$ here represents the firm’s conjecture about the proportional response of industry quantity to own-quantity). In this form, parameter $\theta_i$ captures all aspects of the market (other than the demand elasticity) that bear on the competitiveness of the market outcome. It is sometimes also referred to as the ‘elasticity-adjusted Lerner index’ for the same reason. The Lerner index is the mark-up of price
over marginal cost as a proportion of the price, \( L = (p - C'(q_i)) / p \). Adding \( p \) to both sides of (5.9) and then dividing by \( p \) shows that \( L = -\theta_i / \eta \), or \( \theta_i = L \eta \).

Estimation of this parameter (alongside estimation of the demand elasticity) is common in studies of market power and competitiveness. Comparing (5.9) to the expressions above, it can be seen that a 0 value for parameter \( \theta_i \) represents perfect competition, \( 1/N \) represents an \( N \)-firm symmetric Cournot oligopoly, and \( \theta_i = 1 \) represents the monopoly outcome.

### 5.2.2 Methodological issues

From the mid-1930s concerns with the logical underpinnings of the conjectural variations model began to be expressed. These included the questions of how a firm’s conjectures were formed, and why firms with the reasoning capacity to predict the behaviour of their rivals would fail to jointly realise that the mutually beneficial collusive outcome could be achieved (Giocoli 2003, p199).

More fundamentally, however, the static conjectural variations model contains the conceptual flaw that the inherently dynamic notion of a response, or reaction, is posited within the constraints of a static model. That is, it is illogical that a firm which believes it is playing a one-shot simultaneous game, where it cannot observe its rival’s strategy before choosing its own, is supposed to also hold the mutually inconsistent belief that its rival will observe that choice and respond in a predictable way. For this reason, the literal interpretation of conjectures as ‘beliefs held’ is unrealistic.

A second conceptual problem is that firms in the conjectural variations model are permitted to hold expectations about behaviour that are not realised in practice. To rescue the static conjectural variations model from this critique, the allowable conjectures can be constrained to be those which are the same as the actual slopes of rival firms’ reaction functions in equilibrium:

---

\(^6\)Surveys of methods and applications in this literature are provided by Perloff et al. (2007), Sheldon and Sperling (2003), and Bresnahan (1989).

\(^7\)See, for instance, Friedman (1983, p107-110) or Tirole (2002, p244-245) for discussion.
consistent conjectures’. However a static model with consistent conjectures in equilibrium still melds dynamic and static concepts in an illogical manner, as well as failing to explain how the equilibrium and the consistent conjectures are arrived at.

5.2.3 Conjectures as ‘shorthand’

One modern interpretation of conjectural variations involves dropping the assumption that the conjecture parameters literally represent beliefs, thus avoiding the critiques associated with inconsistency and with ‘shoehorning dynamic interactions into a static game’. In this view, a conjecture parameter is viewed as simply a static-game representation of the degree of competition associated with some unknown form of dynamic interaction. The conjectural variations equilibrium is shorthand for the steady state outcome of an unspecified dynamic supergame. That is, conjectural variations are "best interpreted as reduced form parameters that summarize the intensity of rivalry that emerges from what may be complex patterns of behavior" (Schmalensee 1989, p650).

Dockner (1992) and Cabral (1995) have illustrated this point by deriving expressions for the conjectures \( \chi_i \) that represent the steady state outcomes of two types of dynamic game. These papers show that the steady state equilibria of the games examined are identical to particular conjectural variations equilibria; that is, that the steady-state conditions can be specified in the form of equation (5.5). Dockner and Cabral’s games are, respectively, a continuous-time quantity-setting duopoly with adjustment costs as the ‘friction’, and a repeated game with trigger strategies and punishment. Demonstrating that the steady state outcomes of these dynamic supergames

---

8Amongst other authors, Bresnahan (1981) has explored this concept. The equilibrium with consistent conjectures is characterised by Bresnahan as "firms being right for the right reason", since conjectured responses and reaction functions coincide. This contrasts with a Cournot equilibrium where firms are right in their expectation of the output of their rivals, but "for the wrong reasons" since they assume that output is constant (by conjecturing zero response to their own action) when in truth their rivals have their own (negatively sloped) reaction functions.

9See Figuières et al. (2004) for a textbook-style survey and discussion of the modern uses of conjectural variations either as ‘shorthand’, or within dynamic models.
are equivalent to equilibria of particular static conjectural variations games supports the conjectures as shorthand interpretation. It also allows bounds to be placed on the values of the static conjectures that are potentially representative of these forms of dynamic interactions.\footnote{For instance, for each discount rate in Cabral (1995) there is a single $\chi_i$ which is the conjectural variation equivalent of all linear oligopolies in this form of repeated game, and this $\chi_i$ is shown to be positive. That is, repeated games with trigger strategies and punishment (unsurprisingly) induce less competitive outcomes than Cournot.}

5.2.4 Dynamic conjectural variations

A second, and distinctly different, use of the conjectural variations concept is as an expectation held by a firm within each period of a multi-period game. The conjectures are typically about the future responses of rivals, thereby avoiding the static-dynamic confusion of the original model. To avoid inconsistency, the conjectures can be constrained to be equal to the slope of the rivals’ true reaction functions, thus making the conjectures endogenous to the model, not a parameter chosen by the modeller. Examples of this are the ‘consistent contemporaneous conjecture’ as labelled by Worthington (1990) and the ‘equilibrium dynamic conjectural variation’ of Riordan (1985), which are defined by being both the beliefs of a firm about reactions of rivals, and the actual profit-maximising reactions of those rivals.

Alternatively, rather than enforce consistency, Friedman and Mezzetti (2002) argue that conjectural variations are best interpreted as simple rules held by boundedly rational firms, which capture each firm’s understanding of the nature of their strategic interactions with others. In their model conjectures are allowed to be initially incorrect, but to adapt in the light of observed behaviour until at the steady-state equilibrium they are finally consistent.

5.2.5 Worthington (1989) and Martin (1993)

Worthington and Martin’s models fit within the category of models described in the previous subsection, since they place conjectures within the instantaneous profit function of each period of a dynamic game. The motivation and approach of both papers is similar: the authors’ concern is with the
influences of market structure and conduct on the rate of price adjustment, and both use dynamic models where movement to equilibrium is slowed by adjustment costs. In both cases, a sluggishness coefficient is derived which represents the amount by which the current-period outcome is determined by the previous-period outcome.

The model in Worthington (1989) is an extension of that in Ginsburgh and Michel (1988), with the single modification being the addition of a conjectural variation parameter. Firms in this model choose their output level to maximise single-period profits, leading to the following first-order condition for each firm $i$:

$$\frac{\partial \pi_{it}}{\partial q_{it}} = \alpha - \beta q_t - \beta (1 + \chi) q_{it} - \gamma_i - c (q_{it} - q_{i,t-1}) \quad (5.10)$$

Variables $q_t$ and $q_{it}$ are industry and firm output respectively, while $\alpha$ and $\beta$ are parameters of a linear demand curve, $\gamma_i$ is marginal cost, $c$ is the quadratic adjustment cost parameter, and $\chi$ represents the expected response of the other $n-1$ firms to a one-unit change in firm $i$’s output, $\chi = \sum_{j \neq i} dq_j/dq_i$ (assumed to be the same for all firms). In each period, industry output is shown to be given by the following rule:

$$q_t = \lambda q_{t-1} + \text{constant} \quad (5.11)$$

with the sluggishness coefficient given by:

$$\lambda = \frac{c}{\beta (n + 1 + \chi) + c} \quad (5.12)$$

The novel finding that emerges from Worthington’s paper is that ‘fewness of firms’ and ‘competitiveness of conduct’ influence the adjustment speed, in this modelling framework, in ways which suggest that the net effect of concentration on adjustment may be ambiguous. With fewer firms (lower $n$) adjustment is slower (higher $\lambda$), but with less competitive conduct (higher $\chi$) adjustment is faster (lower $\lambda$). Since both lower $n$ and higher $\chi$ are typically associated with higher concentration, there may exist no simple link between concentration and the adjustment speed.\(^{11}\)

\(^{11}\)It is noted in the paper that generalising the firm’s maximisation problem to an
5.2. CONJECTURAL VARIATIONS

In the model of Part I concentration and competition were effectively bundled together in a single parameter, the conjectural elasticity $\theta$; higher $\theta$ resulted in a faster speed of adjustment. Increasing $\theta$ for any given number of firms represented more collusive conduct by those firms, and the faster adjustment that resulted is in line with Worthington’s finding with respect to $\chi$.

Whereas Worthington finds that fewer firms would slow the rate of adjustment, however, the effect of reducing the number of firms in the model of Chapter 2 could be expected to accelerate the rate of adjustment, holding the competitiveness of conduct constant. In Chapter 2 it was noted that if, for example, firms expected no reaction from rivals (as in ‘open-loop’ strategies where rivals’ actions are unobservable) then $\theta$ would be equal to $1/N$. An alteration in the number of firms would alter $\theta$ inversely, so fewness of firms would raise the rate of adjustment, just as a decrease in the fierceness of competition between an unchanged number of firms would.

For non-zero reactions, however, it might be meaningless to consider $N$ moving independently of the parameter representing competitiveness. Consider Worthington’s model in the case where, for instance, a collusive duopoly ($N = 2$, $\chi = 1$) merged into a monopoly ($N = 1$). Predicting the impact on the adjustment speed by retaining the assumption that $\chi = 1$ is obviously a flawed approach, as any $\chi$ other than 0 is nonsensical in a monopoly situation. A decrease in $N$ in the model of Chapter 2 would also presumably alter the expected reaction of a firm which previously held some (non-zero) expectation about the collective response of industry rivals, making the effect of $N$ on $\theta$ ambiguous. Generally, treating a conduct parameter (like $\chi$) as if it were independent of the structure of the market ($N$) seems problematic. It is not clear how to resolve these problems, and since they arise from the treatment of conjectures as exogenously-determined dynamic reactions – argued below to be a conceptually-flawed approach – it is not obvious that a resolution is possible.

The intertemporal one (where future adjustment cost savings are taken into account) does not change the direction of influence of $n$ and $\chi$, but generalising the adjustment cost function to non-quadratic (less convex) forms creates ambiguity in the direction of the effect of $n$ on $\lambda$. 
Martin’s framework is similar to Worthington’s, but with the firms’ first-order conditions and the speed of adjustment coefficient derived from a price-setting optimisation problem. In this problem, firms minimise a loss function equal to the net present value (over an infinite horizon) of the sum of the adjustment costs incurred in altering the price faster than a particular rate, and the losses incurred from setting the price away from the firm’s preferred level. These losses depend on the curvature of the single-period profit function, which in turn is a function of the conjecture held by the firm. The adjustment speed follows from first-order conditions, as in the other models. Adjustment is slower with fewer firms, and also slower when the conjecture parameter is larger or conduct is more collusive (the latter result is the opposite of that found by Worthington).

5.2.6 Critique of Worthington (1989) and Martin (1993) approach

It is argued here that in light of the discussion in Section 5.2 above, it appears that the use of conjectural variations in these models is not methodologically sound. The conjectures suffer, firstly, from the basic criticism of ‘inconsistency’. That is, no constraint is placed on the value of the exogenously-given conjecture parameters, meaning that they will not be equal to the actual reactions observed. In the environment of complete information in these models, the assumption of exogenous and inconsistent conjectures implies an inability to deduce the correct reaction functions, or irrationality, on the part of firms. However exogenous and inconsistent conjectures are also not easily justified as a form of ‘bounded rationality’ in this situation since the beliefs are assumed to remain constant, without updating in the light of observed behaviour. In a dynamic setting, the assumption that firms fail to modify their own incorrect conjectures held in each period, despite observing the actual reactions of their rivals in previous periods, is clearly problematic. Only in circumstances where the actions of other firms are actually unobservable or firms must ‘pre-commit’ to their quantity or price adjustments would such an assumption be defendable. While this was explicitly
assumed in the model of Chapter 2 (and is an acknowledged limitation of that model), there is no indication in Worthington (1989) or Martin (1990) that their models are intended to apply to boundedly-rational firms in an environment of pre-commitment or unobservable prices.

Secondly, the contemporary interpretation of conjectural variations as representative of an unknown dynamic ‘supergame’ is also unavailable, because the conjectures in question are not simply arguments of the profit function in a single-shot game, but are held by firms in every period of a multi-period model. The ‘conjectures as shorthand’ interpretation in this situation would imply that the firm’s actions in each period were in fact the outcome of another, unmodelled, dynamic interaction (a dynamic game within each period of a dynamic game), which is clearly not intended by the authors.

Finally, the fundamental problem of a response modelled as taking place in a static context remains present in the setup used in these models. Each firm in the Worthington (1989) and Martin (1993) models expects a contemporaneous reaction by its rivals, despite the information on which to base such a response being unavailable (just as the current-period actions of other firms are unknown to the firm in question).

In summary, although the apparently-straightforward meaning of conjectural variations – as a measure of ‘price-making’ ability – make their use in studying the relationship between adjustment speeds and market power natural, there are conceptual issues with the approach that make the choice of modelling technique and the interpretation of the parameter crucial. In a dynamic model where firms hold conjectures about rival responses, these conjectures must satisfy certain conditions. They could be constrained to match actual responses, or be subject to revision based on these responses (as part of an explicitly boundedly-rational theory), or be justified as a representation of a specific informational or commitment environment, in order for the approach to make sense. Although the dynamic models of Worthington (1989) and Martin (1993) do not satisfy this criterion, the next sections show that a market power – adjustment speed analysis can be undertaken on the basis of the alternative interpretation of conjectures, as a static representation of
CHAPTER 5. MARKET POWER AND ADJUSTMENT SPEEDS

5.3 Static conjectures from a dynamic model: Dockner (1992)

5.3.1 Introduction

As described in Section 5.2, the use of static conjectural variations analysis can be justified by supposing not that an arbitrarily-chosen conjecture parameter measures a belief literally held by a firm, but that it is a static representation of the steady-state outcome of some dynamic process. In particular, the conjecture is a representative measure of the competitiveness of the conduct of the industry as determined by the dynamic interactions between firms. This interpretation is well-recognised, and is supported by research that links specific types of dynamic games to particular conjecture values by proving that the steady states are equivalent to conjectural variations equilibria in the form of equation (5.5).

In these papers (Dockner 1992, Cabral 1995, and others cited in Figuières et al. 2004) the value of the static conjecture is obtained as a function of the parameters of the dynamic model. For certain types of model, an adjustment speed measure as a function of these parameters may also be derived. Therefore, by exploiting the dual dependence of the conjecture and the adjustment speed on the underlying parameters of the dynamic model, the direction of the relationship between market power and the rate of adjustment within each model may be observed. This approach may potentially be repeated for any dynamic oligopoly model, helping build a picture of this relationship across different classes of models which vary in their informational environment, strategy space, and assumed rationality of firms, for instance.

This section demonstrates this approach by exploring the model of Dockner (1992), a continuous-time quantity-setting duopoly game where firms are able to adjust their output on the basis of the observed output of their rival (closed-loop or ‘feedback’ strategies). Firstly, the equivalence of the steady state condition to a conjectural variations equilibrium for this generally-
specified game is shown. Because an exact expression for the equivalent conjecture cannot be found without further specification of functional forms, a linear-demand quadratic-cost version is then presented and the equivalent conjecture is derived. The derivation of an adjustment speed parameter and discussion of the observed relationship with market power are left to the following section of the chapter (Section 5.4).

5.3.2 General model

The most general version of the continuous-time adjustment cost model involves two firms which maximise a stream of profits, discounted at a rate $r$, over an infinite horizon. Each firm $i$ chooses an output level $q_i$, has a cost function $C_i(q_i)$, and faces a market price determined by the inverse demand curve $p = p(Q)$ where $Q$ is the industry quantity. The model is given a time dimension by assuming that each firm faces a convex adjustment cost given by function $A_i(\dot{q}_i)$, which creates a tradeoff between the losses from choosing a sub-optimal quantity and the costs of adjusting that quantity. For clarity, the time arguments $(t)$ are generally suppressed in the text below. Except for steady state values which are denoted by the superscript *, the variables $q_i$, $q_j$, $\dot{q}_i$ etc should be read as $q_i(t)$, $q_j(t)$, $\dot{q}_i(t)$ etc.

If firms do not have the option of commitment to a quantity path, and can observe their rival’s output, then closed-loop strategies of the form $\dot{q}_i = \dot{q}_i(q_i, q_j)$ are appropriate (see Dockner et al. (2000) for discussion of how the various choices of strategy space are related to the assumed informational environment of the firm). Under these conditions, the firm’s optimisation problem is a differential game, given by:

$$\max_{\dot{q}_i(q_i, q_j)} \int_{t=0}^{\infty} e^{-rt}[p(Q(t))q_i(t) - C_i(q_i(t)) - A_i(\dot{q}_i(t))]dt$$  \hspace{1cm} (5.13)

Several minimal restrictions on the form of $p(Q)$, $C_i(q_i)$, and $A_i(\dot{q}_i)$ are necessary to ensure that the instantaneous profit function $p(Q) - C(q_i) - A(\dot{q}_i)$ is strictly concave in $q_i$ and $\dot{q}_i$ (these are outlined in Dockner (1992, p381)). In particular, it is assumed that adjustment cost function $A(\dot{q}_i)$
and its first derivative $A'(q_i)$ both take value 0 at $q_i = 0$, and that the adjustment cost is a function of the rate of output change $q_i$ only. These restrictions ensure existence and global asymptotic stability of an open-loop equilibrium (i.e. Nash equilibrium strategies of the form $q_i = q_i(t)$), but not necessarily of a closed-loop equilibrium. When analysing the steady state markup condition, it is assumed that such an equilibrium exists, and Dockner shows that the equilibrium closed-loop strategies derived for the linear-quadratic model are indeed globally stable.

**Steady state markup condition**

The steady state equilibrium is given by the following condition, showing the markup of price over marginal cost (see Appendix C for working):

$$p^* (Q^*) \left( 1 + \frac{S_i}{\eta} \left( 1 + \frac{\partial q_i}{\partial q_i} \frac{\partial q_i}{\partial q_j} \frac{r}{\partial q_j} \right) \right) = C_i' (q_i^*)$$  \hspace{1cm} (5.14)

Superscript *’s on the quantity and price variables represent steady state values. The terms $\partial q_i / \partial q_i$ and $\partial q_j / \partial q_j$ are partial derivatives of the unknown mathematical expressions representing the optimal strategies of the firms. In the open-loop-strategy version of this game, where firms must commit to output plans at the beginning of the game (so that their output choices throughout the remainder of game are time-dependent but not state-dependent), $\partial q_j / \partial q_i$ is zero and the steady state outcome is thus the same as in the Cournot model (Dockner 1992, p383).

It is clear from comparison of equations (5.5) and (5.14) that this dynamic adjustment cost game yields, as its steady state equilibrium, an outcome which is identical to a particular conjectural variations equilibrium. That is, where the term $\chi = -\frac{\partial q_j / \partial q_i}{r-\partial q_i / \partial q_i}$ is used as a firm’s conjecture in a static conjectural variations game, the outcome (quantities and price in equilibrium) is identical to the outcome (quantities and price in steady state) of the general oligopoly game with adjustment costs.

Just as the conjecture $\chi_i$ in a static conjectural variations model is a convenient representation of the competitiveness of the industry, the term $\chi = -\frac{\partial q_j / \partial q_i}{r-\partial q_i / \partial q_i}$ – referred to by Dockner as the ‘dynamic conjecture’ – usefully
5.3. DOCKNER (1992) MODEL

describes the market power of the firms in this dynamic setting as realised in the steady-state markup.

5.3.3 Linear-quadratic case

A higher level of specification of the adjustment cost game is needed in order to place a value on $\chi$, and to observe the speed of adjustment to steady state. Dockner (1992) also presents a linear-quadratic game, which builds on a version first examined by Driskill and McCafferty (1989) by providing a generalisation to asymmetric production costs.

The market demand curve in this game is assumed to be linear:

$$p(Q) = a - Q \quad (5.15)$$

Production costs are quadratic, and may be asymmetric in the linear terms:

$$C_i (q_i) = c_i q_i + \frac{b}{2} q_i^2 \quad (5.16)$$

and adjustment costs are quadratic and symmetric across firms:

$$A_i (q_i) = \frac{k}{2} q_i^2 \quad (5.17)$$

The differential game is summarised as:

$$\max_{\dot{q}_i(q_i, q_j)} \int_{t=0}^{\infty} e^{-rt}[(a - Q)q_i - c_i q_i - \frac{b}{2} q_i^2 - \frac{k}{2} q_i^2] dt \quad (5.18)$$

Feedback rule

Dockner (1992, p391) derives the optimal strategies of the firms using a value function approach (see Appendix C).\textsuperscript{12} Despite the asymmetry in production costs, the optimal strategies are symmetric in their ‘feedback rules’ (with only the intercept terms $\beta_i$ varying across firms):

$$\dot{q}_i^*(q_i, q_j) = \frac{1}{k} [\beta_i + \delta q_i (t) + \sigma q_j (t)] \quad (5.19)$$

\textsuperscript{12}More precisely, the paper finds a unique, globally and asymptotically-stable equilibrium strategy within the class of linear closed-loop (subgame-perfect) strategies. The existence of an equilibrium in non-linear strategies is not necessarily ruled out.
The coefficients $\delta/k$ and $\sigma/k$, and the intercept term $\beta/k$, are complicated functions of parameters $r$, $k$, and $b$, with $\delta$ and $\sigma$ being negative and $\beta$ positive. In this expression, the term of most interest is $\sigma/k$: it represents the effect of firm $j$’s output level on the output choice of firm $i$. In the reaction functions of a static Cournot model, a higher value of firm $j$’s output implies a lower optimal output choice of firm $i$. Analogously, in the feedback rule of this dynamic model a higher value of firm $j$’s output induces firm $i$ to reduce output (seen here through the negativity of $\sigma/k$).

As explained in Sections 5.2.3 and 5.2.4, there are two distinct concepts of conjectures that remain in use: the metaphorical understanding, where the conjecture is seen as a static rendering of the end result of a dynamic process, and the literal understanding of a conjecture as a belief or expectation held by a firm at each point in time. The ‘dynamic conjecture’ $\chi$ derived by Dockner (1992) is an example of the former. However, although it is not discussed, the closed-loop feedback coefficient $\sigma/k$ is also an example of a consistent conjecture of the latter type: a (correct) belief held by each firm about the reaction of the other firm to a one-unit increase in own-output. This interpretation of a closed-loop equilibrium feedback parameter as a consistent conjecture which literally corresponds to a firm’s expectation is suggested by Worthington (1990), who labels the conjecture a ‘dynamic version of a contemporaneous conjectural variation’.

Equation (5.19) is the most important result, for the purposes of this chapter, from Dockner’s model. It describes the dynamics of each firm’s output (and thus the market output) from any initial position in terms of three parameters from the dynamic model: the discount rate $r$, the adjustment cost parameter $k$, and the parameter representing the convexity of the firm’s production costs $b$. It can therefore be used to derive an adjustment path towards steady state equilibrium and a measure of the speed of adjustment to

\[ \frac{\sigma}{k} = \frac{sk}{2s^2k^2 + 2k + bk - 2\sigma^2} \]

The intercept terms $\beta_i$ are not solved for and are not required here. Note the expression for $D$ is from Reynolds (1987); the version in Dockner (1992) is subject to a typographical error.
this equilibrium, in terms of these three parameters (the task of Section 5.4). The second use of (5.19) is to calculate the value of the dynamic conjecture – the parameter representing market power – as a function of $r$, $k$, and $b$ also.

**Steady state markup condition**

The steady state markup follows immediately from the feedback rules and the steady state condition for the general model, equation (5.14): 

$$ p(Q) \left( 1 + \frac{S_i}{\eta} \left( 1 + \frac{\sigma}{rk - \delta} \right) \right) = C'_i(q_i) $$

(5.20)

The dynamic conjecture is given by:

$$ \chi = \frac{\sigma}{rk - \delta} $$

(5.21)

It can be seen from the negativity of $\sigma$ and $\delta$ that the dynamic conjecture $\chi$ is negative, implying that the steady state markup of price over cost is less than in the standard Cournot oligopoly (a value $\chi = 0$ would generate the Cournot markup, while $\chi = -1$ would correspond to the perfectly competitive outcome). That is, this particular form of dynamic interaction – albeit with a simple ‘closed-loop’ strategy space, and thus none of the possibilities for trigger strategies and punishment that can sustain a collusive outcome – generates more competitive outcomes than the one-shot Cournot quantity game.

This result can also be reasoned intuitively from the feedback rules. In firm $i$’s calculus within the dynamic setting, the benefit of reducing output in order to raise the market price would be somewhat offset by firm $j$’s reaction of increasing his own output (according to the negativity of $\sigma$ in the firm $j$ version of (5.19)). In contrast, in the single-shot quantity game where there can be no reaction by firm $j$, firm $i$ has a greater incentive to set output at a lower level with the equilibrium result of less competitive behaviour.
CHAPTER 5. MARKET POWER AND ADJUSTMENT SPEEDS

5.4 The market power – adjustment speed relationship

This section builds on Dockner’s (1992) linear-quadratic solution presented above by deriving an adjustment path from the ‘feedback rules’, and an associated measure of the speed of adjustment. By exploiting the dependence of both the market power parameter $\chi$ and the adjustment speed measure on $r$, $k$, and $b$, the direction of the relationship between these parameters is examined via a numerical exercise.

5.4.1 Adjustment path and adjustment speed measure

Using the optimal strategies in equation (5.19), the adjustment of quantities and prices to steady state can be shown to follow a continuous-time version of a ‘partial adjustment’ process. This was not shown in Dockner (1992) but has been derived in Appendix C. In each period of time the price adjusts by a fixed proportion of the difference between current and steady state output, as shown below:

$$\dot{p} = \frac{-\delta + \sigma}{k} \left( p^* - p(t) \right) \tag{5.22}$$

The parameter representing the speed of adjustment, denoted $\lambda$, is thus given by $\lambda = -(\delta + \sigma)/k$.

The partial adjustment coefficient is frequently used in econometric studies as a measure of the speed of adjustment of prices. Within the literature exploring the competitiveness / adjustment speed relationship, for instance, see Bedrossian and Moschos (1988) or Neumark and Sharpe (1992). Hamermesh and Pfann’s (1996) survey discusses the use of the partial adjustment structure in other areas, as well as noting issues with the underlying theoretical rationale for restricting estimated adjustment patterns to that structure.

Since parameters $\delta$ and $\sigma$ are both negative, it follows that any change in the underlying parameters of the model that reduce $\delta/k$ or $\sigma/k$ (make them more negative) will also increase the rate of adjustment towards the steady state. In particular, those factors that tend to encourage more aggressive responses to changes in a firm’s output (via larger compensatory movements
5.4. MARKET POWER AND ADJUSTMENT SPEEDS

in the other firm’s output, or a larger magnitude of $\sigma/k$) will also lead to faster adjustment. Since these more aggressive responses are associated with more competitive outcomes (as argued in the previous section), it seems intuitively likely that more competitive conduct will be associated with faster adjustment in this model. Although an increase in the magnitude of the parameter $\delta$ is also associated with faster adjustment, the interpretation of this parameter in terms of competitiveness of behaviour is less obvious.

5.4.2 Variation of $\chi$ and $\lambda$ with parameters

The preceding section argued that (at least through the influence of the feedback coefficient $\sigma/k$) more competitive conduct of firms should intuitively be associated with faster adjustment to equilibrium; that is, there is an expectation that the more negative the parameter $\chi$, the more positive the parameter $\lambda$. However the complexity of the expressions for $\delta$ and $\sigma$ appear to make an algebraic proof that $\chi$ and $\lambda$ move in opposite directions as $r$ and $k$ vary impossible.

Dockner (1992, p393) shows that the limit of $\chi$ as $r$ and $k$ approach infinity is 0. That is, as the degree of discounting of future profits or the cost of adjusting output become large enough that the dynamic game approximates a one-shot game, the outcome converges to the one-shot Cournot outcome. Intuitively, as the future is more heavily discounted or the costs of altering output become large, the adjustment of output is also likely to occur ever more slowly, meaning higher $\chi$ would be associated with lower $\lambda$ in the limit. However the limit relationship cannot necessarily be assumed to be identical to that holding at finite parameter values.

As an alternative to a mathematical proof, a numerical exercise was used to uncover the relationships between parameter values. Tables 5.1 through 5.3 report results from this exercise, varying each of $r$, $k$, and $b$ in turn and showing the corresponding values of $\delta/k$, $\sigma/k$, the market power parameter $\chi$ and the adjustment speed parameter $\lambda$. The last row of each table summarises the direction of influence of each parameter. The default values used were $r = 0.1$, $k = 10$, and $b = 0.1$. The default values are irrelevant to the
direction of influence of the parameters, and these values were chosen simply so the relationships between market power and adjustment speed shown in Figure 5.1 were clearly visible on a single plot.

Table 5.1: Relationships between parameter values in Dockner (1992) model, various discount rates

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\delta/k$ ($&lt; 0$)</th>
<th>$\sigma/k$ ($&lt; 0$)</th>
<th>$\chi$ ($&lt; 0$)</th>
<th>$\lambda$ ($&gt; 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-0.42</td>
<td>-0.12</td>
<td>-0.28</td>
<td>0.54</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.38</td>
<td>-0.12</td>
<td>-0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.25</td>
<td>-0.10</td>
<td>-0.13</td>
<td>0.35</td>
</tr>
<tr>
<td>1</td>
<td>-0.17</td>
<td>-0.07</td>
<td>-0.06</td>
<td>0.24</td>
</tr>
<tr>
<td>5</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.00</td>
<td>0.06</td>
</tr>
</tbody>
</table>

$r \uparrow$ $\quad$ $\uparrow$ $\quad$ $\uparrow$ (less competitive) $\downarrow$ (slower adjustment)

Table 5.2: Relationships between parameter values in Dockner (1992) model, various adjustment cost values

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\delta/k$ ($&lt; 0$)</th>
<th>$\sigma/k$ ($&lt; 0$)</th>
<th>$\chi$ ($&lt; 0$)</th>
<th>$\lambda$ ($&gt; 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>-4.26</td>
<td>-1.20</td>
<td>-0.27</td>
<td>5.46</td>
</tr>
<tr>
<td>1</td>
<td>-1.35</td>
<td>-0.38</td>
<td>-0.26</td>
<td>1.73</td>
</tr>
<tr>
<td>10</td>
<td>-0.43</td>
<td>-0.12</td>
<td>-0.23</td>
<td>0.55</td>
</tr>
<tr>
<td>100</td>
<td>-0.13</td>
<td>-0.04</td>
<td>-0.16</td>
<td>0.17</td>
</tr>
<tr>
<td>1000</td>
<td>-0.04</td>
<td>-0.01</td>
<td>-0.08</td>
<td>0.05</td>
</tr>
</tbody>
</table>

$k \uparrow$ $\quad$ $\uparrow$ $\quad$ $\uparrow$ (less competitive) $\downarrow$ (slower adjustment)

When either $r$ or $k$ varies, the market power parameter and the adjustment speed indeed move in opposite directions. That is, in this class of model, less competitive behaviour is associated with slower adjustment to equilibrium, at least insofar as variations in the competitiveness of conduct and the rate of adjustment are driven by variation in either the rate of discounting of future profits or the costs of adjusting output. This is the opposite result to that found in Worthington (1989) and in Part I of the thesis.

This relationship breaks down, however, when changes in $\delta$ and $\sigma$ are driven by variation in $b$, the parameter which determines the convexity of the
5.4. MARKET POWER AND ADJUSTMENT SPEEDS

Table 5.3: Relationships between parameter values in Dockner (1992) model, various quadratic cost parameters

<table>
<thead>
<tr>
<th>$b$</th>
<th>$\delta/k$ (&lt; 0)</th>
<th>$\sigma/k$ (&lt; 0)</th>
<th>$\chi$ (&lt; 0)</th>
<th>$\lambda$ (&gt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>-0.37</td>
<td>-0.12</td>
<td>-0.26</td>
<td>0.49</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.38</td>
<td>-0.12</td>
<td>-0.25</td>
<td>0.50</td>
</tr>
<tr>
<td>0.5</td>
<td>-0.43</td>
<td>-0.11</td>
<td>-0.20</td>
<td>0.54</td>
</tr>
<tr>
<td>1</td>
<td>-0.48</td>
<td>-0.09</td>
<td>-0.16</td>
<td>0.58</td>
</tr>
<tr>
<td>5</td>
<td>-0.78</td>
<td>-0.06</td>
<td>-0.07</td>
<td>0.84</td>
</tr>
</tbody>
</table>

$b \uparrow$  ↓  ↑  ↑ (less competitive)  ↑ (faster adjustment)

firms’ cost functions. Larger $b$ can be interpreted as implying that marginal cost rises more rapidly as output increases. This will tend to encourage more even sharing of the market, since an unequal split will be associated with substantially higher marginal cost for the firm with the greater output. In the context of Dockner’s (1992) model, industries exhibiting this property appear to generate both more collusive outcomes, and faster adjustment to shocks.

Figure 5.1 shows these relationships graphically, plotting $\chi$ against $\lambda$ as each exogenous parameter is varied with the others held constant. More competitive outcomes are to the left, where $\chi$ is most negative (the lower bound for $\chi$ is around $-0.30$, which arises only when $r = b = 0$). As each of the parameters becomes large the outcome converges to $\chi = 0$, identical to the outcome of a one-shot Cournot duopoly.

As shown in Table 5.3 and Figure 5.1, higher $b$ induces a less aggressive reaction to a rival’s output movements in equilibrium ($\sigma/k$ is closer to zero), and this effect also leads to a less competitive steady state (the decrease in $\delta$ reinforces this). It is suggested that the intuition for this less aggressive reaction is that given a decrease in output by a rival which raises the price, a firm is less able to take advantage of this by increasing its own output since it more quickly encounters a higher marginal cost when $b$ is larger. A less competitive equilibrium outcome follows from this less aggressive equilibrium behaviour.

Interestingly, although the effect of a higher discount rate or larger ad-
Figure 5.1: Market power–adjustment speed relationship in Dockner (1992) model, varying parameters $r$, $k$, and $b$. 

![Graph showing the relationship between market power parameter ($\chi$) and adjustment speed parameter ($\lambda$) with varying parameters $r$, $k$, and $b$.](image-url)

- $k \rightarrow 0$
- $b \rightarrow 0$
- $r \rightarrow 0$
- $k \rightarrow \infty$
- $b \rightarrow \infty$
- $r \rightarrow \infty$

- $k = 10$, $b = 0.1$, various $r$
- $r = 0.1$, $b = 0.1$, various $k$
- $r = 0.1$, $k = 10$, various $b$
justment costs in generating less aggressive reactions also resulted in more sluggish adjustment, the effect of greater convexity in costs in inducing less aggressive play is dominated by another effect in determining the adjustment speed. Specifically, the decrease in $\frac{\delta}{k}$ dominates the increase in $\frac{\sigma}{k}$ as $b$ rises.

Although the interpretation of $\delta$ is more difficult than that of $\sigma$, the faster adjustment observed as $b$ rises may be understood as a result of the higher amount of profit lost as a result of a non-optimal output level. If, for instance, a firm’s output is higher than its optimal steady state value, larger $b$ implies a greater reduction in costs from decreasing output towards the optimal level. In the trade-off between avoiding the losses from disequilibrium and incurring the costs of adjustment, the former becomes relatively more important when $b$ rises, thereby encouraging faster adjustment.

With variation in the discount rate this trade-off is not similarly altered, and the effect of the aggressiveness of firm responses on the adjustment speed dominates. With variation in the adjustment cost parameter this trade-off is affected, but the change in the balance of adjustment costs and disequilibrium losses encourages a slower adjustment speed, which only reinforces the effect of reduced aggressiveness of play.

5.4.3 Discussion

Several lessons regarding oligopolistic interactions can be drawn from this analysis. Firstly, a central finding of Dockner’s (1992) paper is worth restating: in environments where firms can be responsive to their rivals’ changes in output – assuming they cannot utilise strategies such as commitment and punishment which tend to aid collusion – market outcomes will be more competitive than when monitoring or reaction are not possible.

Secondly, where larger adjustment costs make such responses more costly, the less aggressive play that results both slows adjustment to shocks and makes the equilibrium outcome less competitive.

Thirdly, greater discounting of the future has similar anti-competitive effects on market outcomes, and encourages slower adjustment to shocks.
When the future profits realised from a rapid response to market conditions diminish in significance relative to the immediate costs of that response, the incentive for aggressive play is reduced with consequently higher markups and slower adjustment.

One way of interpreting variation in the discount rate within dynamic oligopoly models is not necessarily as variation in firms’ preferences across industries, but as variation of the length of the production period inside of which a firm’s output choice is fixed (Cabral 1995, p401). That is, future periods are worth less when $r$ is assumed to be higher because those periods are understood to be further away in time, not because profits in future time are understood to be less valuable vis-à-vis those in the present. An intuitive understanding of why higher $r$ leads to less aggressive responses in Dockner’s (1992) model may be more easily obtained by reference to this interpretation of $r$. Reactions are less aggressive when $r$ is higher because a larger amount of real time elapses before the increased output from a decision to scale up production actually hits the market, and so whatever the rate of time preference the present value of these future sales is diminished relative to the value of current adjustment expenditure.

Finally, where the nature of production technologies cause firms to face sharply increasing marginal costs, the profitability gain from higher market share is lower, and competition will be correspondingly less intense, than where marginal costs are constant. At the same time, because the losses from sub-optimal output choices are higher for all firms, faster adjustment to shocks will occur.

These points suggest that the nature of any relationship between market power and adjustment lags depends on the particular source of market power, and the factors that cause conduct to be uncompetitive. In the model presented here the conditions which disincentivise rapid responses to other firms (higher adjustment costs and longer production periods), resulting in less competitive play, also generally disincentivise quick adjustment to shocks. But one source of anti-competitive behaviour, convex production costs, also makes disequilibrium sufficiently unprofitable relative to adjustment costs that the industry response to shocks will be faster than where production
costs are relatively constant.

More broadly, the analysis above demonstrates a new approach to exploring the relationship between the competitiveness of conduct and the adjustment speed. This approach could be potentially applied to other models of dynamic interactions between firms where the informational environment and strategies available differ, so long as the steady state markup and the rate of adjustment are determined endogenously within each model. For instance, the sticky prices model of Fershtman and Kamien (1987), the alternating moves game of Maskin and Tirole (1987), the strategic investment model of Worthington (1990), and the differentiated goods model of Friedman and Mezzetti (2002) all appear to fit these criteria, and so it may be that insights into the determinants of adjustment speeds can be drawn from further analysis of existing models.

5.5 Conclusion

As described in the introduction, there is a range of plausible but conflicting hypotheses about the effects of concentration and market power on the rate at which prices adjust to shocks to costs or demand. Empirical evidence on the concentration – adjustment speed relationship is also mixed, suggesting that this relationship may be complex and non-monotonic. For these reasons, modelling oligopolistic behaviour in dynamic settings would seem to be a productive way of clarifying the theoretical relationships and disentangling the potentially differing effects of market structure and firm conduct on the adjustment speed.

The use of conjectural variations appears to provide a convenient parameterisation of market power within these models for this purpose. However, as this chapter has argued, there are conceptual problems with conjectural variations that make the choice of modelling technique and the interpretation of the parameter crucial.

This chapter has argued the point, not raised elsewhere, that the way in which Worthington (1989) and Martin (1993) incorporate conjectural variations in their models, as a single-period representation of the expected reac-
tion by other firms to a change in a firm’s quantity of output, suffers from these problems. That is, the conjectures are not constrained to be consistent with (or updated with) actual reactions – the same issue that was the major limitation of the model of Chapter 2 – and there is no indication that the models refer to a specific informational environment that would make this a realistic portrayal of the firm’s expectations.

While transplanting the standard static-model parameterisation of market power to a dynamic setting seems a promising way of modelling the relationship between market power and the rate of price adjustment, the concept of conjectures as beliefs held by the firm through time (without updating) is problematic, just as the literal ‘beliefs’ interpretation of conjectural variations in static models has been replaced by the ‘conjectures as shorthand’ view (at least among theorists). The argument made here is that the conclusions of Worthington and Martin – the latter often cited in support of the idea that competitive markets will adjust more rapidly – are based on methodologically unsound techniques.

The second contribution of the chapter was to suggest a novel approach to achieving the aims of these two papers, drawing on insights from a literature examining the relationship between dynamic models and static conjectures as ‘shorthand’ representations of dynamic model outcomes. The suggested approach relies on the endogeneity of both the adjustment speed and the static conjecture to observe the direction of the changes in the rate of adjustment and market power (as reflected in the price-cost markup) as each structural parameter of the dynamic model is varied. Potentially this method could be applied to a range of dynamic oligopoly models, each differing in their decision variable (prices or quantities), in their ‘friction’ (adjustment costs or sticky prices) and in other aspects; in each case a picture of the relationship between competitiveness and the rate of adjustment as it holds in that model could be observed.

This approach was demonstrated using the quantity-setting oligopoly model of Dockner (1992), and it was shown that variation in adjustment costs and discount rates in this environment caused the competitiveness of steady state outcomes to decrease (or increase) in parallel with slower (or
faster) adjustment. Intuitively, higher adjustment costs and discount rates make aggressive output responses to higher prices more costly, causing both more collusive equilibrium behaviour and slower responses to changes in market conditions. This direction is in fact the opposite of the prediction in Part I of this thesis as well as in Worthington (1989), despite all three models featuring games with quantity-setting firms and convex adjustment costs. The contrast between these results only highlights the importance of basing the modelling technique on solid conceptual ground.

An interesting finding from the analysis of Dockner’s model is that while adjustment costs and the discount rate push competitiveness and the adjustment speed in the same direction, greater convexity in production costs caused competitiveness and the adjustment speed to move in opposite directions. Intuitively, the disequilibrium losses from sub-optimal output choices must outweigh the less aggressive behaviour that also results from a highly-convex cost function, in terms of influencing the speed of adjustment in this model. This finding is significant, as it shows that even within one particular class of model, the nature of the market power – adjustment speed relationship depends on the underlying factors driving the market towards exhibiting competitive or uncompetitive outcomes. It suggests that a deeper understanding of how market power relates to price adjustment will depend on better understanding which are the most important drivers of firms’ output choices and interactions in reality, and how these interactions affect the competitiveness of conduct.
Part III

Taxes on land values: incidence and other topics
Chapter 6

Land-value taxation: Efficiency, development timing, and risk pooling

6.1 Introduction

6.1.1 Background

This part of the thesis considers recurrent taxes on the market value of land, and examines in some detail the well-known argument that such ‘land-value taxes’ are neutral in their effects on resource allocation. As well as discussing the premises of this argument, the current chapter explores some of the policy implications, including the possibility that the tax acts to beneficially ‘pool’ certain risks. Chapter 7 then looks at the efficiency case for the tax in historical context, by examining the views of classical economists and the history of implementation of land taxation in Australia and New Zealand.

As a policy instrument, and as a topic of academic interest, land-value taxation was a marginal rather than mainstream concern throughout much of the 20th century. Over the last decade, however, there has been a small revival of interest in this topic amongst economists and policymakers and in public debate.\(^1\) This has coincided with the increasing attention paid

\(^1\)Work by Coleman and Grimes (2009), for instance, has been influential in recent tax policy debates in New Zealand. Recent theoretical papers have examined the inter-generational efficiency properties of land taxation (Rangel 2005), and how the tax may
to the long-term fiscal challenges facing developed countries, and with the unexpected shock of the global financial crisis of 2008, both of which have prompted a rethink of tax settings and structures.

The first of the major challenges facing tax policymakers stems from globalisation of markets for capital, skilled labour, goods and services. Greater mobility of capital and labour threatens the traditional income tax bases and exacerbates tax-induced distortions, thereby increasing the difficulty of maintaining progressive taxation and reducing the autonomy of governments to substantially deviate from other nations in tax policy. On the expenditure side, globalisation creates social insurance obligations associated with the continuing decline of manufacturing sectors and loss of low-skilled jobs.²

A second long-term challenge is due to population ageing: the demographic transition to an older population in the advanced economies will involve increasing demand for health service and pension provision over the first half of the century. For Australia and New Zealand, small open economies with highly mobile workforces and a reliance on foreign capital to finance investment, these challenges are significant. The combination of internal pressures to increase public expenditure, and external forces which raise the cost of doing so, points towards the need to find new and sustainable sources of tax revenue capable of funding expenditure commitments and obligations, without hampering economic prosperity and thereby destroying that which makes generous social expenditure possible.³

influence foreign investment in open economies (Petrucci 2006). The Liberal Democrats have also brought land-value taxation to the fore in the UK, through their ‘Action for Land Taxation and Economic Reform’ program.

²See Rodrik (1997), McAuley (2009), and Trevitt (2009).

³See, for instance, Whitehead (2009, p6): “Fiscal policy-making will also get increasingly challenging over time. Population aging will increase spending pressures over the next several decades, particularly around health and superannuation. At the same time, on the revenue side economic and technological progress continues to break down national boundaries, bringing many benefits but also increasing the efficiency and productivity costs of raising tax revenues, as activities become increasingly mobile. Policy makers will face increasingly sharp pressures to both increase spending and reduce taxes. That is difficult”. The same arguments feature in the Australian Henry tax review report (Commonwealth of Australia 2009a, p3): “Emerging demographic, health and other pressures on budgets at all levels of government, and expected challenges to our economic circumstances, call into question the durability of the tax and transfer system... increasing globalisation, characterised by increasing mobility of capital and to a lesser extent labour, will have a
6.1. INTRODUCTION

In addition to these long-term trends, the shocks created by the global financial crisis have brought taxation and fiscal policy to the forefront of public debate. The downturn following the financial crisis has had large and unexpected impacts on tax revenues in many developed countries, while policy responses to the downturn (and to financial sector failure in some countries) have also been costly. These events have given questions of taxation settings a new immediacy and relevance in public debate; in the view of the Financial Times commentator Martin Wolf, “managing public finances will govern politics for the foreseeable future... it will be a miserable experience” (Wolf 2011).

For Australia, New Zealand and the UK these challenges have motivated recent reviews of their tax systems, and the intended direction for long-term reform is clear: taxation should be shifted from more to less mobile bases in order to minimise tax-induced distortionary costs, insofar as this is possible given other objectives.⁴ Land constitutes a perfectly immobile base for taxation, and although land taxes raised significant proportions of government revenue in both Australia and New Zealand in the early 20th century, their role today is limited to partial financing of local governments and, in the Australian states, direct taxes on selected high-value (mostly commercially-owned) parcels of land.⁵ As interest grows in broadening the revenue bases of governments, and in replacing the most distortionary forms of taxation currently used, research into the properties and potential of land further profound effect on how investment is taxed”.

⁴Whitehead (2009): “International competition for labour, capital and goods means that our tax system must be in line with other key economies and, ideally, better...we need to shift taxes from bases that are internationally mobile and have the most detrimental impact on growth to those that are less mobile and less damaging to productivity growth”. Henry (2009) expresses similar views: “globalisation means that the things governments tax are becoming increasingly mobile. This has implications for tax system design... increasing numbers of highly skilled individuals are operating in a global labour market. This international mobility will impact on the way that labour is taxed at a national level, let alone at a sub-national level. But there are some inherently immobile tax bases, such as land. The importance of taxing these bases effectively is likely to increase in the future”.

⁵Income from land is also taxed indirectly insofar as it appears in personal income (via rental income), in corporate income (via the portion of company profits which can be attributed to the return on land ownership), and in other indirect ways such as through capital gains tax and stamp duties on real estate sales.
taxes becomes more important.

The basic result regarding the efficiency of land taxes is straightforward, if initially counterintuitive. It is an application of a simple principle of price transmission: where supply is perfectly inelastic, an increase in the cost of supply will not be passed through to prices. Because land is fixed in amount and its services have no cost of production, the income received by landowners is an economic rent rather than a necessary payment for inducing supply. The burden of a tax on land value, which is capitalised land rent, falls entirely on the landowner without being passed forward into higher rental prices. The tax thereby generates no deadweight loss.

In practical terms, land-value taxes have the advantages over taxes on labour and capital incomes of not disincentivising work, employment, savings, or investment, nor inducing people or firms to leave a tax jurisdiction in response to higher rates of the tax. A uniform land-value tax constitutes an unavoidable and non-distortionary tax on wealth held in one particular form at the time that tax is introduced, a characteristic that is the source of both the efficiency advantages and the substantial political obstacles to greater use of the tax.

6.1.2 Outline of Chapter 6

A review of the neutrality argument and exceptions

This chapter seeks to provide two useful contributions to the literature on the properties of land-value taxes. The first is to comprehensively review the current state of economic theory on the neutrality (and specific situations of non-neutrality) of the tax, drawing together disparate threads of work on land taxation. No claims are made of major new insights, but by bringing together the well-known argument about the incidence of the tax – the ‘zero price transmission’ result – with a summary of lesser-known arguments and qualifications to the central argument, and by structuring these in a common framework, the chapter provides a clear and detailed statement of what theory tells us about land-tax neutrality. The aim is to explore the central argument about land tax neutrality more fully, by examining the premises, the
6.1. INTRODUCTION

limitations, and some of the implications of what is an interesting example of a price transmission argument with a clear policy application. The discussion in this review also serves as background for the risk-pooling argument that follows.

The lesser-known arguments discussed here include the effect of exemptions on certain land uses for pass-through, non-neutralities from co-ownership, and the effect of taxing land with depletable-resource characteristics. Also covered are arguments around neutrality with respect to the timing of land development, and the dependence of the neutrality result on land-value assessment methods. These interesting and important adjuncts to the basic argument are typically bypassed in textbook accounts, and in policy discussions. Of the tax reviews recently conducted in Australia, New Zealand, and the United Kingdom, for instance, only in the UK ‘Mirrlees Review’ discussion of land taxation (Institute for Fiscal Studies 2010, chapter 16) is there consideration of the development timing neutrality issue, despite the attention paid to this question in academic journals since the 1980s.

The review, in sections 6.2 to 6.4, also draws some new links between existing results in various parts of the literature, and adds several (minor) points of exposition or clarification. In section 6.2 it is pointed out that the most commonly-encountered argument for neutrality, resting on the concept of returns to land as a rent due to inelastic supply, is not strictly required to establish neutrality of the land-value tax, and unnecessarily confuses what can be a very simple point. To show the lump-sum nature and hence neutrality of the tax it is sufficient to simply ask ‘can a landowner avoid any of the tax by selling their land, or by affecting their land-value assessment?’ Such a presentation also helps frame the potential situations of non-neutrality, which can be seen as due to present-values of expected tax liabilities differing across potential owners, or to flawed assessment methods, or to landowners holding some power to boost demand for their land.

The material in Section 6.3 includes several clarifications stemming from a review by Tideman (1999) of the development-timing neutrality question. One is to explicitly relate the condition identified as necessary for neutrality with respect to development timing – that taxable land value be an estimate
of what the site would fetch if unimproved, without regard to actual development decisions – to the method of valuation used. It is pointed out here that this condition aligns exactly with use of a ‘building-residual’ rather than a ‘land-residual’ method of assessment, and it is also argued that only the former method is conceptually coherent (an issue academic writers on valuation have considered without resolution). This clarifies a point left unresolved in Tideman’s work, namely whether the concept of taxable land value necessary for neutrality actually corresponds to the practices of valuers. Another minor clarification is made regarding a separate hypothesis from Tideman’s chapter, that any timing non-neutrality arising under accurate valuation methods may actually correct a market failure arising from the ‘winner’s curse’. It is pointed out that this logic relies on a particular implicit assumption not made clear in the original exposition, and reasons to consider the assumption realistic are discussed.

**Land tax as risk-pooling**

The second contribution of the chapter is to explore in depth the idea that land-value taxation has ‘risk-pooling’ properties. In a remark within a conference paper, William Vickrey once briefly noted that "one advantage of a high land-value tax is that it would mitigate the impact of zoning changes on the wealth of the affected owners, and at the same time change tax revenues in a way that might act to restrain unduly restrictive zoning adopted for fiscal or exclusionary reasons". The theme of the first part of Vickrey’s comment – that land tax constitutes a risk-pooling device – is substantially expanded on here, by elaborating on the source, nature and consequences of land-value risk, by discussing the difficulty of private responses to that risk, and by examining the effects of land taxes on the distribution of risk.

Vickrey’s point is somewhat self-evident (once stated), but surprisingly no writers have yet explored the idea or drawn out its implications. This is

\[\text{Vickrey (1999 (1992), p16). One other author, Feder (1994, p153), has also noted this point: “[the tax] decreases individual and social risk, since unanticipated gains and losses utterly outside of individual landlords’ control are transferred to the government and pooled with those of the whole community”} \]
6.1. INTRODUCTION

despite the potential for an understanding of the risk-pooling properties to enhance debate about the policy merits of a land tax. Discussion of healthcare and unemployment insurance, for instance, is obviously illuminated by going beyond a superficial description of what these policies ‘do’ (improve the health of the sick, pay benefits to the unemployed) to describe what they ‘are’ (or what they achieve) in a deeper sense – i.e. they pool uninsurable risks. Evaluation of land taxes could be similarly enlightened by an understanding that the tax is not only an efficient revenue-raiser, but that – whatever is done with the revenue – the tax will effectively reduce the private burden of land-value risk and all the negative consequences that go with that, without distorting incentives in any undesirable way.

The points being argued here are a reframing of existing economics, not a new theory about the effects of the tax. They are simple points which can be clearly made without mathematical modelling; they are also self-evident once stated (although that does not diminish the role for stating them in the first place).

In short the argument is that, firstly, land is a risky asset for which diversification and insurance are practically impossible. Privately-borne land-value risk has enormous (under-recognised) costs: the economic consequences alone include speculation, rent-seeking, resource costs of planning disputes, and inefficiently-slow land development.

Secondly, land tax effectively pools land-value risk, by reducing the private value of the risky component of a property asset. The risk-sharing is costless: there are no negative incentive effects (adverse selection or moral hazard), and there are efficiency gains from reducing the costs listed above.

Third, the risk-pooling is a consequence of asset price devaluation, not of distributing the revenues in a particular way. It is thus compatible with a wide range of distributional outcomes depending on the tax and expenditure changes financed with the proceeds; moreover these changes need not be specified in order to understand the risk-pooling property.

Finally, the gains from better allocating risk are a distinct addition to the policy case for land-value tax, only visible when uncertainty is considered. They are not a subset of the efficiency gains from reducing distortionary
taxes, which are present even under certainty. The idea that land tax constitutes risk-pooling is thus both a new perspective from which to understand the consequences of the tax, and an additional relevant point in evaluating its merits.

The overall conclusion is that as well as providing revenue without creating excess burden, land-value taxation improves the allocation of risk in the economy by socialising one risk that serves little to no social purpose when borne privately. That land tax constitutes costless risk-pooling should be recognised, and should sit alongside the basic efficiency property as an important consideration when evaluating land taxes in the context of long-term tax system reform.

The chapter is structured as follows. Section 6.2 presents the basic logic of the incidence and neutrality of a uniform land tax, and discusses how depletable-resource or capital-like properties of land, co-ownership of related sites, or non-uniformity (in the form of exemptions or a progressive rate scale) can upset this incidence result. Section 6.3 discusses the debate over whether the neutrality result extends to neutrality with respect to the timing of redevelopment, and how this depends on the methods of land value assessment used. Section 6.4 discusses the asset price effects of a land tax and argues that the introduction of the tax can be seen as equivalent to the state appropriating an equity share in the value of land. Finally, Section 6.5 presents the argument that land-value taxation constitutes a ‘risk-pooling’ measure, which adds to the efficiency case for the tax.

6.2 Incidence and neutrality with respect to land use

This section lays out the basic efficiency case for land-value taxation. By way of introduction, the argument that the income from land constitutes an economic rent and that a tax on this income is not passed through to the users of land is presented. Next, the argument that a tax on land values is a ‘lump-sum’ and hence non-distortionary tax is discussed. The final subsection explains how exemptions and non-uniformity create the potential for
6.2. INCIDENCE AND NEUTRALITY

Throughout the chapter the income effects of taxes on land rental or land values are not considered. Neutrality here refers solely to the relative price changes and substitution effects induced by the tax. However by altering the distribution of income and wealth amongst households, land taxes might impact on demand and prices in goods markets, if consumption patterns differ across types of households. Such taxes might also affect supply, demand, and prices in factor markets if wealth and income effects lead to changes in consumption–savings decisions and labour–leisure decisions, or to changes in decisions about allocation of savings amongst varieties of risky assets (see Feldstein 1977, Calvo et al. 1979, and Fane 1984).

6.2.1 A tax on the return to land

In the basic argument presented here, the periodic return to land (referred to here as the net return to land, or land income, or similar) is presumed to be determined by the supply of and demand for land in a land rental market. It is shown that the market equilibria before and after the application of a tax on this income are such that the price paid for the use of land is unaffected by the tax, and the full burden of the tax falls on the landowner. It is then explained that a tax on land values has the same non-distortionary property as a tax on land income.

In this chapter the net return to land is deliberately not labelled ‘land rent’. This follows the argument of Tideman (1999, p.110) that the latter should be given a specific interpretation as the opportunity cost of leaving bare land unused (or, what is lost from failing to use the site). Under this interpretation, land rent is also the amount that may be taxed safely without inducing abandonment of the site, since it represents the economic rent available from the site when it is put to its highest-value use. The net return to land in sub-optimal use will be less than this ‘land rent’ but still constitutes an economic rent that may be taxed, if not the highest economic rent the site could provide. Tideman’s definition ensures that land rent is a quantity that is independent of the use to which the land is actually put (independent
of the timing of development, for instance), which is not the case for the net return to land, a quantity which depends on use.

The market for use of land

In any geographically-defined tax jurisdiction, the supply of land is fixed, to a first approximation (land reclamation can increase the quantity of land only to a minor degree). The equilibrium price to rent a unit of land for a period of time is thus given by the intersection of a vertical (perfectly inelastic) supply curve, and the downward-sloping curve representing demand for the use of a unit of land for a period of time. The latter is derived from the demand for the products or services that the land can produce (accommodation services, for example). This price paid for land use, or income received by the landowner, may be observable or imputed. If the former, the price/income refers to the component of a market lease for use of a property that is attributable to land, and if the latter it represents the income foregone by occupying rather than leasing the land.

A tax imposed on this rental income drives a ‘wedge’ between the pre-tax price paid for the use of land and the after-tax income of the landowner. Since the supply curve is perfectly inelastic, the only post-tax equilibrium in which this can occur is where the pre-tax price remains unchanged and the after-tax income of the landowner falls by exactly the amount of the tax. The conclusion from this model is that there is no pass-through of a tax on the return to land into the price paid for the use of land, and the full burden of such a tax falls on the landowner.

In the standard analysis of tax incidence when supply is upward sloping, the wedge created by the tax reduces the quantity traded in the market, raises the price paid by consumers, and lowers the post-tax income received by producers (although not by the full amount of the tax). Since some of the tax is shifted (passed-through) to the consumer price, the welfare burden is shared between producers and consumers. In addition to these effects the tax also creates a deadweight welfare loss or ‘excess burden’, since the quantity supplied to the market has diminished, thus preventing some welfare-improving trade between producers and consumers from taking place.
That this general result extends to land-value taxes is often taken for granted in public commentary. To take one example, in a research note considering a proposed land tax to be applied to all New Zealand properties, economists for a major bank assumed that one-third of the tax would be passed through to rents (National Business Review 2009, Westpac 2009). Similar pass-through claims are frequently made by property-owner lobby groups in Australia, although the non-universality of state land taxes makes such claims more plausible, as discussed further below.

However in the market for use of land, since no supply response can possibly take place there can be no shifting of a tax on rental income to the price paid by consumers of land. That is, for a given demand curve, and assuming the land market remains in equilibrium, the fixity of supply ensures the rental price paid by users remains unaffected by the tax.

A tax on rental income is perfectly efficient: without shifting of the tax to the consumer price, the quantity of land demanded is undiminished, and the amount of land used will remain unchanged. There will therefore be no efficiency cost (excess burden) of the tax; in contrast to any tax for which both supply and demand are less than perfectly inelastic, a tax on land income does not reduce the level of the taxed activity or distort the behaviour of producers and consumers, since relative prices for the use of land and other factors will be unchanged. These properties stem from the immobility and permanence of land. Sites, by definition, are fixed in location and cannot be moved in order to chase higher returns elsewhere, and no investment or lack of investment can increase the amount of land at a given location or cause land to physically depreciate.

**Agricultural land and land-altering improvements**

It should be noted that some aspects of land give it value but do not share this characteristic of being permanently fixed in supply. The ‘locational’ aspect of land (as a site for development which has access to people, markets, public goods, etc) is clearly immutable. But the properties of agricultural land as a productive resource are not, since the fact of soil depletion makes a site’s fertility somewhat dependent on the care and investment of the farmer. Land
in agricultural use thus has both the properties of land as locational space (as with urban land) and of a depletable resource (like timber, water, or minerals).

A tax on the income generated by agricultural land is thus in part a non-distorting tax on an economic rent, and in part a distortionary tax on the return to the farmer’s investment in soil fertility, which acts more like a tax on profit. Taxes on the value of land which include the value added by land-altering improvements such as draining swamps, levelling, or clearing vegetation face the same problem, which is why some jurisdictions (Queensland in Australia for instance) levy taxes on ‘unimproved land value’.

When analyses of the efficiency of land taxes address these types of complications, they sometimes do so by carefully defining ‘land’ as “the permanent qualities of land, exemplified (but not limited to) site” (Gaffney 1994, p46). This is essentially equivalent to Ricardo’s definition of rent as payment for “the use of the original and indestructible powers of the soil” (further discussed in Chapter 7). Taxes on ‘land’ as defined are therefore non-distorting, while taxes on the net return to land – as the word is ordinarily understood – may disincentivise some land-improving activities.

**Understanding the efficiency property**

The conclusion that the return to land can be taxed without welfare loss can be understood from several perspectives. Firstly, as discussed, there is no reduction in the quantity of land services traded and thus no efficiency loss from deterring welfare-enhancing trades. Another perspective is that of excess burdens being a consequence of distortion of behaviour away from that which would prevail in the absence of the tax. From this perspective, the reason there is no welfare loss is that the tax is not passed through to rental prices, meaning the relative prices faced by users of land are unchanged. Demand for land comes from firms for the use of land alongside capital and labour inputs as an input to production, and from households for the use of land alongside other goods and services in consumption. Since the tax is not passed through to the price paid for land use, there will be no substitution away from land by firms or consumers; that is, no distortion to their
6.2. INCIDENCE AND NEUTRALITY

behaviour and therefore no welfare loss.

Another way to understand the logic of the neutrality of a land rent tax is via the concept of the real cost of production. Since the price received by landowners or ‘suppliers of land services’ is reduced by the tax, why is there not an incentive for those suppliers to provide less land to the market, in the same way that an income tax reduces after-tax wages leading to a reduction in labour supplied? The reason is that no real resources are expended by landowners in supplying land to the market. Although real resources are typically used to finance purchase of a land asset, the cost of supplying owned land for use is zero, irrespective of the pattern of ownership.\footnote{Foldvary (2005, p168). Technically, a landowner who supplies land to a formal rental market may incur real costs (a portion of the cost of advertising for and contracting with tenants, for instance). The real costs attributable to the generation of imputed rent in the case of owner-occupancy will be smaller, perhaps limited to the privately-incurred costs of enforcing land title. Note that other periodic expenses typically incurred by landowners (rates, insurance, maintenance, interest payments) constitute taxes, risk-sharing devices, capital investment, and a financing method, none of which are costs necessarily incurred in creating the monetary or imputed income attributable to the land itself.}

That is, there is no opportunity cost to supplying land. Although there is an opportunity cost of employing land in a particular use (the cost being the rental income available in the best alternative use), the opportunity cost of supplying land per se is zero. As Gaffney (1994, p50) has noted, part of the reason for this is that the services of land flow steadily with time, and cannot be brought forward or postponed. When a landowner chooses to withhold land from use, nothing is conserved and no advantage is therefore gained; the act of providing land for use therefore incurs no real cost.

This distinguishes land even from sunk-cost physical capital, from which the present receipt of services comes at the opportunity cost of receiving those services in the future. Even the use of a capital item with no outside value has non-zero opportunity cost at each period in time, since the resource expended in utilising the item is the future use that could have occurred had the item laid idle.

The income received by landowners (strictly, the portion attributable to “the use of the original and indestructible powers of the soil”) is thus entirely an economic rent, since the opportunity cost of providing land services is zero.
Economic rents may be taxed without any distortionary effects, since they are a surplus income which is in excess of that required to induce the income-providing activity to take place. As discussed in Chapter 7, this particular framing of the nature of the return to land – as a surplus, rather than a cost – emerged from Ricardo’s theory of differential rent, and it was central to the classical economists’ view that land rent could be taxed without inhibiting production.

**Heterogeneous land**

The model of land rental and tax incidence described above is subject to one seemingly obvious objection. It treats ‘land’ as a homogeneous factor for which there is a single demand curve and price determined inside a single market for land rental. Clearly, because land inside a tax jurisdiction differs in location, fertility, and amenity, demand for various types of land will in fact be different, meaning that it is unrealistic to refer to a single market for land. However the properties of a tax on rent – zero pass-through, no distortion to household and firm choices, and zero efficiency cost – all carry over to a conceptual model of multiple markets for rental of different types of land. The neutrality of land taxation is thus not an artefact of a simple model which assumes homogeneous land.

The result arises because the most important assumption about the nature of land supply over a whole tax jurisdiction – that supply is fixed because land is immobile – naturally applies also to every geographic unit inside that jurisdiction, right down to a group of homogeneous sites or a single indivisible site. Conceptually, any tax jurisdiction could be broken into a large enough number of areas that the sites in each area are homogeneous, so that a single demand curve exists (or for the case of a single indivisible site, there exists a list of willingnesses-to-pay for the use of the site). The demand curves in each sub-market may exhibit high cross-elasticities (consider residential land in adjacent suburbs, for instance) but because the logic of no tax shifting holds in each sub-market, the fact of this cross-elasticity is unimportant. The incidence of a tax on rent in any one of these sub-markets, or for the jurisdiction as a whole, will be entirely on the relevant landowners.
Note that the no-pass-through result is also not dependent on the land being used in the most profitable manner possible. If a particular site is in a sub-optimal use (one which does not maximise the discounted present value of the stream of net returns after payments to other factors), the demand for use of the site derived from demand for the products or services produced on the land is somewhat less than it could be. However although the net returns to the site in each period will be lower, they will still constitute an economic rent, or a payment over and above the (negligible or zero) amount necessary to keep the site in use.

It was argued above that a tax on land income has no substitution effects, since relative prices for firms and households remain unchanged. The implications of this result for the spatial patterns of land use are straightforwardly seen. Since the land-intensity of firm production and household consumption remain unchanged, and the use prices of each heterogeneous site are unaffected, it is clear that the tax induces no change in land use patterns. The choice of land use that each landowner perceives as maximising the discounted present value of their future net returns from a site will be the same, whether or not there is a tax on these net returns. The type of use (residential, commercial, the type of building, etc) and the intensity of use of each site (the amount of capital and labour inputs applied to the land, which manifests in building height, density of business activity, etc) will be the same after the tax as before. Most notably, this means a tax on land income itself does not reduce the land-intensity of production and consumption activities, so does not lead to higher buildings or more compact cities (although concurrent reductions in taxes on building capital will have this effect; see Brueckner and Kim 2003 and Banzhaf and Lavery 2010).

6.2.2 A tax on the value of land

The reasoning above established that the return to land can be taxed without distortionary changes in prices, since this return is an economic rent resulting from the fixed endowment of land. This reasoning was in terms of a tax on land income, not land values. However the income attributable to land
is generally unobservable, since it is either implicit (in the case of owner-occupiers) or is bundled with payments for the use of buildings. Thus a tax on land values is the more commonly used instrument and the object of policy interest.\footnote{Even those local government jurisdictions that still nominally levy rates on the ‘annual value’ of property frequently use the asset value as the true base; for instance Victorian local governments using ‘net annual value’ as the rating base effectively define this quantity as 5% of the property value (Putland 2008).}

Parts of the literature are somewhat vague about how the neutrality of a land-value tax may be established, and about the definitions of and connections between the return to land, ‘land rent’, and land values. Some writers seamlessly shift between references to the efficiency of taxing rent for public revenue, and the tax on site-values, without explaining the precise relationship between the policy that would be implemented and the income source being taxed (e.g. Foldvary 2005, McCluskey \textit{et al.} 2005). Some argue for the neutrality of a tax on land returns (along the lines of section 6.2.1), before noting that land values are capitalised rent and a tax on the former is equivalent to a tax at a higher rate on the latter (Oates and Schwab 2009, Freebairn 2009, Commonwealth of Australia 2009b). The majority of authors argue neutrality by reference to the inelasticity of supply of land, without explicitly addressing the question of whether land use choices (as distinct from general land availability) might be distorted by the tax.

These common explanations are not wrong, but there is reason to believe they unnecessarily obscure what can be a very simply argument based solely on ‘lump-sum’ reasoning. Reference to the rent of land in explaining the land-value tax is understandable given the historical origins of the arguments (discussed in Chapter 7), but may be one reason misunderstandings about the efficiency of the land-value tax have persisted even while the long-established result about the efficiency of taxes on rent remains beyond doubt. These debates over the neutrality of the tax with respect to development timing are discussed in section 6.3.

Consider some objections that arise from the argument – implicitly made by most writers – that since the tax on periodic returns to land is neutral (as argued in the previous section), the tax on land values (being capitalised
returns to land) must also be neutral. For one, from where exactly does the return to land come? That is, by what mechanism, or in what market, is the periodic return to a site determined? A typical reply (as in 6.2.1) is that this return can be thought of as the equilibrium outcome of a hypothetical market for land use, in which supply is fixed and demand is derived from demand for building use. But how is this demand derived if, for instance, the building on the site is sub-optimal (expensive but lowly valued for rental, say) – is the return to land lesser because the building cost was higher? Would the derived demand and hence the return to an identical site be considered higher if it had a cheaper building that nonetheless fetched the same market rent?

To clear up this confusion, the return to land in this hypothetical market for periodic use is typically supposed to be the return in the ‘highest and best use’ of the site, which is labelled land rent. But the ‘highest and best use’ concept relies on a concept of asset value (the highest and best use being that which maximises the present value of returns from the site, net of the capital and labour costs of developing and using it). Thus to establish the neutrality of the value tax by reference to value as capitalised land rent, where the latter is defined by that which maximises the value, means engaging in a messy circularity with much potential for confusion (which is likely why many accounts blur the argument). At the least this exposition requires precise definitions of a range of concepts – returns to land, land rents, highest and best use, derived demand, the hypothetical market for land use – that substantially complicate the explanation.

An alternative argument, not based on land returns as a rent but on the claim that inelastic land supply means a land-value tax will not cause land to be withdrawn from use, also provokes objections. Without a concept of how land values are determined, how can we be sure the tax liabilities will not induce changes in land use? If the value is not to be explicitly defined as capitalised rent (which would raise the complexities of the argument above), the lump-sum property must be established separately. It is to that logic which attention now turns.

It is suggested here that the clearest way to establish the neutrality of the land-value tax is to use lump-sum reasoning: the tax is neutral – it does not
distort behaviour – if it is a lump-sum from the perspective of the taxpayer (ignoring any income effects, which do not constitute an efficiency loss). Moreover, this reasoning can provide a neat framework for understanding the various situations of non-neutrality.

To show neutrality of the land-value tax it is sufficient to ask two questions. Firstly, can a landowner avoid any of the tax by selling the land? If owners and potential buyers share the same expectations about the present value of the stream of tax liabilities, and buyers capitalise this into their willingness-to-pay, then owners cannot avoid bearing any portion of the future tax stream – the total tax liability is a lump-sum invariant in the decision over whether to retain ownership or sell. In that case neutrality can be argued simply by asking ‘can the landowner affect their land-value assessment?’ The answer (generally) is no: the assessed land value is not under the owner’s control, making it clear that the tax liability is a lump-sum and hence causes no changes in land use, pass-through of the tax into land-use prices, or loss of welfare.

Following this logic, the reasons to expect non-neutrality can be categorised. First, there are those which manifest in different present-day valuations of the expected tax liabilities by different potential owners, causing land to change hands. The effect of exemptions for certain land-uses (as explained in section 6.2.3) is a clear instance of this, but some less-obvious reasons why this might take place are also discussed in section 6.3.2.

Second, if valuers mistakenly assess land value by reference to the particular development in place, rather than by determining the market value the site would have if bare, then the choice of development may influence the tax assessment (without affecting the true land value). This possibility is discussed in section 6.3.1.

Finally, there are specific situations where the landowner’s actions as an individual can affect demand for use of their land, and hence alter the market value. These include investing capital in land-altering improvements or soil quality (discussed earlier), co-ownership of related sites where the choice of use of one alters demand for another (discussed next), and rent-seeking behaviours – exerting control over policy choices, with respect to public goods
6.2. INCIDENCE AND NEUTRALITY

provision and land-use regulation in particular (discussed in section 6.5).

Appealing to the lump-sum property of the tax, it is argued here, is the clearest way of arguing the neutrality of the tax on land values. As that result was in fact re-opened for debate over a twenty-odd year period, due to confusion over the relationships between land returns and land values and what this meant for development timing (as discussed in section 6.3), the importance of clear and rigorous explanations is evident. It has been argued here that the property of land returns as a rent is not necessary to establish the neutrality result (and identify situations where it fails), even though that property can obviously help us understand why the result arises. Since neutrality is about distortions to behaviour, it is simpler to address the issue head-on by considering what scope landowners have to modify their tax burden by changing their behaviour. That is, for expositional purposes the two questions suggested above and the three categories of non-neutrality provide the most straightforward way to explain the general neutrality and specific sources of non-neutrality of the tax.

The market value of land, and the co-ownership situation

In the general case, what determines the value of land, and why is it beyond the owner’s control? Under basic assumptions, the market value of an undeveloped site will be equal to the discounted present value of the net returns to the site generated when the site is developed in the most profitable way possible. Absent any of the sources of non-neutrality discussed, this amount is clearly independent of the actual use the land is put to by the owner, meaning the tax is a lump sum and hence will not alter that owner’s land use decision. As Vickrey (1999 (1992), p15) points out, the lump-sum property can also be achieved with other tax bases which are not exactly correlated with land value (the area or street-frontage of a site are examples of such bases).

For developed sites, a tax on the market value that the site would have if undeveloped has the same lump-sum property. Thus, so long as land value is defined and assessed as being the market value that the site would have if bare, a tax on land value is a lump-sum from the owner’s perspective and
is thus non-distortionary, whether or not the land being taxed is developed. Only if the tax is based on concepts of site value other than this might distortions arise, as discussed further in Section 6.3.1.

Factors affecting the land value include incomes, population, tastes, technology, transport costs, economic conditions, zoning regulations, and provision of local public goods, as well as the nature and use of nearby properties which impact upon the amenity of a site (i.e. externalities). None of these factors are influenced by the way a landowner develops and uses the site in question.9

One exception to this rule is where improvements or building on a particular site enhances the rent and value of nearby sites under common ownership, as for privately planned and developed communities for instance. A similar effect arises for shopping malls composed of separate sites under single ownership where the attraction of an anchor tenant raises the value of the other sites. In these cases, taxing properties under common ownership as if owned separately has some efficiency cost, since it reduces the incentive for the owner to conduct improvements on or choose particular uses of their sites which have positive impacts on the value of the other sites. As Vickrey (1999 (1992), p15) has noted, this occurs because common ownership internalises the externalities associated with land use, and the disincentive to improve a site that land taxation creates in this situation is actually the counterpart of the lack of incentive to undertake activities that create positive externalities, or reduce negative externalities, in the usual situation where ownership is not common.10

9Although the policy-determined aspects of demand may be influenced by landowners if they have some control over policy settings, a fact which encourages rent-seeking as discussed in Section 6.5.

10Since discrimination in tax based on ownership type would cause undesirable distortions and wasteful tax-avoidance behaviour, Vickrey argues, the appropriate policy solution to this common-ownership problem is to independently tax or subsidise the improvements and activities that create the negative and positive externalities. To the extent this is possible, it has the advantage of redressing not only the tax-induced inefficiency but also the general market failure associated with unpriced externalities.
6.2. INCIDENCE AND NEUTRALITY

6.2.3 Exemptions and tax pass-through

While the characterisation of the return to land as an economic rent – a surplus above that necessary to induce the supply and use of the land – may not be strictly necessary to understand the neutrality of a land-value tax, it helps clarify the distortions that arise from selectively taxing particular uses of land. With selective taxation, the after-tax return necessary to keep land in its current use is not zero, but is the rent receivable from the best non-taxable use of the land. Only the difference between the rent of a site in its most profitable taxable use and the rent in the most profitable tax-exempt use may be safely taxed without distorting behaviour by altering the use of the site.

Blaug (1985, p82) puts the principle this way: "the rent that land could earn in one use forms a cost that must be paid when it is used for some other purpose". This means that the level of ‘Paretian’ rent (excess earnings over that required to keep a site in its current use) represents the base for a selective tax on the rent of land in that particular use, while the level of ‘classical’ rent (the entire return to a site) is the base for a non-discriminatory tax on the rent of land in any use.

In Australia around 70 percent of land by value is carved out of the base for state government land-value taxes by virtue of the two largest exemptions, owner-occupied housing and primary production land, with a range of minor uses from child care to caravan parks typically exempted as well. Land holdings employed in non-exempt uses are also subject to a progressive rate scale, with a landholder’s aggregated value of landholdings subject to a tax-free threshold in most states, with higher tax rates as the total value of land held rises.

Such discriminatory application could naturally be expected to distort land ownership and use patterns in favour of the more lightly-taxed owners and uses, although there is a lack of empirical evidence indicating the magnitude of these distortions. In terms of the characterisation of rent above, the imputed rent that a house provides an owner-occupier, for instance, forms a

---

cost that a landlord must cover from after-tax rental income from tenants if the house is to remain in rental market use. A land tax with owner-occupier exemption may reduce a landlord’s net return, and hence the property’s value as a rental investment, below that which it is worth to potential owner-occupiers, resulting in a change in ownership and use. Similarly, the potential after-tax return to a site for a small-business landowner may be sufficiently close to the after-tax return to a large corporation, given the escalating rates on total landholdings, so as to cause a change in ownership in favour of the smaller firm despite the larger firm’s use of the site being the more profitable in the absence of tax.

The existence of exemptions and a progressive rate scale may therefore alter the ‘zero tax pass-through’ result discussed above. The 2010 Australian tax review report argues that "the portion of tax that is not reflected in lower land prices is borne by investors through lower returns, or by their renters through higher rent... in the long run, much of the burden of the tax is shifted to renters, as rents adjust to ensure that investors achieve an adequate return" (Commonwealth of Australia 2009b, p261). The final incidence and burden of land-value taxes applied with such exemptions is clearly an issue of policy importance, since it points not only to the efficiency cost of the tax induced by changes in land use away from the most profitable uses, but to the distributional impacts of the tax as well.

6.3 Land-value tax and the timing of development

6.3.1 Development timing under certainty

In the previous section, it was argued that the lump-sum property prevented a land-value tax from distorting the decisions of landowners. One such land-use decision concerns the timing of development of land. Since structures are durable, the type and size of structures cannot be varied from year to year and landowners face an ongoing choice between immediate development (or re-development) and delay, with the optimal type and intensity of de-
development perhaps differing across time periods. An issue debated in some
detail, beginning with Shoup (1970) and summarised in Tideman (1999) and
Oates and Schwab (2009), is whether the timing of development of land is
brought forward as a result of the presence of a land-value tax, relative to
the timing that would occur in the absence of the tax. This would constitute
a distortion to behaviour that may result in a loss of welfare, all else being
equal.

Prima facie, the argument that land values are independent of landowner
actions and hence that a land-value tax is a non-distorting lump-sum tax
seems to preclude this. However the arguments of Bentick (1979), Mills
(1981), Wildasin (1982) and Tideman (1982) have established that this con-
clusion is in fact sensitive to the method of measurement of land values.

When the land value assessment for tax purposes is an accurate estimate
of the market value that the bare site would fetch in a sale – reflecting the
highest-rent use that the site could be put to, with the timing of development
one of the dimensions of this use – then a land-value tax truly is a lump-
sum tax which does not distort the timing of development. However when
the value for tax purposes is instead based on the discounted rental income
of the landholder from the particular development project chosen (which
may not be the highest-rent use of the site), this is not the case. With
the latter tax, landowners have the ability to reduce their discounted tax
liabilities by choosing projects with more immediate payoffs over projects
which delay development but receive higher returns in the future, even where
the discounted rental income of each project is identical in the absence of the
tax. Thus the timing of development would tend to be brought forward
relative to the tax-free case.

By way of simple example, consider land which is used for only two periods
in one of two uses, with second-period returns discounted at a rate of 0.5.
Use A yields an income of 1 in the first period and 0 in the second, and use
B yields 0 in the first period and 2 in the second. The discounted returns to
the two uses in the absence of taxation are identical. However in the second
period, development B retains some value since there is income yet to be
received. A land-value tax, with value assessed as discounted rental income
from the particular use chosen, would thus generate a tax liability in both periods for land in use B, but only in the first period for land in use A. Use A would become the preferred use despite use B being equally profitable in the absence of taxation.

However a distorting tax as described is not actually a tax on ‘land value’ (the market value of a site if it were bare) but, as argued by Tideman (1999), is a tax on the ‘present value of planned net income (PVPNI)’. Such a tax in practice would involve a land value assessor estimating the value of a developed site according not to what it would fetch were it free of buildings, but to the net present value of the future returns from the particular development project in place, after attributing a portion of these returns to the improvements. It would also (counter-intuitively) require land assessors to assess the value of an undeveloped site not according to its market value, but with regard to the particular plans of the owner for its later development.

For this reason, Tideman (1999, p117) argues that such a tax is simply "a theoretical construct that does not correspond to any real or plausibly potentially real tax". The conclusion is that a land-value tax based on an unbiased assessment of the price that a site, if unimproved, would receive upon sale is in fact neutral with respect to development timing, neither bringing forward nor delaying development by virtue of the tax being independent of the use the site is actually put to.

In the example above, a correctly-applied land-value tax would impose a liability in period 2 based on what the site would sell for at that time if bare. Landowners in period 1, anticipating this, would remain indifferent between use A and use B.

The debate over development timing has generated much (misplaced) doubt over the status of the neutrality result, the relevant concepts of land values, and the implications for policy. To take one (important) example, the Mirrlees report suggests that "[a land-value tax] can create a bias towards excessively rapid development... This bias could be avoided by taxing the ‘best use’ value of the land rather than its market value" (Institute for Fiscal Studies 2010, p372). This claim is problematic on several grounds: if market value means what the market would be willing to pay for the site if bare, the
logic of section 6.2.2 means the tax will (contra this statement) be neutral. Moreover, it is not clear how a ‘best use’ value could ever be objectively determined other than from evidence of what the market would be willing to pay.

**Conceptual methods for valuing land and improvements**

This logic naturally prompts the question of which methods of land appraisal actually generate unbiased estimates of what unimproved sites would fetch in the marketplace, and whether the commonly used methods of assessment are at all influenced by the current use of the land, rather than reflecting only the hypothetical market value if bare.

Although Tideman (1999) claims that the PVPNI tax is a theoretical construct, he also agrees that:

"It is possible that assessors, who are supposed to assess land according to the value it would have if unimproved, are instead influenced by the development decisions that owners of land make, and if such a bias in assessments is predictable, it motivates inefficient investment decisions." (p121)

Hudson (2001), addressing a separate topic (the accuracy of statistical estimates of US land price movements), describes two methods of decomposing land and improvements values from the joint value of a developed property. The ‘building-residual’ approach estimates site values using actual sale prices of nearby bare land (and sales of property where the buildings are subsequently demolished), and attributes the difference between the assessed property value and the estimated land value to the building. Conversely, a ‘land-residual’ approach (also known as the ‘residual site value’ method) subtracts from the property value an estimate of building value based on construction costs (replacement cost or historic cost, usually reduced to reflect the depreciated state of the building), leaving the land value as a residual. The same methods are described by Hendriks (2005), in the context of a debate amongst writers on valuation practice over whether the market value of a property can or should be apportioned amongst its components, and
whether depreciated replacement cost methods are suitable for this task (see also Wyatt 2009 and Boyd and Boyd 2012).

Given the durability of buildings, the value of a site developed in a sub-optimal way (that is, with a use which does not maximise the discounted present value of pre-tax rent of the site) may be underestimated where the land-residual approach is used, making a tax levied on such estimates more like a tax on PVPNI. Consider a property which, if the site was in optimal (highest rent) use with a particular building costing $100,000, would be valued at $200,000; that is, it would generate discounted returns to the property owner of this amount. If the site were (hypothetically) sold bare it will tend to trade at a price of $100,000, which represents the highest discounted rent achievable from this site. Suppose the site is currently sub-optimally developed with a different building also costing $100,000. Suppose also that because the site is not in its highest rent use it generates discounted returns of only $150,000 (meaning the developed property trades for this amount). Using the land-residual method, the value of the site in sub-optimal use is estimated at $50,000 rather than the $100,000 which is the market value of the undeveloped land. A land tax based on this value assessment would tax owners of sub-optimally used land a smaller proportion of the true land value than owners of optimally used land.

Where buildings are still far from the end of their lifespan, so that the property value largely reflects the returns received in the current mode of use of the site, the difference between residual site value and land value (i.e. the market value of bare land) may be quite large. This implies that a land tax with liabilities based on residual site value is closer to Tideman’s PVPNI tax than to a true land-value tax. If it is predictable that tax liabilities will be determined this way, there will be a distortion to development timing as described, whereby projects with early payoffs are favoured over those which delay construction or which otherwise feature a payoff stream loaded towards the future.

12 Conversely the value of a property with a ‘tear-down’ building will largely reflect the land value, which in turn reflects the potential rent available from the site if developed optimally. The residual site value in this situation will be close to the land value (market value of the land if bare).
6.3. THE TIMING OF DEVELOPMENT

Land valuation methods and implications for neutrality

The debate over theory – whether the land-value tax is neutral with respect to
development timing – appears to have reached a straightforward conclusion:

"[The claim] that taxes on the value of land distort land de-
velopment decisions by advancing the time of development ineffi-
ciently... rests sometimes on logical errors and sometimes on
analysis of taxes that are not taxes on the value of land" (Tide-

However an answer to the practical question of whether such taxes – that
‘are not taxes on the value of land’ – may in fact be erroneously implemented
in practice is less clear. Whether that is the case evidently depends on
the method of land assessment used, and in particular whether the current
use influences the assessor’s appraisal of land value. It has been argued
here that of the two conceptual methods of land valuation only use of the
‘building-residual’ method will ensure that land valuations and hence tax
assessments are invariant to the use of the site, while the ‘land-residual’
method of valuation will result in a land tax actually resembling Tideman’s
PVPNI tax.

This prompts the question of whether there are grounds for regarding the
building-residual method as more conceptually sound than the land-residual
method. With the land-residual method, structures fixed to a particular lo-
cation must be attributed value on the basis of construction costs, since their
value is not observable in a marketplace (by their nature – fixed – the struc-
tures cannot be removed and traded). However it can be argued that such
attribution misses a crucial distinction between the cost of production and
the value in use; although profit motives will generally drive these quantities
together for new buildings, sub-optimal development decisions or unforeseen
changing circumstances over the long life of a building mean that equality of
cost and value is not guaranteed.

In this view, land can have value independent of structures that are on
it (that value, being equal to the discounted rent in its best use, is what
it will usually sell for when bare), but structures cannot be attributed any value – only a cost – independently from their location. The implication is that to decompose the value of a land-structure combination meaningfully, the land value must be estimated first with the value of those structures in that location being given as a residual.

Another way of arguing this point is to note that to ask what a site would sell for if bare – and to define site value accordingly – is not a hypothetical question, since that state can be achieved. In contrast, a definition of the value of fixed structures based on what they would fetch if not fixed (a quantity which in practice is approximated by construction costs) is one based on an entirely hypothetical question. The only sensible definition of structure value, therefore, is that of its value in its current location; that is, what the structure adds to an independently-determined value of the site it occupies.

The intuition for this result also becomes obvious with simple examples: consider a million-dollar hotel, built on zero-value desert land where its inaccessibility makes the land-structure package worthless. Under a land-residual valuation the structure value, based on construction costs, is $1m, while the previously-worthless land is assessed as contributing negative $1m to the value of the joint package.

The point is expressed by Tideman (1999) as follows:

"Any attribution of value to improvements that cannot be economically removed from a site, other than what the improvements add to a previously determined value of the site, would be inappropriate. If, as in the real world, unanticipated changing circumstances generate fluctuations in the value of the combination of land and improvements, any independent attribution of value to improvements would be even more dissociated from economic meaning". (p117)

The implication for valuation methods of the argument above is that only the ‘building-residual’ method is conceptually sound. The connection being drawn here is to take Tideman’s condition for a meaningful valuation
6.3. THE TIMING OF DEVELOPMENT

of land (one which is also sufficient to ensure neutrality with respect to
development timing) and to match it explicitly with a particular method
of land appraisal. Assessing improvements values first on a cost basis is
conceptually faulty, confusing the concepts of cost and value. Moreover a
land tax based on such ‘land-residual’ assessments may instead resemble
Tideman’s PVPNI tax, with potential non-neutrality in the effect of the tax
on the optimal timing of development. For both conceptually coherent land
valuation, and for neutrality of the tax in practice, there is a clear implication
for valuation methods: only the ‘building-residual’ method clearly satisfies
both criteria.

A paper on this issue by Arnott (1998) characterises the assessment prob-
lem as a tradeoff between the neutrality of a true land-value tax and the ad-
ministrative feasibility of a residual site value tax, seemingly without recogni-
tion of the feasibility of accurate building-residual calculations based on sales
of nearby bare land and ‘tear-down’ properties. He asserts that the defini-
tion of land value as ‘what the site would be worth if there were no structure
on it’ has the disadvantage of not being able to be "simply calculated or
inferred on the basis of market observables", and demonstrates instead that
a neutral tax can be approximated with taxes at different rates on each of
three more easily observed values: pre-development bare land value, post-
development residual site value, and structure value. However, the method
proposed seems to be subject to the same key methodological objection that
makes the land-residual method of valuation problematic, that attribution
of value to structures based on construction costs ignores the distinction be-
tween the cost of production and the value in use, which may differ when the
choice of long-lived structure is a suboptimal use of the land, or becomes so
over time.

6.3.2 Development timing with uncertainty

Oates and Schwab (2009, p60) point out that there is an apparent inconsis-
tency between the neutrality of land-value taxes with respect to development
timing, and the often-quoted claim that by introducing a ‘holding cost’ that
must be paid irrespective of land use, such taxes will ‘force the hand’ of idle landowners and speculators and draw more land into productive use or development. The latter of these seemingly mutually exclusive results was a key plank of Henry George’s arguments in favour of land taxes, but the argument that a tax on properly-assessed land values is a lump-sum from the owner’s perspective appears to preclude this outcome.

Two separate arguments provide reasons why, although land taxes may be neutral in the simplest theories, in practice they may force earlier development of land than would otherwise occur. The first relies on sub-optimal behaviour on the part of landowners, and the different effects on the wealth and liquidity of landowners of a tax liability which must be paid in cash versus an equal value of foregone interest. Although the optimal use of land might be the same after the imposition of a land tax as before, parcels of land will likely be more quickly developed both because the owners face a cash drain which would require credit to continue financing if they continue to hold off development, and because “[this] cash drain... conveys information to many owners who are only vaguely aware that they are holding a resource of high salvage value to society” (Gaffney 1999, p78). Landowners perceive their holding costs via taxation as much higher than the value of the interest they forgo from underuse of a site, and consequently develop or sell the land to willing developers more quickly than they would in the absence of the tax.

A second argument which supports George’s claim, as explained by Tideman (1999), begins with the (realistic) assumption that the path of future rents of a site is not perfectly known. Different potential owners will form different expectations about the materialisation of future demand for the site, and hence about the most appropriate (profit-maximising) timing and type of development.

The nature of markets for sale of land mean that those investors with the most optimistic expectations about future demand assign the site the highest value in the present, and will thus be willing to bid highest. In the calculus of these (most bullish) investors, Tideman goes on to argue, the optimal use of the site will typically be to hold the land undeveloped or in low-intensity use until it is known with more certainty whether demand which is higher
than the rest of the market expects will materialise. If the bullish investor’s projection is correct, a larger development can be undertaken in later periods than that which would have been committed immediately to the site by the less-bullish investors who were earlier outbid.

However a land-value tax may in fact rearrange the rankings of these initial bids for land, by making the earlier and less-intensive development plans relatively more profitable to those who expect low future demand than the later more-intensive plans are for those that anticipate high future demand. The reason is that, in the expected profit calculation of bullish investors, the future land value and associated tax liability will be large, since high future rents are available. However in the calculation of conservative investors, low future demand implies a low future land value, so the present day impact of a commensurately lower tax liability on this investor’s bid is small. The lower bid in the absence of the tax becomes the higher in the presence of the tax, and development takes place sooner.

The claim that land-value tax can ‘re-rank’ bids for land that correspond to different land uses seems at first to conflict with the argument in the previous subsection that the tax, with land values accurately assessed, is neutral. That argument was based on properly defining land value to be use-invariant, and on future land values being known with certainty (or, at a minimum, on all agents’ expectations about future land values being equal). A land-value tax capitalised into present-day valuations of alternative uses thus caused these valuations to all fall by the same amount, preserving the ranking of profitability of uses. However under conditions of uncertainty, the present-day impacts of the tax will differ for individuals with differing expectations of future land values. Where those individuals favour different land uses, it is the re-ranking of alternative uses planned by different individuals that induces a change in land use favouring faster development, not the re-ranking of alternative uses for the same individual (Tideman 1999, p126 provides a numerical example of this).

The theory presented in the section above, that characterised land-value tax as a lump-sum with no distortion to the optimal timing of development, can thus be reconciled with the Georgist idea that land-value tax forces land
CHAPTER 6. LAND-VALUE TAXATION

held for speculation into productive use. It is the addition of uncertainty about future rents to the basic theory, coupled with the idea that optimistic investors are also those that favour delayed development, that lends theoretical support to the claim that land-value tax encourages earlier development.\textsuperscript{13}

**Alternative expectations and preferred development timing**

The key premise in Tideman’s argument is clearly the claim that there is a correlation across different bidders between the optimism of their expectations about future demand and their preferences for delayed development. Tideman provides little support for this claim, but there are three reasons suggested here to explain why such a correlation may arise in reality.

Firstly, in a rising land market the potential buyers of land may be of two types: genuine property developers, whose comparative advantage lies in producing appropriate buildings to realise the highest income from capital investment, and professional speculators whose skills lie not in building development, but in the identification of sites most likely to rise in value. The latter group are searching for income from land price appreciation, and their optimal business strategy involves purchasing as many sites as possible given their equity base and credit constraints, rather than tying up capital in building development. If the second group (the pure speculators) are more optimistic about future demand than the first group (the developers), then development will be slowed.

Secondly, certain bidders may hold optimistic expectations about future demand because their specialisation is in enhancing demand for sites through expertise in manipulating planning systems, in order to obtain permission for more intensive development than is currently permitted. Relative to the other bidders who take planning constraints as given, bidders who are optimistic about their prospects for altering planning constraints will naturally prefer

\textsuperscript{13}In case this extension of the basic theory seems arbitrary, or contrived, a later paper by Tideman argues: "Economic theory is sometimes developed under an assumption of perfect foresight. But if everyone had perfect foresight, there would be no possibility of gain from any form of speculation. So if speculation does occur, a theory of speculation should not be evaluated in terms of a theory that incorporates an assumption of perfect foresight." (Tideman 2004, p1094)
6.3. THE TIMING OF DEVELOPMENT

not to build until after achieving success in relaxing these constraints, which takes time.

Thirdly, it is likely that in many situations of significant uncertainty about future demand, the optimal strategy for bidders with slightly-more optimistic expectations than the rest of the market is to ‘wait-and-see’, rather than to immediately develop at a higher intensity than that which would have occurred had the less-optimistic bidders acquired the site. This is because of ‘lumpiness’ in the spectrum of site uses; between the 2-storey building that is optimal for low demand and the 3-storey building that is optimal for high demand, there exists no option of a ‘2.5 storey building’ that an investor who places a slightly higher subjective probability on the high-demand outcome may choose. Instead, the optimum for this investor may be to wait, before observing demand and selecting between options.

Welfare implications

The welfare implications of these two effects – whereby land ends up in the hands of those with the most extreme (optimistic) beliefs about the future, and whereby the tax acts to redistribute land away from these investors towards more conservative ones – are somewhat uncertain. In some instances the optimistic investor will be correct and the withholding of land (‘speculation’) is in fact socially optimal, since the investor has prevented inefficient and wasteful early commitment to a land use which proves inappropriate for the site. This is what leads Blaug (1985, p85) to say that "land speculation performs an economic function: people differ in their expectations of the future economic development of particular locations, and the profits of those who have forecast correctly are, of course, matched by the losses of those who have not".

However, as Tideman (1999, 127) points out, this effect could also be associated with a ‘winner’s curse’: the winning bid will be made by the investor with the most extremely optimistic (and incorrect) expectations. With the assumption that the investors optimistic about future rents are also those that favour delayed development, this winner’s curse will also result in a socially inefficient outcome – a market failure – since in aggregate there will
be an artificial scarcity of land for current development; a scarcity greater than that justified by the need to prevent inefficiently early development.

Since the land-value tax reduces the present-day expected value of a site to both ‘useful’ speculators (who correctly foresee opportunities missed by others) and ‘wasteful’ speculators (whose extreme beliefs are incorrect and whose withholding of land is socially inefficient), it may correct a market failure in some instances and create welfare losses in others. The question of how to devise institutions to neither under nor over-reward investors for forestalling premature development remains largely open, although Tideman (1999, p129) suggests the promotion of markets whereby "those with a skill for identifying future opportunities can sell insurance to those who want to use land now against the possibility that future increases in rent will reduce the value of their capital".

6.4 Asset price effects of land-value taxes

As background for the argument that follows, this section presents several simple expressions that describe the impact of the introduction of a land-value tax. These expressions relate the tax rate, discount rate, and expected growth rate of rents to the value and income from a land asset in the absence and in the presence of the tax, and to the stream of tax revenues and their present value to the government. These expressions can be found in various forms in Oates and Schwab (2009), Gaffney (2009), Coleman and Grimes (2009), and elsewhere.

They show that the impact of a land-value tax can be summarised with a simple analogy – the introduction of the tax is economically equivalent to the government appropriating an equity share in the value of the land asset. That is, the tax revenue received by the government, and the present value of the stream of these revenues, exactly match the reduction in the landowner’s net-of-tax income and the fall in the private asset value, just as if an ownership share of the asset with a corresponding dividend stream had been transferred to the state. This analogy is reasonably self-evident, and similar statements have been made by other authors (e.g. Gaffney 2009,
6.4. ASSET PRICE EFFECTS

Several assumptions are used here for simplicity in demonstrating the key relationships. First, the expected growth in rents $A_t$ is assumed to be at a constant rate $g$ for all time periods $t$. Second, the market value of a land asset – the base for the tax – is assumed to equal the discounted present value of all future after-tax rents over an infinite horizon; this is a natural reflection of the life of the asset, not an approximation to the optimisation horizon of its owners.\textsuperscript{14} Third, a single discount rate $r$ represents the interest rates, which would in reality differ, on both borrowing and saving for landowners as well as for the government (a simplification which avoids the asset value being dependent on the amount borrowed for its purchase). Finally, the timing of valuation and cash payments is as follows: the asset value assessed at the beginning of period $t$ includes the rent and taxes for that period, which are received or paid at the end of the period (hence are discounted in the net present value calculation).

6.4.1 Pre-tax asset values and income

The expression for the period-$t$ asset value, given the above assumptions, is:

\[ V_t = \sum_{s=0}^{\infty} \frac{A_{t+s}}{(1 + r)^{s+t}} \]  \hspace{1cm} (6.1)

Given that $A_{t+s}$ grows at growth rate $g$, the expression can be converted to

\[ V_t = \sum_{s=0}^{\infty} \frac{A_t(1 + g)^s}{(1 + r)^{s+t}} \]  \hspace{1cm} (6.2)

This can be solved, using the expression for an infinite sum, to show that:

\[ V_t = \frac{A_t}{r - g} \]  \hspace{1cm} (6.3)

Equivalently, this relationship can be derived using the ‘fundamental equation of yield’ and ‘guessing’ at an expression for $V_t$. In this equation, the

\textsuperscript{14}“Even those who preach that the end-time is near buy and sell land at the market” (Gaffney 2009, p348).
value $V_{t+1}$ must be such that an investor is indifferent between investing $V_t$ in a non-land asset at rate $r$, or purchasing the asset at $V_t$, receiving income $A_t$, and holding an asset worth $V_{t+1}$ at the end of the period:

$$V_t (1 + r) = A_t + V_{t+1} \quad (6.4)$$

It can be shown that equation (6.3) is the only expression that satisfies (6.4) for all $t$.

There are two components to the comprehensive (or Haig-Simons) income received by the owner of the asset: the current income $A_t$, and the growth in asset value $V_{t+1} - V_t$ (the ‘capital’ gain). Asset values as given in (6.3) grow at the same rate as rents, $g$, so that the latter is given by:

$$\frac{gA_t}{r - g} \quad (6.5)$$

The asset value is commonly described in terms of two quantities: the income it generates, and a ‘capitalisation rate’. Given the two components of income, there are two ways the land value can be expressed: as the current income $A_t$ divided by the capitalisation rate $r - g$, or as the comprehensive income $A_t + gV_t$ over the capitalisation rate $r$. The latter expression highlights the fact that unrealised ‘capital’ gain and current income are equivalent in terms of how they are capitalised into the value of an asset.

### 6.4.2 Post-tax asset values, incomes and tax receipts

With a land-value tax at rate $\tau$, the discounted present value expression for the post-tax asset value $V_t'$ is:

$$V_t' = \sum_{s=0}^{\infty} \frac{A_{t+s} - \tau V_{t+s}'}{(1 + r)^{s+1}} \quad (6.6)$$

This is not readily solvable using the infinite sum expression due to the right-hand-side presence of the $V_{t+s}'$ term. More straightforward is to guess at an expression that satisfies the post-tax fundamental equation of yield:

$$V_t' (1 + r) = A_t - \tau V_t' + V_{t+1}' \quad (6.7)$$
6.4. ASSET PRICE EFFECTS

It can be shown that this equation holds across all $t$ when $V'_t$ is given by the expression:

$$V'_t = \frac{A_t}{r + \tau - g} \quad (6.8)$$

The tax liabilities in each period $t$ are thus given by:

$$\tau V'_t = \left( \frac{\tau}{r + \tau - g} \right) A_t \quad (6.9)$$

This leaves the landowner with a net-of-tax current income of:

$$\left( \frac{r - g}{r + \tau - g} \right) A_t \quad (6.10)$$

The present-valued sum of this net-of-tax income, over an infinite horizon, is the post-tax private asset value shown in equation (6.8). The net present value of the tax receipts is equal to:

$$\left( \frac{\tau}{r + \tau - g} \right) \frac{A_t}{r - g} \quad (6.11)$$

The difference between the pre-tax asset value in (6.3) and the post-tax value in (6.8) can be seen to equal the present value of the stream of the tax receipts in (6.11). This equivalence underlies the claim that the tax payments, current and future, are ‘capitalised’ into the asset value, as the latter falls by the present value of the former upon introduction of the tax.

Because the private asset value initially falls, but continues to grow thereafter at the rate of growth of rents $g$, the second component of the landowner’s comprehensive income – the growth in asset value or ‘capital’ gain – is also lower in absolute terms after introduction of the tax. This gain, initially given by (6.5), is now given by:

$$\frac{g A_t}{r + \tau - g} \quad (6.12)$$

6.4.3 The government’s ‘equity share’

The relationships between incomes and assets, private and public, can be summarised in the following expressions which support the claim that introducing a land-value tax is akin to the state appropriating an ‘equity share’ in land.
The first expression elates (6.3), (6.8) and (6.11), and shows the split of pre-tax land value between the landowner, in terms of post-tax private asset value, and the government, in terms of the present value of tax receipts:

\[
\frac{A_t}{r - g} = \frac{A_t}{r + \tau - g} + \left( \frac{\tau}{r + \tau - g} \right) \frac{A_t}{r - g}
\]

or

\[
\frac{A_t}{r - g} = \frac{A_t}{r + \tau - g} + \left( \frac{\tau}{r + \tau - g} \right) \frac{A_t}{r - g}, \quad \text{or}
\]

\[
\frac{A_t}{r - g} = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] \frac{A_t}{r - g}, \quad \text{or}
\]

The second expression shows that the current income of the land asset, \(A_t\), is split between the landowner’s net-of-tax income given in (6.10) and the tax payment given in (6.9), in the same proportions as in the expression above:

\[
A_t = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] A_t, \quad \text{or}
\]

\[
\frac{A_t}{r - g} = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] \frac{A_t}{r - g}, \quad \text{or}
\]

\[
\frac{A_t}{r - g} = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] \frac{A_t}{r - g}, \quad \text{or}
\]

Finally, the pre- and post-tax growth in asset values, quantities (6.5) and (6.12), are related as shown below:

\[
\frac{gA_t}{r - g} = \frac{gA_t}{r + \tau - g} + \left( \frac{\tau}{r + \tau - g} \right) \frac{gA_t}{r - g}
\]

\[
\frac{gA_t}{r - g} = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] \frac{gA_t}{r - g}, \quad \text{or}
\]

\[
\frac{gA_t}{r - g} = \left[ \left( \frac{r - g}{r + \tau - g} \right) + \left( \frac{\tau}{r + \tau - g} \right) \right] \frac{gA_t}{r - g}, \quad \text{or}
\]

The final term in the expression is the growth in the net present value of tax receipts, or the growth in the value of the ‘asset’ possessed by the government that is bestowed by the right to levy taxes. This quantity has some significance: it represents the amount over and above the cash receipts
that may be borrowed using debt and spent by the government in the present
day, without reducing the value of the government’s asset balance. It is “part
of what the public may spend currently from the tax base without reducing
the net worth of the public equity, or damaging public credit” (Gaffney 2009).

As with the current income generated by the land asset, and the value of
the asset itself, the proportion of the ‘capital’ gain (growth in asset value) that
is privately retained, and the portion effectively transferred to government,
are \((r - g)/(r + \tau - g)\) and \(\tau/(r + \tau - g)\) respectively. The latter of these
terms is the government’s effective ‘equity share’ in land, taken by imposing
the tax.

The asset value effects of introduction of the tax can be seen from the
expressions above to be unequal across different properties with different
growth profiles. Where the expected asset value growth is a large component
of the total return to the land asset (i.e. where \(g\) is large and \(r - g\) is small),
the proportion of the private asset value that is lost to the landowner on
introduction of the tax is large. Where the expected growth is a relatively
small portion of the rate of return to the land, and the current yield is
relatively more important, the devaluation is much less. For example, with
a discount rate of \(r = 0.1\) and a tax rate of 1 percent (\(\tau = 0.01\)), a land
asset where the ‘capital’ gain was three-quarters of the total return and the
current yield a quarter (\(g = 0.075\)) would fall in value by around 29 percent,
whereas if the expected asset value growth were only a quarter of the total
return the private asset value would only fall by around 12 percent.

6.5 Land-value taxation as ‘risk pooling’

This section argues that land-value taxation acts to transfer some of the un-
avoidable risk associated with fluctuations in land values from landowners
to the government and, because this transfer is achieved without undesirable
effects on incentives, effectively constitutes a ‘risk-pooling’ measure without
detrimental efficiency consequences. In the previous section it was explained
that the introduction of a land-value tax is equivalent in effect to the govern-
ment appropriating an equity share in privately-held land assets; the ‘risk-
pooling’ argument here suggests that this equity transfer has benefits in its own right, by also re-allocating risk, and that this adds to the efficiency case for land-value taxation.

6.5.1 The source and nature of the risk

When householders and firms purchase property assets they take on a financial risk, since these assets can unexpectedly appreciate or depreciate in value for reasons entirely outside of their control. The demand for the use of a particular property, and hence also the value of that property, are influenced by policy changes and other trends that affect the amenity of particular locations, as well as by events that tend to affect the value of all property.\footnote{For owner-occupiers who do not anticipate selling, the risk faced in the short and medium run is of a gain or loss in the amenity of their property rather than a loss of rental income or monetary wealth; for simplicity the discussion in this section refers only to a financial risk, but this can be taken to encompass risks of changes in amenity which have an implicit monetary value as well.}

Location-specific fluctuations in property values can result from land use regulation (zoning) changes, as well as from actual changes in use. For instance, permission for an expanded range of uses on a particular site will raise the value of the site in question, and will raise or lower the value of nearby sites depending on whether the externalities associated with the anticipated changes in use are positive or negative. Actual changes in use will clearly alter the amenity of nearby properties and thus affect their values. The benefits to some locations (and detriments to others) from provision of local public goods and infrastructure will similarly be capitalised into the values of properties in those locations.

These effects are well-known, and there are many studies examining the capitalisation of local public goods or planning regulations into property prices (see McMillen and McDonald 2004, Gibbons and Machin 2008, and Grimes and Liang 2009 for instance). The concepts of ‘value capture’ and betterment taxation (as one instrument for value capture) are also of increasing interest to policymakers concerned at the unviability of infrastructure projects, or at the windfall gains granted by rezoning land at the fringe of growing cities (see Fensham and Gleeson 2003). These property-price effects...
are simply being framed here as a risk, an ongoing possibility of windfall gains or losses more or less beyond the control of the landowner (unless they have significant lobbying power).

Non-policy-related trends in the desirability of different locations can also affect property values, either on a small scale (such as when demand for inner-city sites rises as transport costs increase) or on a larger scale (such as when the economies of entire cities decline as key industries disappear). Economy-wide trends, such as interest rate shocks or cyclical movements in real estate values driven by shifts in confidence, expectations or credit availability can cause unanticipated changes in property values over an even larger geographic scale. The general growth of rents and property values as incomes and population grow against a fixed amount of land is obviously anticipated by property owners, but unanticipated variations in this growth are another source of risk.

Most types of fluctuation in the value of a property – and all those which affect the value of particular locations – are actually fluctuations in the value of the underlying land component of the property, not the improvements. As such, the risk being discussed is associated more precisely with land values rather than property values. This point can be understood in two ways. Firstly, it can be seen that all the drivers of changes in the value of developed property that are mentioned above would tend to affect the prices of bare land in the affected locations as well.

A second explanation deals with the question of whether and how a change in a built-property value can meaningfully be separated into changes in value of the components of the property. Although the decomposition of the market value (and changes in that value) of developed property into land value and capital value components might at first appear arbitrary, since it is only the land plus improvements package that is ever traded, it can be argued that there do exist sensible definitions of the value of a site and the value of a building, even when the two are combined so that directly observing the market value of each component separately is impossible. Land value can be defined simply as what the land would sell for if bare, while building value is defined as the difference between the market value of the entire property and
the land value (that is, what the building adds to the previously-determined value of the land). That is, it is valid to claim that the risk associated with unanticipated changes of this type is largely associated with the land component of property, not the capital component.

Some small measure of the idiosyncratic risk of property-ownership is probably still associated with the building, not the land, and this is the risk that the building becomes inappropriate for the area it is situated in (no longer the ‘best and highest use’ of the site). Consider the gentrification of an area which both raises land values and changes the composition of demand towards larger residences. In this case a block of small apartments might increase in market value, but by less than the land underneath has risen in value, meaning the building value itself has fallen as a consequence of no longer being the optimal use of the site.

Once having considered how to meaningfully decompose a built-property value into its components, it seems obvious that property price fluctuations as well as appreciation and depreciation over a boom and bust cycle will largely be driven by changes in land values (excepting those property price movements resulting from capital investment, i.e. maintenance/renovation). Buildings, after all, may be rebuilt, which makes their replacement cost an upper bound on their value. Replacement costs will be relatively stable, as material and construction costs change slowly, and markets for these likely have quite elastic supply except over the very short term (see Caplin et al. 2003, p20). The services of buildings are also delivered over a finite lifespan, meaning the horizon over which a calculating investor might estimate the present value of returns is limited. Land, in contrast, is infinitely-lived, making expectations of its value today highly sensitive to implicit assumptions (or prejudices) about growth and scarcity over a longer timeframe.

---

As previously noted in Section 6.3, the conceptual basis for these definitions is that while a site can have a market value that exists independently of the particular building on it, a building cannot have a market value independently of the particular site it stands on since it usually cannot be removed; this means that land value must be defined first with the value that the building adds to the site calculated as a residual. Since land values can be distinguished from property values, changes in property values can be attributed to either the land or the building; if the land would have changed in value if it were bare, then it is the land underneath the building that has changed in value.
There is no upper bound value (equivalent to a building replacement cost) to constrain expectations, meaning estimates of land values will be often determined by simple assumptions, and swayed by various cognitive biases (these are documented in Shiller 2000 and 2008, and Akerlof and Shiller 2009).

The conclusion that land is the major component of property price fluctuations also emerges from empirical studies examining the correlates of house price volatility, where it is labelled the ‘land leverage hypothesis’ (from Bostic et al. 2007). Houses with a greater share of land value to total value are found to exhibit higher price volatility (Bostic et al. 2007, Zhou and Haurin 2010). Rates of appreciation and depreciation as the overall market rises and falls are also exaggerated for houses with higher ‘land leverage’ (Bourassa et al. 2009).

How significant is land-value risk? In a lengthy study exploring the potential for ‘shared-equity’ home financing, Caplin et al. (2003) consider the question of the volatility of prices of individual Australian residential properties (although they do not consider land prices directly). They strongly emphasise that volatility at the individual household level is markedly different to volatility for the asset class, citing previous studies that find variance in individual property values to be around four to six times larger than variance at a city-wide level (p115, 119). Their study estimates the distribution of capital-growth returns at the individual level by looking at repeat sales from 1984 to 2002, and finds that more than one in four home owners in the sample lost money (in real terms) when selling their property, with around one in ten sellers losing more than 13 per cent (2003, p117).17 They conclude that "by any reasonable measure real estate risk is of immense importance to the typical owner" (p126), and their figures support the view that land assets

---

17Those proportions were for the entire repeat-sales sample. When partitioning the sample according to the tenure time – i.e. years between sales – the distribution of real capital returns remained reasonably steady (although with some evidence of lower returns, and lower variance in returns, over intermediate holding timeframes vis-à-vis short or long timeframes). For instance, the median real growth rate within the bottom decile of returns for a 0-3 year tenure was minus 17 per cent, and the equivalent rates for 6-9 and 12+ year tenures were minus 25.2 per cent and minus 14.9 per cent respectively. The median within the top decile for 0-3 year tenures was a capital growth of 154 per cent, for 6-9 years 125.8 per cent, and for 12+ years 159.7 per cent.
are subject to significant idiosyncratic risk. While some of the variation in their sample could be due to differences in maintenance and renovations between sales, it is reasonable to suppose that differences in land value growth account for much of the variation in overall property value growth.

### 6.5.2 Consequences of land value risk

Many households and businesses that purchase property may be unaware of the extent of this land-value risk, or would prefer not to bear this risk but treat it as a necessary accompaniment to the security of tenure and other benefits that arise from owning the premises that they live (or operate) in.

Diversification or insurance against the risk associated with land ownership is also difficult. Because real estate assets are indivisible and large in value, diversification against the risk associated with individual properties is possible only by purchasing a portfolio of properties, which requires ‘deep pockets’, and still leaves the landowner to bear the risk associated with non-location-specific trends and policy changes (widespread real estate booms and slumps, interest rate changes, etc). And there appear to be no insurance markets available to landowners that would allow them to insure against the land value risks they take on through property ownership.\(^{18}\) Only through defaulting on their debt and becoming bankrupt can highly-leveraged landowners who have seen large falls in the value of their assets avoid some of the downside of this risk.

One house-financing concept with risk pooling properties is the ‘shared equity mortgage’, offered in Australia by both state governments (for certain types of buyers on low incomes) and by at least one commercial provider. Shared equity finance involves house buyers borrowing a proportion of the property value, with no immediate repayment obligation but with the finance provider being entitled to receive (or bear) a portion of the capital gain (or loss) upon sale or at some agreed time horizon.\(^ {19}\) The idea behind this rel-

---

\(^{18}\) See Caplin et al. (2003, section 1.3) for more discussion.

\(^{19}\) For instance, the Adelaide Bank / Rismark Equity Finance Mortgage lends house buyers up to 20% of the purchase price, in exchange for a 40% share in the capital gain (or 20% share in the capital loss) after either 25 years, or when the house is sold, or if the
6.5. **RISK POOLING**

tively new financial product is a “desire to eliminate the ‘indivisibility’ of the housing asset”, so that households may realise the benefits of homeownership without holding a 100 percent stake in the equity of their houses.\(^{20}\) Although these products are typically promoted as reducing barriers to homeownership – since the traditional (debt) mortgage required is smaller, households can purchase a house with a relatively smaller deposit – they also act to pool some of the risk otherwise borne by the house owner in the absence of insurance markets capable of doing so.

It was stated early in this section that the fluctuations in land values that create the financial risk landowners face are out of the control of the landowner. However this is only approximately true; there is much costly activity undertaken by property owners which can be seen as an attempt to manage this risk in their favour by preserving and enhancing land values. A clearer statement is that while the choice of land use and the productive activity undertaken by a landowner on their site do not influence its value,\(^{21}\) there is an incentive for unproductive rent-seeking or ‘rent-preserving’ activity in order to affect policy settings which impact upon demand for the use of a site. This is an undesirable consequence of property owners bearing the risk of land value fluctuations.

This point is seen most clearly for professional land developers: since the uplift in land values from rezoning of agricultural land at the urban fringe is so large, there exists a strong incentive for developers to lobby for such rezoning; in other words, to influence the direction of land value fluctuations in order to be on the ‘upside’ of land value risk. Seeking permission to exceed previously set height and density limits within built-up areas, as well as lobbying for provision of infrastructure to service new developments, are similar actions which influence land values. At the household level, landowners can expend significant time and cost in preventing higher-density development on nearby sites, seeking permission for higher-density development on their own site,

---

\(^{20}\)Quote from Caplin *et al.* (2003, p25).

\(^{21}\)Except in the cases previously discussed of common ownership of related sites, agricultural land, and land-altering improvements.
and lobbying for local public goods and changes in regulations that enhance the value of their land (traffic calming measures, parking restrictions, etc).

The extent of the risk that property-owners face, and the costly risk-mitigation actions taken in response, can be usefully illustrated via case studies. Several recent (newsworthy) examples from Melbourne highlight the consequences of weak and unpredictable planning regulations, and the property-value risk this presents.\textsuperscript{22} In one case, apartment owners are facing development plans for a 36-storey building on the adjacent plot that will not only block their views and sun, but will abut the original building, enclosing its balconies within a 20-storey deep light shaft that will be their sole source of fresh air. In another case, an apartment developer keen to guarantee buyers that their views could not be similarly extinguished by future development of the neighbouring block was forced to purchase and hold a top-floor apartment in that block, later (unsuccessfully) attempting to subdivide this title into an apartment portion and an ‘air rights’ portion. Given the magnitude of estimates of the ‘value of a view’,\textsuperscript{23} it is clear that success or failure in these sorts of planning disputes comes with very high stakes that, judging from the comments of affected owners, are often not appreciated when the risk is taken on.

A final ‘cost’ associated with land-value risk, which is not a form of rent-seeking but still constitutes an inefficient distortion to behaviour, is the effect that variation in land prices over cycles in real estate prices may induce on the timing of sale of property by households who wish to avoid realising unexpected paper losses (or who seek to take advantage of unexpected paper gains).

\textbf{6.5.3 Pooling risk through land tax}

A land-value tax pools a portion of the risk of unexpected changes in land value. This can be seen in two ways. Firstly, unexpected windfall gains or losses in land values lead to commensurately higher or lower land tax

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{22} From *The Age*, "No room for a view", 14 May 2012.
\item \textsuperscript{23} Samarasinghe and Sharp (2008) estimate that a wide water view adds up to 44 per cent to property prices in Auckland, for instance.
\end{itemize}
\end{footnotesize}
liabilities. Landowners receiving the ‘upside’ of this risk have to share some of the upside with the state; landowners on the downside are somewhat compensated through a lower ongoing tax expense.

Secondly, because future land taxes are capitalised into present land values, higher rates of land tax make land values smaller relative to building values. The risky asset, land, thus forms a smaller proportion of the wealth tied up in the real estate owned by households and businesses, meaning that the risk taken on with the purchase of a single asset is lower. The risky asset also constitutes a smaller proportion of the total asset holdings of these households and businesses, making the risk less significant in the context of the household’s overall budget or the business’s balance sheet. At the extreme, in a system with very high land tax rates, capitalisation of the expected tax liabilities would cause land values to be so small in comparison to building values that real estate would trade for an amount not much greater than the value of the building or other improvements, and fluctuations in land values would have a negligible influence on a property owner’s total asset wealth.

6.5.4 Consequences of risk-pooling

Putting aside questions of the fairness of landowners bearing windfall gains and losses, there are benefits to such risk-pooling even in a pure efficiency sense. Firstly, households and firms would be better able to control the risk composition of their wealth holdings at the same time as enjoying the benefits of property ownership, while firms desiring security of tenure would also not be forced to hold substantial amounts of equity (which could otherwise be invested in capital) in the form of land.

A second positive consequence is that the incentive for the costly rent-preserving and rent-seeking behaviour discussed above, which is associated with the full risk and return to land being made private, will be diminished. That is, "the higher the tax rate, the lower the windfalls for which speculators and developers compete" (Feder 1994, p153).24 One possible objection to

---

24It might be argued that although land-value tax would reduce the incentives for rent-seeking lobbying over land use policies and spending on public goods, the enormous gains that would accrue to landowners from subsequently achieving reductions in the land tax
CHAPTER 6. LAND-VALUE TAXATION

the claim that this would be desirable is that this private activity, despite its unproductive costs and distortionary effects on public land-use and spending policies, may be an important source of information for government decision-making and a force acting against poor policy or spending choices that would be absent when downside risks are socialised. However in response it has been noted, by Vickrey (1999 (1992), p16) among others, that the effective transfer of jurisdiction-wide land price gains and losses to public revenue may improve incentives on government to pursue policies which lead to net land value gains, since these would raise tax revenue and thus allow for further spending or tax reductions elsewhere.

Are there any negative consequences of the transfer of the risk of land value fluctuations from landowners to the government, akin to the inefficient behavioural responses that can be associated with risk-pooling through insurance? There is no possibility of adverse selection, since the risk-pooling is created by a compulsory tax, not an opt-in arrangement. Moral hazard, where risk-pooling induces more risky behaviour, is also irrelevant since the extent of the risk of land value fluctuations is not under the control of the landowners who bear the risk. On this point, although individual acts of lobbying clearly influence the probability that land value fluctuations will be in favour of the lobbyist, they do not alter the overall level of risk, which is a function of the frequency of change and general unpredictability of government policies and spending choices. Arguably, socialising the unexpected land gains (and losses) from policy changes would, through discouragement of rent-seeking, reduce the frequency of such unexpected changes as government policy came to be more predictably and consistently driven towards maximising total land values. This is essentially a claim that a system which

\[ \text{rate would spur even greater and more concentrated lobbying efforts in favour of this end. Smith (2000, p3) argues a similar point: “one can only conclude that even if the rational and ethical community charge were introduced, governments do not have the integrity and sophistication to sustain it”}. \]

\[ ^{25}\text{Here we do not consider the negative consequences of reducing the incentive for efficient development (maximisation of joint returns) of co-owned sites, nor of the reduced incentives for undertaking land-altering improvements or improving the fertility of agricultural land. The focus is only on the consequence of reducing the extent of unexpected fluctuations in land values (i.e. risk).} \]
lowered the private stakes in public decision-making over land use would produce better outcomes.

In the discussion of development timing under uncertainty, Section 6.3 hinted at one useful function that fluctuations in land values may perform. Investors that withhold land from development (‘speculate’ in land) act to prevent inefficiently premature development in those situations when generally-unexpected increases in demand materialise, where early development would otherwise waste capital by leaving the site with an inappropriate building. When future demand is higher than generally anticipated, the increase in land values rewards the socially-useful behaviour of these speculators in preventing such loss of capital value, and when demand unexpectedly falls and speculators suffer a loss of land value they are punished for their act of holding the land in inefficient underuse.

As explained in Section 6.3 and here, a land-value tax will limit these gains and losses and thus perhaps create one negative consequence: an under-rewarding of speculators. However, as argued by Tideman (1999, p129), in the absence of the tax the reward accruing to these investors will always outweigh the social value of their actions, so that some level of land tax will improve alignment of incentives. This occurs because while the social value of forestalling premature development is the avoided loss of building capital that will be rendered prematurely obsolete by increased future demand, the private value is the entire unexpected increase in the present value of future rents that would have risen or not regardless of the speculator’s activity. These two quantities differ, and are equal "only if early development would eliminate all of the gains that the currently unforeseen possibility would allow" (Tideman 1999, p128).

6.5.5 Distributional consequences

There appear to be two distinct differences between the risk-pooling aspects of taxing land-values, and taxing and redistributing income. The first was discussed previously: land-value tax produces no negative incentive effects associated with pooling risk (adverse selection, moral hazard) that would
mean the risk-pooling properties come at a cost in terms of efficiency. Land-value tax – except in the selected circumstances mentioned – induces no change in land use type, capital investment, labour supply or other behaviour (aside from any changes resulting from income effects upon introduction of the tax). In contrast, the social insurance benefits of an income tax and transfer scheme come at an efficiency cost, meaning that optimal tax settings for efficiency must consider the trade-off between risk-pooling benefits and undesirable incentive effects (see Varian 1980).

A second difference is that the distribution of revenue from a land tax is not crucial to its risk-pooling characteristics. An income tax and redistribution scheme would obviously not reduce risk from the perspective of individual taxpayers if it were not accompanied by payments to those with low incomes (or those who suffer income shocks). But a land-value tax can be seen to do so, since it is the devaluation of the (risky) land asset in the property-owner’s overall portfolio that reduces the financial risk they face, not the expectation of receiving transfer payments in the event of downside outcomes.

For an established land tax (i.e. ignoring the windfall losses upon introduction of the tax) the financial risk faced by a property-owner is thus lower, whatever the distribution of revenues, simply because the private value of land held is lower than if the tax had not been in place when the asset was purchased. It is not the case that the privately-borne risk remains, and the tax and transfers ameliorate its effects on individuals; rather, the presence of the land tax reduces the extent of the private risk, transferring it to the government instead. From a jurisdiction-wide perspective, there is less of the risky asset available to be held by individuals, since the government has effectively taken a share of the ownership (and risk) of land.

This is why the statement that land-value tax acts to pool risk – or more precisely, that landowners bear less risk in a system with land tax than in a system without (to clarify that the transitional effects have been ignored) – is a useful description of the properties of the tax, even without specifying how the tax revenues are used.

Is it possible that transferring land-value risk to government does not re-
reduce its overall negative consequences, merely transferring risk from landowners to taxpayers? Or, to consider a related objection, might it be argued that land taxes (levied by local government, for instance) will fail to effectively insure landowners against the risk of generalised growth or decline in land values, since tax rates will be adjusted each year to balance the budget? In the latter case tax rates would rise as a proportion of land value as the market declines and fall as the market rises.

The extent to which risk is pooled, rather than merely transferred (and passed back to landowners or onto the general taxpayer) clearly depends on whether the risks are correlated ('systematic' risk) or uncorrelated (non-systematic or idiosyncratic risk). It also depends on whether governments pool risks inter-temporally, by saving tax revenues and smoothing spending so that rates are not constantly adjusted.

While the risks faced by individual landowners associated with broad land price cycles are obviously highly correlated – although less so the larger the jurisdiction – the risks associated with land-use changes and provision of local public goods would be less so. Moreover, many changes in planning regulations or land-use permission that affect property values (as in the examples in section 6.5.2) would appear to be the source of negatively correlated risks for the affected owners – one landowner’s gain is their neighbour’s loss, typically. So the transfer of risk to government is to a substantial extent a transfer of idiosyncratic risk, meaning the benefits of pooling these risks across different landowners (and through time) are real.

Because risk-pooling through land tax is not due to the redistribution of revenue, it is also clearly consistent with a wide range of distributive outcomes, depending on how the revenue is used. For instance, one option would be to make introduction of the tax (as well as the ongoing effect of collecting and spending the revenue) entirely distributionally neutral, by distributing all revenues to landowners in proportion to their site’s original share of the land value of the entire jurisdiction. This would ensure the same income stream and asset value for the affected owners as if the rent of their land had grown steadily over time at the jurisdiction-wide average, a pure risk-pooling benefit for landowners with no other distributional implications.
Clearly, however, less-than-full compensation must be provided to the initial set of landowners if the efficiency gains from reducing the rates of distortionary taxes are to be realised. In this case the distributive effects depend on the profile of the initial landowners, who effectively lose a portion of their wealth, and also on the incidence of the distortionary taxes for which rates are reduced. Thus it is evident that the distributive effects of a shift towards greater use of land-value tax depend not only on the characteristics of the affected landowners, but also on how the tax revenue is used.

The overall risk-pooling properties of a tax and transfer system incorporating a land tax will equally be dependent on how the land-tax revenue is used. However it is clear that so long as less-than-full compensation is granted to the initial landowners, the government gains capacity to ease the risk-pooling vs efficiency trade-off that is implicit in the tax system as a whole.

6.5.6 Summary

It has been argued that the bearing of the risk of land value fluctuations by individual landowners serves little social purpose. It does not encourage entrepreneurship, optimal use of land, capital investment, work, or saving, but it does encourage costly rent-seeking and ‘rent-preserving’ behaviour as landowners attempt to exert what control they have over policy settings in order to prevent ‘downside risk’, and to encourage ‘upside risks’ to materialise in their favour. It may encourage efficient use of capital by preventing the losses associated with premature land development, but the receipt of the full gain or loss in land values probably over-rewards the activity of withholding land, resulting in market failure. Because diversification against land-value risk requires holding a (expensive) portfolio of land assets, and because there appear to be no relevant insurance markets, the usual private responses are difficult or impossible in this case. Individual owners are forced to bear this risk in order to realise the benefits of property ownership, and many likely misunderstand its magnitude when taking it on.

It was argued also that by devaluing the private asset value of land and
reducing its significance within an individual property-owner’s asset portfolio, land-value taxation reduces the financial risk associated with holding the land component of property. The transfer of this risk to the government does not induce undesirable efficiency consequences normally associated with risk-pooling, since adverse selection is impossible and moral hazard is irrelevant, and there may be positive efficiency consequences through reductions in rent-seeking. To a point, this transfer of risk also improves incentives to use land efficiently in an inter-temporal sense, by reducing the incentive for speculators to create an artificial scarcity of land (the subject of section 6.3.2).

Under conditions of certainty, the only efficiency benefit of land taxation comes from a reduction in distortionary taxes. Introducing uncertainty helps makes sense of phenomena such as rent-seeking and speculation, and can support the argument about market failures in development timing (as in Tideman 1999). But it also allows for the possibility of welfare being enhanced by reducing the variation in outcomes that ownership of risky assets presents, a possibility that has been barely discussed.

The basic economics of land taxation supports the argument that the tax acts to pool risk with no efficiency cost, and this argument constitutes a point in favour that is distinct from the general efficiency argument which rests on reductions in distortionary taxes. Moreover, because the devaluation of land – not the promise of compensating transfers for downside outcomes – lies behind the reduction in privately-borne risk, we can understand the risk-pooling point without considering offsetting effects through other parts of the tax system. For the same reason, the risk-pooling benefits might be realised alongside any one of a wide range of redistributive outcomes, including the pure risk-pooling outcome with full compensation of existing owners and no other redistribution.

It is suggested here that this perspective, of land-value tax as a risk-pooling policy without undesirable side-effects, could be useful for better explaining and presenting the case for the tax in a policy context. Although the premises and the economics underlying these arguments are generally well-known and quite obvious, so that the idea that land tax acts to pool risk is somewhat self-evident, few writers have mentioned this property (Vickrey
1999 (1992), p16, and Feder 1994, p153 being exceptions). None have yet framed the issues of the possibility of windfall gains and losses, and the incentive this possibility creates for rent-seeking and rent-preserving action, as a risk and a natural private response to that risk which might be treated as such in government policy.

Recent thought on the role of the state has reframed the traditional areas of government intervention and service provision as collective responses to the risks faced by individuals that markets cannot solve (Quiggin 2007). Particularly in Australia and New Zealand, where ownership of property remains relatively dispersed and continues to be encouraged through policy, land values are a large component of the wealth of many households and the risk this situation creates seems to be under-recognised as such. It is suggested here that seeing land-value taxation as a policy akin to unemployment insurance, for example, in having risk-pooling properties is a useful and novel perspective on the tax. Land-value taxation could be regarded as complementary, and having similar effects, to other direct methods of pooling land-value risk by capturing windfall land value gains from public expenditure and regulatory changes.

6.6 Conclusion

This chapter has offered a new review of existing ideas on the neutrality of the land-value tax and, building on these ideas, an original perspective on the tax itself. This perspective – of land-value taxation as ‘pooling’ and thus improving the allocation of risk, where markets fail to do so – presents a distinct case for the tax, it is argued, as well as providing a useful summary to aid understanding of its effects.

The incidence of a land-value tax is a unique instance where the textbook example of perfectly inelastic supply and ‘zero price transmission’ is actually a plausible approximation to reality, and this result underlies the efficiency properties that make land-value taxation worthy of research and policy attention. Recent research, discussed in Sections 6.2 and 6.3, refines our understanding of when this incidence result, and hence neutrality, holds
6.6. CONCLUSION

in practice. This chapter has endeavoured to clearly lay out the core neutrality argument, and to understand the limitations of that argument: that is, to identify and categorise those situations where neutrality does not hold.

It has been suggested here that a much simpler approach to explaining the neutrality of the land-value tax is to appeal to the ‘lump-sum’ property; reference to the (often-confused) concept of land rents and their relationship with land values has in the past perhaps unnecessarily obscured the simplicity of the lump-sum argument as applied to taxable land values. Questioning whether the landowner’s tax burden is truly a lump-sum can also usefully structure our understanding of the sources of non-neutrality: these can be seen to arise either from different subjective valuations of the discounted tax liabilities, from faulty assessment, or from some ability of a landowner to affect demand for a site.

The first of those categories encompasses the effect of exemptions and progressive rate scales on the ownership of land, but also the effect of uncertainty described by Tideman (1999), whereby differing subjective perceptions of the present value of net-of-tax returns may be re-ranked by the tax. That argument was critiqued in section 6.3.2; it was suggested the result relied on an implicit assumption not made clear in the original exposition, but for which there are nonetheless various reasons to support.

The second of the categories of non-neutrality includes a possibility highlighted by the debate over development timing, that where assessors systematically value land by reference to the particular development project in place, the land-value tax may introduce a bias towards developments with early pay-offs. It was pointed out here that the criteria for neutrality in light of this issue – that assessed land values be a reflection of what the site if bare would fetch for sale – means that only the ‘building-residual’ and not the ‘land-residual’ method ensures neutrality. Since only the former reflects a meaningful conception of land and building values, the methods for good assessment practice and tax neutrality align; where neutrality is not guaranteed it is the assessment method at fault, not the tax.

The final category encompasses a range of well-established causes of non-neutrality: co-ownership of related sites (where land-use externalities are
internalised), land-altering improvements which boost demand for the site itself, and other capital-like characteristics that boost the value of land (such as fertility).

To these can be added ‘rent-seeking’ behaviour; the windfall gains that rezoning or local public goods bestow on landowners are a primary motivation for efforts to influence policy decisions, and the reduction in windfalls due to the tax could accordingly be expected to reduce the intensity of lobbying behaviour.

The potential for land tax, in a world of uncertain future returns, to disincentivise both speculative holding of land and rent-seeking lobbying over land-use and public-goods policies, means the tax offers distinct benefits aside from just replacing distortionary taxes. To a point the first of those effects may correct a market failure, but at too-high a tax rate it may also inefficiently under-reward speculators who provide some social benefit by preventing premature land development and the loss in capital value this causes.

This possibility provides the single qualification to the argument, developed in sections 6.4 and 6.5, that land-value taxation may be viewed as effectively socialising a privately-borne risk that serves no useful purpose. This argument was presented by showing, firstly, that introducing the tax is equivalent in effect to the government taking an equity share in land assets and, secondly, that privately-borne land-value fluctuations are reduced as a consequence. It was then argued that socialising these gains and losses has no incentive effects (a corollary of the general ‘lump-sum’ neutrality argument), other than a reduction in the incentive for rent-seeking expenditures targeted at influencing land-use policy. In summary, aside from the possibility that high enough tax rates may spur inefficiently-fast development, risk-pooling via land taxation improves the allocation of risk where markets appear unable to do so.

This point is based on a well-established body of economic theory about the consequences of the tax; the risk-pooling argument constitutes a new interpretation of that theory rather than an extension of it. Nonetheless it is suggested that the interpretation of land-value taxation in terms of risk
provides a novel and useful perspective from which to view the consequences of the tax, as well as another element relevant to the policy case.

The introduction to this chapter noted that the lump-sum nature of the tax, from which the efficiency property follows, also creates the major practical obstacle to the tax: the effective appropriation of private wealth which occurs upon introduction. The next chapter explores this topic further, by presenting a pair of brief historical surveys intended to highlight the limited influence of the efficiency arguments in the face of distributional (as well as administrative) concerns with the tax. These surveys discuss the development of thought amongst economists throughout the 18th and 19th centuries on the efficiency, justice and practicality of land taxation, and how these arguments played out in the rationale for introduction and subsequent reasons for abandonment of land taxes in several countries throughout the 19th and 20th centuries (Australia and New Zealand in particular).
Chapter 7

Land taxation in historical context

7.1 Introduction

Chapter 6 discussed the basic result regarding the efficiency of a land-value tax, and several implications of this result. One key implication is that the entire burden of the tax – present and future payments – falls upon the current owner, making introduction of the tax akin to the government appropriating an equity share in land assets.

The current chapter explores the significance of this redistributive effect and the relative importance of efficiency arguments both in the history of academic thought and in selected episodes of implementation of land taxation. The first section sketches the debates amongst classical and early neo-classical economists about the economic merits and the justice of land taxation, while the second section discusses the playing out of these themes in the implementation of land taxes, and the rationale and motivation for this in Australia and New Zealand in particular. These surveys are included to highlight some of the distributive and political considerations that may explain why the efficiency case for land taxation has not led to greater support for land taxes in practice.

As discussed early in the previous chapter, contemporary debates about tax reform are largely focussed on the question of how to minimise the distortionary costs of taxation (whilst preserving revenue levels and attaining
social goals). But, as outlined below, the efficiency merits of land taxation have been well-understood for over two centuries, yet land taxes are little-used today and have been gradually abandoned even in those places where they were once well-established. A genuine question remains about why policymakers and the public have seen so little attraction in tapping the revenue base of land. This chapter asks, firstly, how the classical economists viewed land-value taxation in terms of both its economic and moral aspects, before considering how influential these arguments have been in the adoption of land taxes, and how relevant they have been to the gradual erosion and repeal of these taxes.

7.2 A history of economic thought

The desirability and feasibility of taxes on land rent was a major preoccupation of many of the prominent classical economists: Smith, Ricardo, and James and John Stuart Mill in particular, and before them the Physiocrats. Somewhat in continuation of their line of thought, and somewhat in response to the popularity of Henry George’s *Progress and Poverty* (1879), the merits of land taxation were also debated by several of the early neoclassicals (Marshall and Pigou, amongst others). For nearly all of these writers, the salient questions regarding land tax were less about its incidence or economic effects (for which a broad agreement existed, at least from Ricardo onwards), but about the justice of private ownership of land rent and the fairness or unfairness of the various tax proposals designed to socialise this rent.¹

7.2.1 The Physiocrats and Smith

A core concept of the Physiocrats was their theory of tax incidence – the idea that all taxes levied in an economy ultimately reduce the rent paid to landlords.² In their logic, this followed from the belief that only agricultural

¹See Prest (1981), Tideman (1994), and the relevant chapters in Blaug (1985) for discussions of historical thought on land rent and land taxation.

²Similar ideas had been put forward by John Locke and other writers (Gaffney 2009, p374).
activities were ‘productive’ in the sense of yielding a surplus of product over the costs of production, while the processing and trading undertaken by manufacturers and merchants (labelled the ‘sterile class’ by Quesnay) were merely transformative, adding no net value. The natural implication of their tax incidence theory was a policy of rationalising tax collection by imposing a single tax on the net product of land (rent), thereby reducing collection costs and avoiding impediments to trade—that is, deadweight loss (Blaug 1985, p24-28; Tideman 1994, p105-106).

Adam Smith, although criticising the “disagreeable discussion of the metaphysical arguments by which they support their very ingenious theory” (2007 (1776), p643), nonetheless produced a series of arguments on the incidence of particular taxes that, in sum, amounted to an endorsement of the Physiocratic theory. Smith reasoned separately that taxes on the produce of land (agricultural output), on the rent of houses, or on the number of windows on a house, on commercial profits and interest, on the wages of labour, and on the transfer of land titles, must all ultimately reduce the income finally received by the landlord as the owner of an immobile productive asset.\(^3\) To illustrate the style of reasoning consider an example of a tax on the profits of capital stock (entrepreneurial profits or compensation “for the risk and trouble of employing the stock”; p657). If this capital were employed in agriculture, the owner might increase his gross rate of profit, in order to retain the same net profit, only by reducing the rent paid to the landlord (a logic which implicitly assumed that agricultural prices may not rise). If the owner were a manufacturer, in order to pay the tax he must pay a lower rate of interest on money borrowed, or accept a lower rate of profit (again assuming that prices cannot be raised). However because stock (capital) is internationally mobile, reductions in profit or interest would drive the manufacturer or moneylender abroad, thereby lowering wages and ultimately land rent in the taxing country.

Given his view of incidence, Smith therefore favoured taxes on ‘ground-rents’ (the income attributable to the land under houses) and on the rent

\(^3\)See pages 648, 651, 657, 671, and 669 respectively in Smith (2007 (1776)) for these arguments.
of agricultural land, for the practical reasons identified by the Physiocrats (avoiding impediments to trade). However he also favoured land taxes for moral reasons, consistent with the views laid out in his four ‘maxims of taxation’. These elements – practical and moral – are seen mostly clearly in the quotation below, which illustrates both efficiency concerns and application of the benefits-received principle of taxation to the income of landowners:

"Both ground rents and the ordinary rent of land are a species of revenue which the owner, in many cases, enjoys without any care or attention of his own. Though a part of this revenue should be taken from him in order to defray the expenses of the state, no discouragement will thereby be given to any sort of industry. The annual produce of the land and labour of the society, the real wealth and revenue of the great body of the people, might be the same after such a tax as before...

Ground rents, so far as they exceed the ordinary rent of land, are altogether owing to the good government of the sovereign, which, by protecting the industry either of the whole people, or of the inhabitants of some particular place, enables them to pay so much more than its real value for the ground which they build their houses upon; or to make to its owner so much more than compensation for the loss which he might sustain by this use of it. Nothing can be more reasonable than that a fund which owes its existence to the good government of the state should be taxed peculiarly, or should contribute something more than the greater part of other funds, toward the support of that government."\(^4\)

Smith argued too that movements in the rents and land values were driven by, and were an indicator of, the general progress of society rather than being under the control of the landlords:

"Every improvement in the circumstances of the society tends either directly or indirectly to raise the real rent of land, to increase the real wealth of the landlord... [landowners] are the only

one of the three orders whose revenue costs them neither labour
nor care, but comes to them, as it were, of its own accord, and
independent of any plan or project of their own." (p198-199)

7.2.2 Ricardo and the Mills

In contrast to Smith, and despite acceptance of his by-then-established claims
about the efficiency of taxes on ground-rents, David Ricardo was not a strong
supporter of land taxation policies. Ricardo was somewhat critical of the
justice of appropriating accumulated wealth, whilst also maintaining a belief
that uncertainty over future taxation would cause land to be driven into the
hands of speculators (Connellan 2004, p14; Tideman 1994, p110; Blaug 1985,
p84). Nonetheless he made several lasting, albeit indirect, contributions to
the theory of land taxation. The first was to clearly delineate ‘pure’ rent
from those elements of contractual rent which constitute returns to invest-
ment and labour. This delineation is summarised in the statement that rent
is a payment for “the use of the original and indestructible powers of the
soil”; that is, the return to the fixed, non-reproducible, aspects of land which
may be safely taxed without resulting in decreased production or increased
agricultural prices.

A second contribution was to provide a superior formal presentation of
the theory of rent, which explained the existence of rent as a consequence
of diminishing returns to variable inputs, and quantified the level of rent as
the excess product over and above outlays on variable inputs (equivalently,
the difference between average and marginal products of the variable input,
multiplied by the number of units used). By this theory, the equilibrium

5Elements of this theory had been previously developed by Anderson in 1777, and the
basic theory was actually presented by each of Ricardo, Malthus, West, and Torrens in
the same year, 1815 (Blaug 1985, p77; O’Brien 2004, p45,146). The magnitude of rent for
a particular site could be defined with reference either to the ‘extensive margin’ (the least
fertile land in use: marginal, or rent-free land) or the ‘intensive margin’ (the final unit
of a variable input applied to sites inside the extensive margin). Rent is seen to be the
excess of what is produced on a site over what the variable inputs used on that site could
produce if either (a) applied to sites at the extensive margin of production, or (b) applied
equally to other infra-marginal sites at their intensive margins. Since marginal products of
variable inputs will be equalised at the intensive and extensive margins, these definitions
level of rent was clearly understood to be a surplus or residual, and not a cost of production from a societal point of view (even if it were such from the perspective of a renter of land). This positioning of rent provided further theoretical support for the case that taxes on rent were uniquely harmless in their effects on the economy, since it explained clearly why they would not be transmitted into goods prices (Tideman 1994, p109).

Like Ricardo, James and John Stuart Mill were concerned with the justice of taxing land rent, and both promoted an adapted land tax proposal first suggested by James Mill. Using the reasoning that the existing level of land rent had been bought and paid for by the landowner, and confiscation of this would be unjust, it was proposed that increments in rent above some previously ascertained level be subject to tax (Blaug 1985, p84; O’Brien 2004, p299; Tideman 1994, p111-118; Prest 1981, p10-12). It was recognised by both writers that changes in legislation, population, and prosperity caused windfall gains in land values. Taxation of these gains was thereby seen as a legitimate expression of the benefit principle, with J.S. Mill arguing that:

"[landlords] grow richer, as it were, in their sleep, without working, risking or economising. What claim have they, on the general principle of social justice, to this accession of riches?"

Both Mills, however, made a conceptual error in their advocacy of a land-tax-with-threshold, by ignoring the fact that anticipated rises in rent could be capitalised into sales prices, thus rendering unfair (by their own standard) taxation of these gains (Tideman 1994, p111-118). McCulloch, an influential classical writer on taxation, shared with the Mills the concern over
imposing "particular burdens on the land" which amounted to "confiscation of a portion of the property of the landlords"\textsuperscript{7}, yet still opposed their mooted solution on the grounds that distinguishing between the returns to the soil and to the improvements on the land was practically impossible (Tideman 1994, p113-116; O’Brien 2004, p300).

\subsection*{7.2.3 From Henry George onwards}

The writer responsible for popularising the policy of land-value taxation, and credited even with determining (in a reactionary way) the direction of early neoclassical economics (Czech 2009), was the American Henry George, a self-taught economist and populist social reformer. George’s 1879 \textit{Progress and Poverty} was motivated by the observance of simultaneous economic growth and increasing poverty, which he attributed to land speculation, particularly the artificial scarcity created by land barons withholding land from production and banking on capital gain (Tideman 2004, p118). His remedy was a single tax on land, at sufficiently high a level to replace all other taxes. Despite the enormous impact on his contemporaries, Georgist ethics today persist only at the fringe of political debates, and George is infrequently discussed in histories of economic thought. The latter is less surprising if it is the case, as Prest (1981, p13) claims, that George’s "whole argument [is] a general moral theory of property rights rather than a piece of economic or statistical reasoning."\textsuperscript{8}

Despite Ricardo’s characterisation of rent as accruing to the ‘indestructible powers’ of land, and his theory of differential rents as determined by the quality of inhomogeneous land, he and later writers failed to explicitly apply his theory to urban land rents. For urban land, the ‘powers’ referred to simply consist of (literally indestructible) space to build upon, while the quality of inhomogeneous land is easily associated with proximity to urban centres or amenities (i.e. people, markets and public goods). The logical extension

\textsuperscript{7}McCulloch (1863 (1845), p59).

\textsuperscript{8}For more on George, see Andelson (2003) and Laurent (2005). One paradoxical aspect of his legacy is that despite his strong libertarian leanings, George’s speeches and advocacy are credited as inspiring much of the support for trade unionism and socialism in the late 19th century (Connellan 2004, p17; Prest 1981, p65).
CHAPTER 7. LAND TAXATION IN HISTORICAL CONTEXT

of differential rent theory from its agricultural origins to the urban land context was provided by Marshall, who demonstrated that the concepts of the intensive and extensive margins of production also applied in urban areas to site-use intensity and the urban perimeter, respectively (Prest 1981, p14; Connellan 2004, p14). Marshall provided several other nuanced contributions to rent theory: he recognised that well-executed planned (or ‘comprehensive’) development, as opposed to site-by-site development, could produce higher aggregate land rents; when such development planning were undertaken by a private entity some of the observed rent would actually be akin to profits. He also argued that, due to the fact of capitalisation of public expenditure into rents, a site value tax in conjunction with increased provision of public services could potentially raise rather than lower land values (Prest 1981, p15).

Marshall, like the Mills, recognised the extent of receipt of windfall gains by private owners, and favoured land-value taxation but worried about distributional issues; he proposed several other schemes for socialising rent, such as nationalising land after a 100-year period (Connellan 2004, p16; Prest 1981, p16). Arthur Pigou, too, favoured land-value taxation for reasons of efficiency but suggested that a separate tax on windfall increments in land values (a ‘betterment tax’, with equal compensation for windfall losses) be also implemented alongside the annual charge. The two taxes in parallel would have the advantage over a single site value tax of capturing a large share of unearned capital gains, without necessarily appropriating large amounts of savings held in land form; the combination would be both efficient, and distributionally more sound (Prest 1981, p18).

7.2.4 Summary

Several themes emerge from the classical and early neoclassical economists’ debates over land taxation, and to a large extent these mirror the preoccupations of writers and policymakers regarding land taxation today. The first point is that the theoretical case that taxes on pure rent are neutral in their resource allocation effects was unambiguously accepted, and all writ-
ers agreed that land rent was thus a good subject for taxation according to an efficiency criterion (even if it were recognised, as Marshall did, that land rent was but “a leading species of a large genus” (Marshall 1893)). A second theme was fairness and justice, both of taxing away accumulated savings in a particular class of asset, and of private receipt of windfall gains derived from societal progress. Disagreements over the nature of just appropriation were a major focus for writers from Ricardo onwards, and spawned a variety of distinct policy proposals for socialising land rents and windfalls. Finally, some writers questioned the ease with which pure land rent could be distinguished from the return to improvements – issues of land value assessment – which would be revisited later in the context of problems with public acceptance of the tax.

7.3 A history of implementation

7.3.1 The contemporary role of taxes on land

A brief examination of taxation systems around the world today prompts the suggestion that much of the intellectual effort and many of the ideas of the classical economists regarding land taxes were contributed in vain. This is because, firstly, direct taxes on land today constitute no more than a minor portion of government revenue in any country. Moreover, the efficiency properties of land taxes are frequently forgotten or ignored (and are evidently not seen as so important as they were to the classical economists), the economic logic of publicly-conferred land-value gains is not widely understood and rarely features in policy discussions, and the accompanying moral position regarding land rent taken by many of the classical economists appears even less often.

Taxes on land are used today primarily by local governments, in the form of the property tax (known as ‘rates’ in the UK, Australia, and New Zealand). Across the OECD, taxes on property (very broadly defined) raise between 1 and 13 percent of total government revenues, with an OECD average of around 6 percent (Australia raises 9 percent, New Zealand 5 percent, and
the US and UK 11 and 13 percent respectively). However taxes on immovable property (i.e. real estate) contribute on average only half of the revenue attributable to this broad category. The US raises 10 percent of all tax revenue from the local property tax, the highest of all OECD countries. In comparison, the OECD average is around 3 percent, and the figures for Australia, New Zealand, and the UK are 4.5 percent, 5 percent, and 9 percent respectively. With the exception of France, European countries rely less on property taxation than do the countries of the Anglosphere (the European figures are mostly lower than for the OECD as a whole). For the OECD as a whole over the last three decades, there has been no noticeable trend in the shares of total tax revenue of either the broadly defined ‘taxes on property’ category, or of the ‘recurrent taxes on immovable property’ subcategory (Heady 2009, p24).

Of the latter subcategory, which provides no more than 10 percent of tax revenue in any OECD country, only a portion actually constitutes taxation of land rents. That portion depends on the relevant base of the property tax, the breakdown of property values between land and improvements, and assessment practices in the tax jurisdictions in question.

In Australia rates are the sole tax instrument available to local governments and, in contrast to most other countries, many local governments levy rates on land values rather than on the capital value of properties. Australia is also an outlier in making use of a state-level land tax (albeit with numerous exemptions), which raises between 5.7 (Queensland) and 11.5 (NSW) percent of state-level tax revenue, except in the Northern Territory which does not impose a land tax. Only these local government rates and state-level land taxes, which account for the miniscule 4.5 percent of Australia’s total tax

---

9 The OECD ‘Taxes on Property’ category includes four major subcategories: recurrent taxes on immovable property (i.e. the property tax / rates); recurrent taxes on net wealth; estate, inheritance and gift taxes; and taxes on financial and capital transactions. Only the first subcategory (and the second, to the extent that landholdings are a component of overall wealth) incorporates a tax on land rents or land values (this subcategory also includes the building-capital portion of taxes on property). Figures are for 2007, from OECD (2009, p80,116).

10 The allowable ratings bases vary by state, and councils in some states have discretion over the choice of rating base; see Commonwealth of Australia (2008a, p141).

11 Commonwealth of Australia (2008a, p76).
revenue provided by ‘recurrent taxes on immovable property’, are similar to
the kind of land taxes supported by the Mills, Henry George, and the other
economists discussed in the previous section. The other half of Australia’s
revenue from ‘taxes on property’ comes from conveyance duty (‘stamp duty’),
a tax on transfers of real estate which bears little relation to taxation of land
rents.

Franzsen (2009) provides a short list of other countries imposing special
taxes on land (or allowing higher rates on land than on improvements as
part of ordinary property taxation). As well as a number of developing
countries, France and Denmark raise some revenue from taxes specifically
on land (2.5 and 1.9 percent of their overall tax takes respectively\textsuperscript{12}), and
some municipal governments in New Zealand, Canada and the USA use a
land value base or levy higher rates on land than on improvements. Not
mentioned in Franzsen’s list are Hong Kong and Singapore, which both levy
property taxes on land and improvement values equally. Despite having no
recurrent taxes levied solely on land these two jurisdictions raise much public
revenue from land rents, arguably much more effectively than Australia’s
heavily eroded land taxes, via a number of different instruments: a land-plus-
improvements property tax, public land ownership, a range of development
charges and betterment taxes and, in Hong Kong in particular, staggered
sales of land leases which are timed to maximise land sale revenue (Hui et
al. 2004).

In sum, direct taxes on land play only a minor role in modern tax sys-
tems, taking their most visible and important form in concert with a tax
on improvements capital as part of the property tax.\textsuperscript{13} Land rent, to the
extent that it appears in ordinary personal income, may also be taxed in
this form. However generous tax treatment for income derived from real
estate means that in practice land rent is lightly taxed: the imputed rent
to owner-occupiers is typically exempt from income tax, for instance, and
in some countries there is concessional treatment of capital gains income,

\textsuperscript{12}OECD Revenue Statistics (2009, p116).
\textsuperscript{13}This is what led Vickrey (1999 (1995), p17) to label the property tax as “a combination
of one of the worst taxes – the part that is assessed on real estate improvements – and
one of the best taxes – the tax on land or site value”.
together with the ability to offset real estate investment losses against other income in determining income tax liabilities. In these countries the norm is for current income from landholdings to fall short of the interest payments incurred in leveraged purchases of the land, so that no tax is paid on land rent.\footnote{New Zealand’s total stock of residential rental property, valued at around $200 billion, produced a $500 million net loss and thus contributed around negative $150 million in tax receipts in 2008-09. That is, because much of this property was purchased with borrowed money with debt repayments exceeding rental income, the sector not only contributed no tax revenue but also allowed investors to avoid $150 million in tax on income from other activities (Fallow 2009; NZ Treasury 2009, p15). In Australia in 2005-06, rental property owners in aggregate lost $5.1 billion; it is estimated that negative gearing cost the federal government at least $1.5 billion or around 1 percent of the personal income tax take (Commonwealth of Australia 2008b, p61 fn39).}

This snapshot of the contemporary role of land taxes poses a puzzle: why have the ideas of the early economists on land taxation apparently had so little influence on the design of tax systems today? The discussion in Chapter 6 showed that the efficiency arguments of Smith and Ricardo remain largely intact today; the limited use of land taxation is not a result of intellectual advances rendering their views redundant.

Equally, however, this snapshot masks an interesting 20th century history involving debate, implementation, erosion, and repeal of taxes on land, particularly in New Zealand, Australia, and the US. Many of the arguments and concerns present in the writings of the early economists are reflected in the reasons for the rise and fall of land-value taxation over this period. In particular the arguments about the feasibility and fairness of land taxes for new countries with rapidly growing economies, and about the benefits of exempting improvements from taxation as a spur to development, were influential for the promoters of land-value taxes; the concerns raised about accuracy in land value assessments also proved important, in hindsight, to public perceptions of the fairness of the tax.

Before turning to the experiences of these countries, an early but significant instance of economic thought influencing land tax policy is worth mentioning. James Mill, recognising that appropriating land rent under a well-developed system of property rights would be difficult and unjust, ar-
gue that for a new country, or one without the strength of recognition of private property of the UK, servicing the entire government revenue need by claiming rent could be practical and efficient: “industry by that means would not sustain the smallest depression... and the expense of government would be defrayed without imposing any burden upon any individual”.\textsuperscript{15} Mill worked for the East India company and consequently his ideas on rent as public revenue significantly influenced the administration of India; Winch (1966, p167) claims that “Mill’s influence on the Indian land revenue system represents perhaps the single most important application of Ricardian economics in practice”.

\subsection*{7.3.2 New Zealand and Australia}

In New Zealand, the arguments of the classical economists (and later, Henry George) were clearly a major force behind adoption of both a dedicated land tax, and land-value rating as the basis of local government finance. As early as 1853, Governor George Grey indicated a wish to impose a land tax “to prevent the acquisition of large areas of unoccupied land”. In 1878 Grey (as Premier) and the treasurer, John Ballance, introduced a land tax which, after being converted to a property tax, took its final form with improvements exempted in 1891.\textsuperscript{16} The reasoning of James and J.S. Mill about the suitability of taxing land in a rapidly growing economy, and the desire to avoid hindering development and capital investment, are evident in Ballance’s justification for the tax:

“We believe that no form of wealth is more legitimately called upon to contribute a portion of the public revenue of the colony than the value of land minus improvements, which for brevity, I shall call the unimproved value, as no other commodity increases so rapidly in value from the increase of population and the natural progress of a country. By exempting improvements, we award

\textsuperscript{15}Mill (1824 (1821), p243, Ch IV, Section V).
\textsuperscript{16}Keall (2000). Keall also notes that Grey met and discussed land taxation with John Stuart Mill, and corresponded with Henry George (p422).
a premium to industry and discourage a system of speculation which thrives only upon the labour of others.”

As the statements from Grey and Ballance attest, another force behind the imposition and design of the tax was an egalitarian concern for preventing concentration in land ownership and for encouraging productive use of land. To that end the New Zealand land tax was levied with a progressive rate schedule, with 16 different rates rising from a minimum of around 0.4 percent to a maximum of 1.25 percent for those estates valued in the highest bracket. The explicit purposes of the tax were both to encourage large landowners to contribute a greater share of revenue, and to spur them to improve their lands (Seligman 1895, p317-322).

In addition, the 1891 version of the tax contained two further innovations which can be seen as attempts to maintain fairness in allocation of the burden of the tax. Firstly, landowners holding mortgage debt were assessed on the value of their land less the amount of mortgage outstanding, whilst mortgagees were taxed on the value of the mortgage; that is, mortgage assets were treated as ownership stakes in real estate for tax purposes. Secondly, landowners were offered an option to escape faulty or unfair assessments of land value in determination of their tax liability. A landowner would submit his own valuation, but would be required to pay tax on a separate government-assessed valuation; if the owner thought the government assessment were too high, he could call on the government to purchase the land at his own submitted valuation (Seligman 1895, p319).

In local government finance, too, from the early days of the colony lawmakers encouraged rating on an unimproved land value basis. Until 1876 the provinces raised money both from land sales and from rates levied on a variety of bases – capital value including improvements, unimproved site value excluding improvements, but most commonly annual rental value (the English rating system of the time) – with Wellington the first to adopt land-value rating in 1849 (Franzsen 2009). The replacement of the system of provinces

---

17 Dwyer (2009, p103).
18 One penny in the pound was the base rate; for the largest estates the rate was three pence in the pound (Seligman 1895, p318).
7.3. A HISTORY OF IMPLEMENTATION

by local governments with different powers in 1876 was accompanied by laws to achieve national uniformity in rating systems, using the English (annual value) rating base.

There were various reasons why this uniform system was abandoned shortly afterwards in favour of a system of local choice, with one option being land-value rating. The first was a desire to discourage idle and speculative holding of land (undeveloped land having no rental value but a high market value), and to encourage improvements (as these tended to raise rental values and thus tax liabilities). The English system, although well-suited to the predominant landlord-tenant tenure pattern there, was also difficult to administer when most land was purchased not for rental but to be occupied and worked by the owner (Keall 2000, p425). From 1896 there began a process of local governments switching to land-value rating; by 1985 land taxation was the dominant choice of local government tax, with 80 percent of local authorities using the land-value rating system.

In the Australian colonies the sentiments leading to land taxation were similar: a desire to encourage pastoral development without reducing incentives for investment (Smith 1993, p19-20), and a growing need for funds as tariffs were lowered and as revenue from land sales diminished (Smith 1993, p19-24). Between 1877 and 1915 each of the states introduced a land tax, and from 1910 a federal land tax co-existed with these. As in New Zealand, rating for local authority revenue was initially based on the English standard of ‘annual rental value’, which included capital improvements, but all states bar Tasmania eventually mandated either an optional or compulsory switch to unimproved site value rating. The goals of stimulating construction and penalising absentee owners of idle land were central here too (Anderson 1996, p8).

A central (and explicitly stated) aim of the land taxes was to break up the extreme concentration of land in the hands of the dominant landholders, the squatters who had acquired large tracts from the colonies cheaply by purchase and by conversion of leasehold tenure to freehold. One aspect of this goal was forcing speculatively held land to be brought into cultivation –

---

to ‘unlock the land’ for small-holding farmers; another was redistributive, with popular discontent at the growing concentration of land wealth and an increasing spread of Georgist and socialist ideas (Smith 1993, p19-21; Worrall 1981).

The passage of legislation was fought against by land-owning interests in the Parliamentary upper houses (Legislative Councils), and was substantially delayed in several of the colonies. In Western Australia the first three attempts to introduce the tax were defeated, and in Queensland the tax was finally implemented in 1915 after more than 40 years since first being debated (Smith 1993, p22-23). Although there was variety in the form of the taxes, in line with their redistributive aims all had a progressive rate schedule in the form of either a tax-free threshold or graduated rates, with the federal tax and Queensland’s land tax being the most strongly progressive.

Thus the key economic and moral ideas of the classical economists were clearly reflected in the arguments which brought about the heyday of land taxation in Australia and New Zealand: exempting improvements and holding down the rates of other taxes would encourage development and not inhibit industry (Smith); in a new country with a growing economy the ‘unearned increment’ may rightly be claimed by government without unjust abuse of property rights (James and J.S. Mill; in Australia the removal of ill-gained land rights was seen as a correction, not a creation, of injustice); and a penalty for unproductive speculative landholding would hasten the development of the land (George). The taxes that resulted from the interplay of these ideas and other practical considerations (such as the breaking up of the large estates) with the politics of the time were clearly tools of social policy as well as efficient revenue-raisers.

\[20\] In NSW it was concluded that the land tax had forced much unproductive land into use, and that “the holders of idle land do not complain as they perceive that the remedy for a condition of which they may be tempted to complain lies in their own hands”, as the NSW Statistician observed in 1909 (Smith 1993, p34 fn28).

\[21\] The “Single Tax Leagues”, adherents to Henry George’s doctrines, were influential in garnering public and political support, but their support for George’s economic orthodoxy in the form of proportional tax rates without exemptions led to a schism with the socialists and the labour movement, whose aims were more explicitly redistributive and who thus supported progressive rate schedules (Smith 1993, p19-24).
7.3. A HISTORY OF IMPLEMENTATION

However by the middle of the twentieth century the importance of the land taxes had declined, and the second half of the century has seen the acceptance, popularity, and use of land taxation greatly diminished.\textsuperscript{22} The state and national land taxes in Australia and New Zealand have been eroded by exemptions and concessions (or removed altogether), the local government rating base has frequently been extended to include capital improvements, and it is clear that arguments about the efficiency or justice of collecting land rent are either unknown or hold little force for policymakers and taxpayers.

Recounting the gradual erosion of land-value taxes, it seems that perceptions of unfairness regarding the scope and nature of the tax, particularly once exemptions became widespread, have driven a general decline in political support. Compounding factors, including poor administration and assessment practices leading to seemingly arbitrary jumps in tax liabilities, and low revenue yields, have contributed to these perceptions of unfairness and irrelevancy. The disappearance from public debate of the notion of an ‘unearned increment’, or of any efficiency or equity case for special taxes on land, has also helped the rollback of land taxation to proceed with little public resistance.

In Australia the federal land tax was ‘returned to the states’ in 1952, as compensation for their abandonment of income taxation; however over the next half-century rather than further exploit the land base the states instead exempted various categories of users, raised rates on the remaining base, and steepened rate schedules. Smith (1993) describes inflation in the 1970’s as the catalyst for this process of erosion. Inflation in land prices and bracket creep gave rise to politically embarrassing cases of hardship following steep rises in tax liabilities, and lags between valuation and the issuing of tax bills in a fluctuating property market made liabilities for taxpayers and revenue for governments volatile and unpredictable.

As land became a less important store of wealth, the land taxes also became seen as discriminatory wealth taxes: not only were other forms of wealth untaxed, but the exemptions for residential owner-occupiers and pri-

\textsuperscript{22} By 1945 the New Zealand land tax contributed only 1.5 percent of total revenue (Reece 1993, p226).
mary producers meant land taxes were levied on a narrow base at higher rates, thereby increasing public resentment of the tax. The final result of this process: a tax contributing only around 3 percent of state budgets, primarily from commercial and industrial landholders.\textsuperscript{23} Smith (1993, p78) labels it “an apology for the absence of any wealth tax... it is but a reminder of the grander visions of nineteenth century liberals and social reformers”.

Similar sentiments in New Zealand led to the complete repeal of the land tax in 1990. Previous governments had exempted owner-occupied and agricultural land and had raised the rate on commercial land. A 7-year land price boom had caused the tax take to rise substantially, and after a rapid fall in prices the tax assessments based on valuations issued at the peak of the boom caused widespread objection. The infrequency of revaluations in the rapidly rising property market of the 1980’s, which made revisions in tax liabilities large when they occurred, was one part of the problem; another was that the boom had pushed the value of many small landholdings over the tax-free threshold. Without popular support the tax became a political sticking-point and was quickly repealed (Reece 1993; Keall 2000, p423). Reece (1993) documents the change of tack taken by the reformist government of the time: faced with a choice of broadening the base or abolishing the tax altogether, they initially opted for the former by removing some exemptions and cutting the threshold substantially. However this brought additional landowners into the tax net and just one year later, with an upcoming election and an opposition promising abolition, it was decided that the repeal of this ‘nuisance tax’ was rather more politically advantageous.

At the local government level in New Zealand, a steady shift away from land-value rating towards capital-value rating and the use of other charges has also taken place since 1985: the 80 percent of authorities using land-value rating in that year had fallen to 44 percent by 2007 (Franzsen 2009, p38). A restructuring of tax jurisdictions in 1990 which reduced the number of local authorities by two-thirds, and central government legislation which

\textsuperscript{23}The Australian states and territories collected $4.4 billion in land taxes in 2006-07 (excluding local government rates); this is 9 percent of total state tax revenue of $49 billion (Commonwealth of Australia 2008a, p13), 3 percent of total state revenue of $154 billion (p297), and 1.4 percent of Australia’s total tax revenue of $320 billion (p13).
removed the requirement to alter the rating base only after a poll, encouraged this shift. One consequence of the restructuring was the amalgamation of previously separated rural and urban areas, which created (under land-value rating) an apparent cross-subsidy of urban services by rural landowners (Keall 2000, p427). Thus the shift to capital value rating may partly be explained by a desire to equalise the tax burden across land-intensive rural and capital-intensive urban properties. McCluskey et al. (2006) also note that the jurisdictions that switched tax bases were those with higher median incomes, and although no explanation is given for this relationship it is possible that land valuation difficulties in the built-up urban (and presumably wealthier) centres made capital value rating more attractive to local governments, and more transparent and hence acceptable to ratepayers (Franzsen 2009, p38).

Another theory for the decline in attractiveness of land-value rating to ratepayers and local politicians relates to the changing nature of the public goods provided by local government, particular in urban areas. The services traditionally provided (such as investment in and maintenance of water supply, sewerage, and roads) had an obvious corollary in higher land values, and site value rating was therefore widely seen as fair, in the sense of being consistent with the benefit principle of taxation. However as local governments have expanded the scope of their activity into ‘people services’ (cultural, social, and economic development activities), which are less obviously related to land-value uplift, voters and politicians have come to perceive residents rather than landowners as the primary beneficiaries of local public goods (McCluskey and Franzsen 2005, p141).

Thus capital-value rating, which imposes more equal per-household burdens than land-value rating, with any remaining disproportionality in tax contributions directly related to the most visible measure of household wealth (total property value), is seen as a more equitable system of financing. Other tax policies such as per-household charges, and differential rating by land use type, are also increasingly used with similar ‘benefits received’ reasoning. Although there is little distinction in theory between ‘traditional’ and ‘modern’ services in terms of impact on land values – both operate through the same
mechanism, an increased willingness to pay to reside in the local government area which is capitalised into the price of land – support for land-value rating on benefit grounds has evidently disappeared.

In Victoria, Australia, a site value rating option was granted to local governments in 1914, and by 1993 approximately half of the urban authorities (but only 7 percent of rural authorities) had adopted site-value rating, typically following a poll (Putland 2008). After an amalgamation process in the early 1990’s and a 1993 report recommending councils adopt the capital-value rating system (with legislated incentive to do so, in the form of permission to use differentiated rates by property type), the majority of councils switched to capital value rating.

In 2010, the last local government using site-value rating (the City of Monash) abandoned this in favour of the capital improved value rating base. The arguments presented for the change exhibit some of the ‘modern’ interpretations of benefit and equity that McCluskey and Franzsen (2005) discuss in the New Zealand context: that under land-value rating, there is a large variation in rates across residential household types (land-intensive houses versus capital-intensive apartments), services consumed apparently bear no relation to tax burdens, and ‘ability to pay’ is less closely correlated with land value than with capital value (MacroPlan 2009 p16-20; City of Monash 2010; Putland 2009).

The final explanation offered for the decline of land taxes and land-value rating downplays the role of democratic choice and supposed dissatisfaction with these taxes, instead emphasising interest group politics over a reasonably mundane and low-visibility policy choice as driving the changes seen. As Keall (2000) argues for New Zealand:

“wherever land-value rating applies, it has been adopted by a poll of ratepayers representing a great amount of work and profound social concern. Wherever capital or annual value rating applies, it has been imposed by government or local councils, contrary to the express wish of the ratepayers in almost every case”. (p427)
Legislative changes, including a failed central government bill in 1990 that would have made capital-value rating irreversible where adopted, have been intended to sideline the principles of local choice that were central to the adoption of land-value rating a century earlier. Trends in Victoria are similar: of the 90 plebiscites over rating systems undertaken between 1920 and 1986, 70 were in favour of site-value taxation. And as with the legislative changes in New Zealand, in Victoria the laws enabling residents to force a change in ratings bases via plebiscite, and requiring local governments to switch systems only after such a poll, were dropped (Putland 2008).

Under this view, the trends described in local government financing are another manifestation of the same interest group politics, favouring those with large wealth holdings in land, that fought and delayed the imposition of land taxes in the late 19th century. While organised landowner opposition to state and central government land taxes has always been overt – Reece (1993, p236-238) describes the real estate lobby’s influence in New Zealand’s land tax repeal, for instance, and the various property lobbies in Australia regularly push for the abolition of land tax – the trends in local government choice are also the result of a fundamentally similar but less visible process, according to the commentators cited above. In this view, the disappearance from public discussion of the ideas of the Mills and Henry George regarding the economic and ethical case for land taxes has enabled the decline of an organised counterweight to landowning interests. This imbalance, combined with other obviously important trends (such as the diminishing importance and more egalitarian distribution of land wealth today, as compared to a century ago) is the fundamental reason for the quiet shift away from land taxation and land-value rating seen over the last two decades.

7.3.3 The USA and the UK

Property taxes in the United States have long been (and remain) a more significant source of revenue for local governments than their equivalents in Australia and New Zealand, and naturally changes in the form of these taxes have been more politically contentious. While no jurisdictions in the USA
levy property taxes on land alone, there have been experiments in implementing ‘split-rate’ taxes – where land is taxed at a higher rate than improvements – in parts of Pennsylvania and Hawaii. The experience of these states shows that assessment issues – particularly timeliness, and public confidence in the accuracy of assessments – are crucial to maintaining support for the more complicated system of split-rate taxes over the traditional property tax.

Faster development of bare land was the motivation for exempting improvements in Australia and New Zealand; in Pennsylvania, the anticipated spur to redevelopment of blighted inner-city sites drove decisions to adopt split-rate taxation. Twenty-three jurisdictions in Pennsylvania (mostly cities) have adopted split-rate taxation at some point since the 1970’s, primarily to encourage building investment and thereby reverse the decline brought about by the loss of manufacturing industries (Bourassa 2009, p14). The best known of these experiments was in Pittsburgh. In the early twentieth century Pittsburgh (and another large city, Scranton) had actually adopted a split-rate tax, a change driven by supporters of Henry George. In 1979, to encourage inner-city revitalisation, this ratio was sharply increased so as to tax land at about five times the rate of improvements.24 This sharp rate differential remained for two decades, and is cautiously credited by Oates and Schwab (1997) for enabling a construction boom throughout the 1980’s that marked Pittsburgh out from a number of other similar ‘rust-belt’ cities, all of which saw declining building activity over the same period.

Pittsburgh’s split-rate tax was suddenly repealed in 2000-01, having been made a scapegoat for deficiencies in assessment and in rate-setting procedures. While land and property value reassessments in Australia and New Zealand are typically issued annually or biannually, reassessments in the USA occur infrequently and inconsistently. An overdue reassessment in 2000-01 resulted in large increases in tax bills, compounded by perceptions from residential taxpayers that commercial landowners were successfully avoiding tax by shifting property values away from land and onto buildings, via appeal to

24 The overall land-improvements tax rate ratio faced by owners of land in Pittsburgh city was not so high as this, however. Taking into account county-level and school-district property taxes, which were levied equally on land and improvements (the norm in the USA), the ratio was closer to 2 to 1 (Oates and Schwab 1998, p137).
the courts (Bourassa 2009, p16). Similarly clumsy assessment practices led to repeal of a split rate system in the city of Amsterdam, New York, after only one year of operation. After 36 years since the previous assessment a reassessment was scheduled for the same year as the split-rate introduction, and the redistributive effects of the reassessment were blamed on the new tax system (Bourassa 2009, p21). Similar assessment issues have plagued introduction of split-rate taxation elsewhere in Pennsylvania, leading to repeal in several cities. The experiences of these US jurisdictions with a limited style of land-value taxation, in contrast with the history of land taxes in Australia and New Zealand, seems to support the notion that “the only good tax is an old tax”: when a new and poorly understood tax change causes redistribution (and particularly where careless implementation makes the redistribution appear arbitrary), political resistance is likely to scuttle the reforms.

The history of attempts at land-value taxation in the UK, with its long-established landowner class, illustrates this point further. Connellan (2004, p59) argues that the main reason for the failure of various attempts to introduce land taxation was “a lack of political willpower in the face of opposition from various professional groups and landowners, each with their own taxation agendas”. Over the course of the twentieth century a number of legislative bills, including from municipal authorities wishing to adopt land-value rating, were introduced but failed to pass, and the Finance Acts of 1910 and 1931 did allow for land-value taxation but the powers were never used (Connellan 2004). A succession of committees in the post-WWII period also failed to establish a consistent position on land taxation, and on the appropriate relationship with local rates and with the development (‘betterment’) charges which were allowed for under various pieces of post-war legislation.

As well as the landowner opposition Connellan mentions, the perception that the practical obstacles were too large further contributed to legislative inertia. For instance, several of the committees tasked with investigating site-value rating (particularly the 1952 Simes committee and the 1976 Layfield report) assumed that the practical difficulties inherent in valuing land were insurmountable, despite extensive international experience in land valuation
by that time. Several local experiments in land-value rating had also shown that the process of assessment could be undertaken at the same or even a lower cost than under the traditional rating system (Connellan 2004, p56-58). Prest (1981, p102-3) summarises this disconnect by stating that “it is a salutory reflection that none of the previous participants in a discussion which has raged on and off for a century had taken the trouble to see whether, and if so how, the system could be applied in the field”.

7.3.4 Summary

Under each of the themes of efficiency and economic development, fairness and political viability, and practical difficulties, a number of points arise from this historical survey.

The idea that exempting improvements from property taxation encourages development has been an important and influential driver of the adoption of land-value taxation. This is evident from the stated motivations of policymakers in the Australian and New Zealand colonies, and in the run-down cities of Pennsylvania. However the efficiency properties of land taxation vis-à-vis other taxes, such as the income tax, have apparently not been seen as important enough to warrant its use other than by local governments; nowhere do recurrent taxes on land form a major part of central government revenues.

Several classical economists argued that land rent ought to be taxed specially because of its ‘uneared’ nature. While this may have had some influence on policymakers in the nineteenth and early twentieth centuries, the idea evidently exerts no influence over tax design and is not widely supported by the public today. Special taxes on land are instead considered discriminatory (in the pejorative sense) because they leave non-land forms of wealth untaxed.

However the idea that uplift in land values – particularly where caused by some identifiable public expenditure – ought to belong to society and be used to finance government was clearly central to the early land taxes, and continues to find acceptance. Methods of capturing windfall increases in land
values other than by the imposition of recurrent land-value taxes, such as via development charges, betterment taxes, and other forms of ‘value capture’ (which have not been discussed here) are still widely used.

In local government finance there appears to be evidence of an increasingly-common perception that the benefits of public expenditure are not captured by landowners in particular, but are shared by residents at large. Whether this is a reality somehow connected to changes in the type of public goods provided, or reflects a lack of understanding of how public benefits are capitalised into land values is not clear. Whatever the reason, this view seems to have eroded public sentiment towards special taxes on land and contributed to the shift away from land-value rating in Australia and New Zealand.

This changing view of benefit in local-government taxation, in conjunction with a view that property taxes ought to be levied on ‘ability-to-pay’, not ‘benefits-received’, seems to have caused a shift in perceptions of fairness in local government taxation. This shift has driven capital-value rating, user charges, differential rates by land type, and poll tax measures, all in place of land-value rating (and even in place of pure ad valorem property taxation) as the financing mechanism for local government.

The ideas of Ricardo, the Mills, Marshall and other early economists who stressed the unfairness and confiscatory nature of taxes on existing landholdings, particularly in countries with well-established property rights in land, seem to have retained their force as a theory predicting the difficulties of political change towards greater land taxation. Clearly, the confiscatory nature of new taxes on land wealth mean that efforts to implement them will be resisted. Even in the new antipodean colonies the emergent landowning minority successfully resisted taxation for some time, while landowning lobbies openly seek to exert influence still.

The experiences of Pittsburgh and other US cities, as well as of New Zealand and the Australian states (with regards to eroding support for their land taxes), is that timely and reliable assessment of land and building values is crucial for perceptions of the fairness and hence political viability of land-value taxation. The history of these taxes suggests that a hint of unfairness in assigning tax liabilities can bring about significant public mistrust in a tax
for which the methods of operation and the underlying rationale are poorly understood.

7.4 Conclusion

The historical surveys above have highlighted the limited importance of the bare efficiency arguments for land taxes, both in the historical debates over the desirability of the tax and in the practice of policymaking from the 19th century onwards. Broadly, the discussion of land taxes by the early economists – from Adam Smith’s time until the 20th century – has been a conversation not about the properties of the tax, for the basic efficiency argument was well agreed-upon, but about the justice and feasibility of the tax and how a transition towards land taxation could be managed. In practice, even in those times and places where the economic advantages of taxing land were well-recognised (relative to today), the economic arguments have often proved secondary in policymaking priorities to concerns about fairness in burden and administration. The history of the decline in use of land taxation suggests that these perceived unfairnesses are powerful counterinfluences to the strength of efficiency-based arguments in determining the general popularity of land taxes.
Chapter 8

Conclusion

The three parts of this thesis, as explained in the introductory chapter, are linked by a common motivation: to better explain the causes and understand the consequences of price adjustment processes. For a variety of scientific and policy-relevant questions where transmission of price shocks plays a central part, improving our understanding of why markets adjust as they do and what this means for economic outcomes is an important goal.

The thesis has not attempted to consider the economics of price pass-through and market adjustment dynamics comprehensively. Rather, a focus on price transmission was the common element in a selection of case studies on distinct questions.

As such the thesis offers no overarching conclusion. Rather, the nature of the contribution is to provide within the case studies a variety of smaller insights, modelling innovations, suggestive theoretical predictions, syntheses of lines of enquiry, new arguments or novel ‘re-packaging’ of old ideas, and clarifications to (or clearer exposition of) existing arguments. These original aspects of the thesis and where they lie relative to the questions of interest in each case study are summarised here.

Part I, on asymmetric price transmission, sought to further one side of the research agenda laid out in Meyer and von Cramon-Taubadel’s (2004) review. With an existing body of theory on asymmetric price adjustment described as a "bouquet of often casual explanations", and providing little basis for empirical tests to distinguish between potential causes of asymmetry,
these authors argue the importance of theoretical work which draws out more precisely what form (or pattern) and what extent of asymmetry in price responses we expect to see associated with each of the various explanations for asymmetry.

The particular focus of Part I was on Peltzman’s (2000) hypothesis, in which price asymmetry is a natural reflection of the asymmetric costs of output adjustment resulting from differences in ‘hiring and firing’ costs for labour inputs, capital (capacity) levels, or material inputs to production. Part I aimed to estimate what degree of asymmetry would ‘reasonably’ reflect plausible asymmetries in adjustment costs in a quantity-setting oligopolistic market, typical of the supply-chain intermediaries on which much interest in asymmetric pass-through centres.

Motivated by the utility of the ‘equilibrium displacement modelling’ approach in understanding the determinants of long-run price transmission elasticities through static modelling, Chapter 2 sought to model the dynamics of adjustment-cost driven price adjustment in a way that incorporated the same key feature. The idea was to construct a model solely in terms of (potentially-quantifiable) elasticity parameters, thus allowing the influence of different parameters to be assessed via numerical exercises using plausible parameter values.

This modelling approach, chosen for a specific goal – to estimate the degree of asymmetry in price responses, and the dependence of this on market characteristics including perceptions of price-making power – was the novel aspect of Chapter 2. The chapter usefully highlighted some difficulties in achieving this goal: in particular, the treatment of market power (which in a dynamic model must be endogenous), and the difficulties of solving the model when functional forms are not linear. While perhaps rendering the model’s predictions ‘suggestive’ rather than ‘definitive’, these difficulties also point towards both opportunities for improvement and the obstacles to doing so. The particular challenge raised is that of modelling firm behaviour with the usual game-theoretic norms, whilst also giving the model predictive content beyond simply describing directions of influence of parameters. The question of interest, after all, is about how much asymmetry in price responses is
expected in any given setting.

The contribution of Chapter 3 was complementary. Since the simplest specification of convex adjustment costs (the quadratic cost function) does not, by itself, contain an economically-meaningful adjustment cost parameter, Chapter 3 suggested two functional forms with easily-interpreted parameters (which could also potentially be estimated from data). In the first functional form, which was used in the numerical exercises of Chapter 4, the parameter scales the percentage rate of change of output into the percentage ‘cost premium’ on new production that adjustment costs effectively impose. Equivalently, this parameter represents an elasticity of the average cost of new units of output (including adjustment costs) with respect to the rate of change of output. The adjustment cost parameter in the second function was similar, representing the elasticity of marginal cost with respect to the rate of change of output. This original and intuitive framing of the adjustment cost burden proved useful for the modelling exercise of Part 1 where a quantifiable parameter was required, and it may potentially find useful application elsewhere.

Chapter 4 contained two distinct contributions. The central task was to draw out several predictions about the nature of adjustment-cost driven asymmetry from the model developed in Chapter 2. An interesting finding that (if it were verified elsewhere) could provide an important stylised fact for empirical studies was that the predicted degree of asymmetry appeared relatively invariant to characteristics of the firm and market other than the asymmetry of adjustment costs. That is, in quantitative terms the adjustment-cost asymmetry – pricing asymmetry relationship remained quite steady: a cost asymmetry factor of two generated asymmetry of 0.15-0.20 (under two of the main measures), whatever the nature of the market. Notably, the degree of asymmetry in price responses did not increase proportionately as the degree of asymmetry in costs was increased.

A second contribution of Chapter 4 spoke to another line of enquiry promoted by Meyer and von Cramon-Taubadel (2004). These authors suggest greater use of cross-sectional studies to search for correlates of the degree of asymmetry, which might then point towards the likely drivers of asymmetry.
The only major study to have done this already (Peltzman 2000) reached some interesting conclusions, including no evidence of greater asymmetry in more concentrated markets (contra the dominant market power-based theories).

Chapter 4 argued an important point not yet raised in the literature, that the standard quantification of asymmetry used by Peltzman and most other researchers – the difference between positive and negative price responses – is ill-suited for use in multi-industry studies, because it biases measured asymmetry according to the size of the long-run price transmission elasticity. Market power (in particular) reduces the long-run pass-through of shocks, and hence reduces the potential difference between positive and negative responses; failure to adjust for this will thwart any attempts to uncover a general correlation across industries between competitiveness and asymmetry. Chapter 4 offered a range of alternative measures of asymmetry without this drawback, using them to illustrate the predictions of the model of Chapter 2.

Part II touched on some of the modelling concepts of Part I, despite beginning with different aims. The concern of Chapter 5 was the relationship between market power in an industry and the speed of adjustment to shocks, and the question of whether (and how) dynamic oligopoly models might be used to examine this.

An apparently-promising approach to exploring this question is to incorporate into dynamic models an ‘exogenous conjectural variations’ parameterisation of market power, reflecting the common use of conjectures to represent market power in static price transmission models and econometric work. At least two previous authors – Worthington (1989) and Martin (1993) – have adopted this approach. As Chapter 5 argued, however, the methodological soundness of conjectures within dynamic models depends crucially on how they are incorporated. Treating conjectures as exogenous ‘beliefs’, as in those two papers, was argued to be conceptually problematic (an issue also acknowledged to apply to the model of Chapter 2).

In addition to this critique, Chapter 5 offered a second (and more constructive) contribution, demonstrating a method by which the goals of Wor-
thington’s and Martin’s analyses might be realised free of these methodo-
logical problems. Logically, any model-based analysis of adjustment speeds
and market power requires a dynamic set-up (to examine the former), and a
measure of market power (to represent the latter). Since an exogenous pa-
parameterisation of market power poses problems, Chapter 5 suggested a good
approach to marrying these two requirements would be to use the alterna-
tive understanding of static conjectural-variations parameters, as ‘shorthand’
representations of the competitiveness of the outcome of a dynamic model.

The suggested approach exploits the joint reliance of the endogenous
conjectural-variation parameter (which represents the competitiveness of the
dynamic steady state), and the endogenous adjustment speed parameter, on
the underlying parameters of the model. By observing how these two vari-
ables of interest move as underlying parameters change, insight might be
gained into the market power – adjustment speed relationship as it holds
within each class of model.

This was demonstrated using Dockner’s (1992) model of an adjustment-
cost duopoly. More competitive outcomes were shown to be associated with
faster adjustment to equilibrium in this model, at least as the discount rate
and adjustment cost parameters were varied. Significantly, however, the
relationship as driven by a third exogenous parameter (the convexity of pro-
duction costs) was in the reverse direction, indicating that even within one
model the relationship depends on the underlying factors causing the market
to exhibit more or less competitive outcomes.

Chapter 5 brought together two previously-unconnected streams of work
– the emerging literature relating dynamic models to static conjectural vari-
ations, and the theoretical enquiry into the influence of market power on
adjustment speeds. The idea that the latter might benefit from the insights
of the former is original, and the method demonstrated here offers a promis-
ing approach to exploring the market power – adjustment speed relationship
and its dependence on the nature of strategic interactions between firms. A
range of existing models might potentially be analysed with this approach,
although the indeterminacy of the relationship even within the single model
examined perhaps indicates that the search for general tendencies in the re-
relationship might best be undertaken through empirical work, rather than theory.

Part III considered the properties of land-value taxes. The central result that such a tax is not passed through to prices for land use is long-established, but with implications no less striking for that fact. It lies behind the major efficiency advantages of land tax over other taxes (and also, paradoxically, the substantial political obstacles to greater use of the tax). The chapters of Part III explored this pass-through result (i.e. the argument for the neutrality of the tax), in order to better understand its premises, its implications, and its relevance in the policy debates of economists and lawmakers over time.

The review of the neutrality argument in Chapter 6 offered a comprehensive summary of the existing literature, as well as a collection of (minor) original arguments and expositional points arising from that work. In particular, in explaining the ways neutrality is typically argued and the objections and exceptions often raised, a logical framework for understanding the neutrality of the tax and classifying the various situations of non-neutrality was proposed. By asking simply whether landowners may avoid some part of the tax burden by either selling their land asset, or by influencing its assessed value, the neutrality result can be seen to hold generally, and to fail only in three particular circumstances. These are when different potential owners value the expected tax liabilities differently, when land value is incorrectly assessed by reference to the development in place on the site, or when the true value is amenable to the owner’s influence in various specific ways. The simplest way to explain the efficiency of land-value taxes thus need not invoke rent, or inelastic land supply.

Chapter 6 also considered another ‘framing’ of the properties of a land-value tax. Introducing (or increasing) the tax is effectively equivalent to appropriating a share of private equity in land assets, and a corollary of this is that a share of private risk is also transferred to the public purse, an idea examined in some detail here. Although the concept of land-value tax as a risk-pooling measure has been noted before, the argument and its implications have not been studied in any depth. Chapter 6 explored how this result arises and what it implies, and argued that the risk-pooling properties
of land taxation differ fundamentally from those inherent in redistributive tax and expenditure systems. The chapter concluded that in evaluating the policy case for land tax, the risk-pooling idea is not only a useful tool to comprehend the nature of the tax, but also suggests a benefit of the tax quite distinct from the general efficiency advantages it offers.

Finally, Chapter 7 reflected on the relative importance of the neutrality argument, as seen in the writings of classical economists and their positions on the merits of the tax, and in the reasons for the adoption and erosion of land taxes in various places and times throughout the twentieth century. This review considered these arguments and historical developments from a somewhat novel perspective: the question of interest considered throughout was about the comparative importance of economic efficiency arguments (or those that would be labelled as such today), as against claims and theories of equity and justice, or the influence of interest groups and other political forces.

Where Chapter 6 sought to better understand the efficiency logic for land taxes, Chapter 7 sought to better understand the origins and the historical impact of this logic. The coverage of thought and policy was not exhaustive, but was sufficient to illustrate clearly the limited relevance of bare efficiency arguments in the reasoning of intellectuals, and in the development of policy through time. In the debates of the classical economists and in the introduction and abandonment of land taxes it has been views of fairness, not efficiency, that have been dominant, a conclusion that puts the contributions of modern economics on the topic into a somewhat different perspective.

The work in Part III prompts the idea that the development of policy proposals sensitive to all considerations – efficiency, equity, and political – is likely to be the most productive areas for economists to direct future research efforts towards. The efficiency case for land taxation is clear and well-understood (although quantifying efficiency benefits and communicating this case to policymakers remains a difficult and important task). But historical and contemporary experience shows that this case is usually trumped in political decision-making by perceived (or actual) issues around the fairness of the tax. The question of how a transition towards land-value taxation
could be managed to negate these issues, without sacrificing all potential revenue from the shift, thus remains one of the most important challenges facing economists and policy analysts working in this area.
Appendix A

Solving the dynamic and static games of Chapter 2

Interpretation of $\theta$ as a market power parameter

The steady state price level $p^*$ – actual, not conjectured – is given by:

$$ p^* = p(Q^*) = p(Nq_i^*) $$

where $Q^*$ is the steady state industry quantity, and $q_i^*$ is the steady state firm quantity. The latter is implicitly determined by equation (2.11) and is therefore dependent on the value of $\theta$.

To show that higher values of $\theta$ result in a larger markup involves demonstrating that, for any $\theta' > \theta$:

$$ \frac{p(Nq_i^{*\prime})}{c} > \frac{p(Nq_i^{*})}{c} $$

where $q_i^{*\prime}$ and $q_i^{*}$ are the steady state firm level quantities that arise when the parameter values are $\theta'$ and $\theta$ respectively. Given that $p(Q)$ is decreasing in $Q$ and that $N$ and $c$ are fixed, it is sufficient to show that $q_i^{*\prime} < q_i^{*}$.

Substituting $\theta'$ and $\theta$ sequentially into equation (2.11), and equating the two steady state conditions that result, yields:

$$ p(Q^c(q_i^{*\prime})) \left(1 - \frac{\theta'}{\eta}\right) = p(Q^c(q_i^{*})) \left(1 - \frac{\theta}{\eta}\right) $$

(A.3)
or equivalently:

\[
\frac{p(Q^c(q^*)^\prime)}{p(Q^c(q^*))} = \frac{1 - \frac{\theta}{\eta}}{1 - \frac{\theta^\prime}{\eta}} = \frac{\eta - \theta}{\eta - \theta^\prime}
\] (A.4)

The ratio on the right hand side is greater than 1 when \(\theta^\prime > \theta\), implying that \(p(Q^c(q^*)^\prime)) > p(Q^c(q^*))\). Given that \(p(Q)\) is decreasing in \(Q\), and that \(Q^c(q_i)\) is increasing in \(q_i\) (so long as \(\theta\) is positive), \(q^*_{\theta^\prime}\) will be smaller than \(q^*_{\theta}\). That is, when the value of the conjecture \(\theta\) is larger the steady state … rm level quantity and the industry quantity are smaller, resulting in a higher steady state price and higher markup over cost. It is appropriate, therefore, to interpret \(\theta\) as a market power parameter even though the exact condition relating the steady state markup to \(\theta\) is not known.

**Linearisation of equation system (2.10)**

The equation system (2.10) describes the dynamics of the two variables \(x_i\) and \(q_i\) (which are closely linked since \(x_i = \dot{q}_i\)). To solve for the adjustment path \(q_i(t)\), the equation system is first simplified by way of linearisation (deriving a linear approximation around the steady state value defined jointly by \(x^*_i = 0\) and by equation (2.11)). This approximation is given by the first order Taylor series expansion:

\[
\begin{align*}
\dot{x}_i & \approx g^1_{x_i}(x^*_i, q^*_i) (x_i - x^*_i) + g^1_{q_i}(x^*_i, q^*_i) (q_i - q^*_i) \\
\dot{q}_i & \approx g^2_{x_i}(x^*_i, q^*_i) (x_i - x^*_i) + g^2_{q_i}(x^*_i, q^*_i) (q_i - q^*_i)
\end{align*}
\] (A.5)

where the subscripts indicate partial derivatives, and where the \(g\) functions are given by:

\[
\begin{align*}
g^1(x_i, q_i) &= \frac{1}{A_{x_i}} \left( rA_{x_i} + A_{q_i} - A_{x_iq_i} x_i - p^c \left( 1 - \frac{\theta}{\eta} \right) + c \right) \\
g^2(x_i, q_i) &= x_i
\end{align*}
\] (A.6)

The first partial derivative of \(g^1\) (with respect to \(x_i\)) is:

\[
\begin{align*}
g^1_{x_i}(x_i, q_i) &= -\frac{A_{x_i}}{A_{x_i}^2} \left( rA_{x_i} + A_{q_i} - A_{x_iq_i} x_i - p^c \left( 1 - \frac{\theta}{\eta} \right) + c \right) \\
&\quad + \frac{1}{A_{x_i}} \left( rA_{x_i} + A_{q_i} - A_{x_iq_i} x_i - A_{x_iq_i} \right)
\end{align*}
\] (A.7)
Evaluated at steady state, the first bracketed term disappears, since this term is just the steady state condition and since $A_q$ and $A_x$ are assumed to take zero value when $x_i = 0$. In the second bracketed term $A_{q_i x_i}$ and $A_{x_i q_i}$ cancel each other so at the steady state this term reduces to $r A_{x_i x_i}^*$ and the whole expression becomes:

$$g_{x_i}^1(x_i^*, q_i^*) = r$$  \hspace{1cm} (A.8)

The second partial derivative of $g^1$ (with respect to $q_i$) is more complex and is derived below.

Note first that since the demand curve may not have constant elasticity, the (absolute value of the) elasticity of demand, $\eta$, changes with movements in the price level. Parameter $\omega$ is defined as the proportional change in $\eta$ as $p$ changes, $\omega = \frac{d\eta}{dp} \eta$. The value of $\omega$ depends on the form of the demand curve ($\omega = 0$ for a constant-elasticity demand curve, $\omega = 1$ for a semi-logarithmic demand, and $\omega = 1 + \eta$ with linear demand). Recalling that $Q = N q_i$ along the adjustment path, the derivative of $\eta$ with respect to $q_i$ is thus:

$$\frac{d\eta}{dq_i} = \frac{d\eta}{dp} \frac{dQ}{dq_i} = \frac{\eta}{p} \left( \frac{dQ}{Q} \right) \frac{dQ}{dq_i} = -\frac{\omega N}{q_i}$$  \hspace{1cm} (A.9)

Therefore the partial derivative of $g^1$ with respect to $q_i$ is given by the following (with the substitution $\frac{d\eta}{dq_i} = -\frac{\omega}{q_i}$ made):

$$g_{q_i}^1(x_i, q_i) = -\frac{A_{x_i x_i}}{A_{x_i x_i}^2} \left( r A_{x_i} + A_{q_i} - A_{x_i q_i} x_i - \frac{dp^c}{dq_i} \left( 1 - \frac{\theta}{\eta} \right) + c \right)$$

$$+ \frac{1}{A_{x_i x_i}} \left( r A_{x_i q_i} + A_{q_i q_i} - A_{x_i q_i q_i} x_i - \frac{dp^c}{dq_i} \left( 1 - \frac{\theta}{\eta} \right) - \frac{\omega}{q_i} \right)$$  \hspace{1cm} (A.10)

Substituting $\frac{dp^c}{dq_i} = -\frac{\theta}{\eta} \frac{p^c}{q_i}$ (from equation (2.6)) the second bracketed term becomes:

$$r A_{x_i q_i} + A_{q_i q_i} - A_{x_i q_i q_i} x_i + \frac{\theta}{\eta} \frac{p^c}{q_i} \left( 1 - \frac{\theta}{\eta} \right) + \frac{\omega}{q_i} \left( \frac{\theta}{\eta^2} \right)$$

$$= r A_{x_i q_i} + A_{q_i q_i} - A_{x_i q_i q_i} x_i + \frac{\theta}{\eta} \frac{p^c}{q_i} \left( \frac{\eta - \theta + \omega}{\eta} \right)$$  \hspace{1cm} (A.11)
Evaluated at the steady state, the first bracketed term disappears, as before. Substituting $p^c = \left( \frac{n}{\eta - \theta} \right) c$ (the steady state expression) into the second bracketed term, the partial derivative becomes:

$$
g^1_{qi}(x^*_i, q^*_i) = \frac{1}{A^*_x} \left( rA^*_{x,qi} + A^*_{q,qi} + \frac{c}{q^*_i} \left( \frac{\eta}{\eta - \theta} - \frac{\eta - \theta + \omega}{\eta} \right) \right)
= \frac{1}{A^*_x} \left( rA^*_{x,qi} + A^*_{q,qi} + \frac{c}{q^*_i} (1 + \frac{\omega}{\eta - \theta}) \right) \quad (A.12)
$$

The partial derivatives of $g^2$ are $g^2_{x_i}(x_i, q_i) = 1$ and $g^2_{q_i}(x_i, q_i) = 0$.

The linearised equation system is thus:

$$
\begin{pmatrix}
\dot{x}_i \\
\dot{q}_i
\end{pmatrix} =
\begin{bmatrix}
r & K \\
1 & 0
\end{bmatrix}
\begin{pmatrix}
x_i - x^*_i \\
q_i - q^*_i
\end{pmatrix}
\quad (A.13)
$$

where

$$
K = \frac{1}{A^*_x} \left( rA^*_{x,qi} + A^*_{q,qi} + \frac{c}{q^*_i} (1 + \frac{\omega}{\eta - \theta}) \right).
$$

**Stability properties of the linearised system**

The determinant of the matrix is $-K$, but without an explicit formula for the adjustment cost function the stability properties of this system are difficult to establish. Proceeding under the assumption that the partial derivatives $A_{x,qi}$ and $A_{q,qi}$ take zero value at the steady state output level, it is clear that the stability properties are entirely determined by the term $\frac{1}{A^*_x} \frac{c}{q^*_i} \frac{\theta}{(\eta - \theta)}$.

Therefore as long as $A$ is convex in the variable $x_i$ at the steady state ($A^*_{x,x_i} > 0$), the sign of $1 + \frac{\omega}{\eta - \theta}$ determines whether the system is saddlepath stable or globally unstable. A sufficient condition for $1 + \frac{\omega}{\eta - \theta}$ to be positive and the determinant to be negative, implying saddlepath stability, is that $\eta > \theta$. Even when this does not hold, it is necessary that the elasticity of demand $\eta$ fall within the range $\theta - \omega < \eta < \theta$ for $1 + \frac{\omega}{\eta - \theta}$ to be negative and the system to be unstable.

Instability thus requires, firstly, that demand be highly inelastic and the market be very uncompetitive (the right hand side inequality). The left

---

1Note that this assumption is satisfied by the adjustment cost function introduced in Chapter 3 and used in the numerical exercises in Chapter 4.
hand side inequality says that even in such situations a certain curvature of the demand curve is required – it must be flatter to some extent than a constant-elasticity demand curve ($\omega > \theta - \eta$ must hold).

Note also that for any adjustment cost function where $A_{x_i q_i}^* = A_{q_i q_i}^* = 0$ fails, the stable range of parameters will be different to that calculated here. So assuming that stability holds (the most obvious scenario to achieve this being that partial derivatives $A_{x_i q_i}$ and $A_{q_i q_i}$ take zero value at the steady state output level and the demand curve is elastic enough that $\eta > \theta$ holds), $K$ is positive and the adjustment path that is the solution to the linearised equation system is saddlepath stable.

**Solution to the linearised equation system**

The eigenvalues $\mu_1$ and $\mu_2$ of the matrix in equation (2.12) are given by the solutions to the characteristic equation:

$$c(\mu) = -\mu(r - \mu) - K$$

which are:

$$\mu_1 = \frac{1}{2} \left( r - \sqrt{r^2 + 4K} \right) < 0$$

$$\mu_2 = \frac{1}{2} \left( r + \sqrt{r^2 + 4K} \right) > 0$$

and which have the signs indicated, so long as saddlepath stability holds (as discussed above, $K > 0$ is the condition for saddlepath stability).

Using an eigenvector representation of the solution:

$$\begin{pmatrix} x_i(t) \\ q_i(t) - q_i^* \end{pmatrix} = \begin{bmatrix} \mu_1 & \mu_2 \\ 1 & 1 \end{bmatrix} \begin{pmatrix} B_1 e^{\mu_1 t} \\ B_2 e^{\mu_2 t} \end{pmatrix}$$

the output and rate of change of output timepaths for firm $i$ are:

$$q_i(t) = q_i^* + B_1 e^{\mu_1 t} + B_2 e^{\mu_2 t}$$

$$x_i(t) = \mu_1 B_1 e^{\mu_1 t} + \mu_2 B_2 e^{\mu_2 t}$$
The requirement of convergent timepaths in a saddlepath stable system dictates that the coefficient \( B_2 \) on the positive eigenvalue \( \mu_2 \) be set to 0, and the initial value of the firm’s output \( q_i(0) \) determines the other coefficient \( (B_1 = q_i(0) - q_i^*) \). The convergent solution is thus:

\[
x_i(t) = \mu_1 (q_i(0) - q_i^*) e^{\mu_1 t}
\]
(A.20)

\[
q_i(t) - q_i^* = (q_i(0) - q_i^*) e^{\mu_1 t}
\]
(A.21)

Substituting the right hand side of the second equation into the first, and replacing \( x_i(t) \) with \( \dot{q}_i(t) \), firm \( i \)'s output can be seen to follow a ‘partial adjustment’ process with adjustment coefficient \( \lambda \), where \( \lambda = -\mu_1 \):

\[
\dot{q}_i(t) = \lambda (q_i^* - q_i(t))
\]
(A.22)

That is, the change in output over a period is a constant proportion of the difference between steady state output and current period output.

**Equilibrium displacement model working**

Equations (2.11), (2.18), and (2.20) to (2.24) constitute a system of 7 equations in 7 endogenous variables – \( q_i \), \( q_i^A \), \( q_i^M \), \( p \), \( c \), \( p_c \), and \( Q_c \) – as shown below:

\[
p^c \left( 1 - \frac{\theta}{\eta} \right) = c
\]
(A.23)

\[
q_i = f(q_i^A, q_i^M)
\]
(A.24)

\[
p = p(N q_i)
\]
(A.25)

\[
\bar{p}^A = c_f A
\]
(A.26)

\[
\bar{p}^M = c_f M
\]
(A.27)

\[
p^c = p(Q_c)
\]
(A.28)

\[
Q_c = Q_c(q_i)
\]
(A.29)

Each equation is totally differentiated and expressed in logarithmic form. It is assumed the changes are small, so that the elasticity parameters may
be assumed constant, and so that the proportional and log changes are approximately equal, i.e. \( dX/X \approx d\ln X \) for each variable \( X \).

The first equation (the steady state condition relating marginal cost to the conjectured price, which implicitly defines the firm’s quantity choice in equilibrium) becomes:

\[
\begin{align*}
   dp^c \left( 1 - \frac{\theta}{\eta} \right) + \frac{p^c \theta}{\eta^2} d\eta &= dc \\
   \frac{dp^c}{p^c} \frac{p^c}{c} \left( 1 - \frac{\theta}{\eta} \right) + \frac{p^c}{c} \frac{\theta}{\eta^2} \frac{d\eta}{dp^c} \frac{dp}{p} \frac{p}{\eta} &= \frac{dc}{c} \\
   d\ln p^c + \left( \frac{\omega \theta}{\eta - \theta} \right) d\ln p &= d\ln c \quad (A.30)
\end{align*}
\]

The second equation (the production function) becomes:

\[
\begin{align*}
   d\ln q_i &= \alpha d\ln q_i^A + \beta d\ln q_i^M \quad (A.31)
\end{align*}
\]

The third equation (the relationship between firm \( i \)'s output and the price, via the inverse demand function) becomes:

\[
\begin{align*}
   d\ln p = -\frac{1}{\eta} d\ln q_i \quad (A.32)
\end{align*}
\]

The fourth and fifth equations (the input cost minimisation conditions, with the change in the exogenous agricultural input price \( \tilde{p}^A \) constituting the shock to the system of equations, and assuming no change in \( \tilde{p}^M \)) become:

\[
\begin{align*}
   d\ln \tilde{p}^A &= d\ln c - \frac{\beta}{\sigma} \left( d\ln q_i^A - d\ln q_i^M \right) \quad (A.33) \\
   0 &= d\ln c + \frac{\alpha}{\sigma} \left( d\ln q_i^A - d\ln q_i^M \right) \quad (A.34)
\end{align*}
\]

The sixth equation (the relationship between the conjectured value of industry output and the conjectured value of the price, via the demand curve) becomes:

\[
\begin{align*}
   dp^c &= \frac{\partial p(Q^c)}{\partial Q^c} dQ^c \\
   \frac{dp^c}{p^c} &= \left( \frac{\partial p(Q^c)}{\partial Q^c} \frac{Q^c}{p(Q^c)} \right) \frac{p(Q^c)}{p^c} dQ^c \\
   d\ln p^c &= -\frac{1}{\eta} d\ln Q^c \quad (A.35)
\end{align*}
\]
The seventh equation (the relationship between the conjectured value of industry output and the actual value of the firm’s output) becomes:

\[
\frac{dQ^c}{c} = \frac{\partial Q^c}{\partial q_i} dq_i, \quad \frac{dQ^c}{Q^c} = \left( \frac{\partial Q^c}{\partial q_i} \frac{q_i}{Q^c} \right) dq_i, \quad d \ln Q^c = \theta d \ln q_i \tag{A.36}
\]

The whole equation system expressed in matrix form is:

\[
\begin{bmatrix}
0 & 0 & 0 & \frac{\omega \theta}{\eta - \theta} & -1 & 1 & 0 \\
1 & -\alpha & -\beta & 0 & 0 & 0 & 0 \\
\frac{1}{\eta} & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & -\frac{\beta}{\sigma} & \frac{\beta}{\sigma} & 0 & 1 & 0 & 0 \\
0 & \frac{\alpha}{\sigma} & -\frac{\alpha}{\sigma} & 0 & 1 & 0 & 0 \\
-\theta & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
dl \ln q_i \\
dl \ln q_i^A \\
dl \ln q_i^M \\
dl \ln p \\
dl \ln c \\
dl \ln p^c \\
dl \ln Q^c
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\tag{A.37}
\]

**Derivation of equation (2.32)**

The relationship between the size of the shock to the agricultural input price $d \ln \tilde{p}^A$ and the size of the movement in the firm’s steady state output level $d \ln q_i$ is found via Cramers rule. The latter is given by $d \ln q_i = \frac{\det A'}{\det A}$ where $A$ is the coefficient matrix above and $A'$ is the matrix with the first column replaced by the vector on the right hand side.

The determinants of $A$ and $A'$ are:

\[
\det A = -\frac{1 + \frac{\omega \theta}{\eta - \theta}}{\sigma \eta} \tag{A.38}
\]

\[
\det A' = \frac{\alpha}{\sigma} d \ln \tilde{p}^A \tag{A.39}
\]

Hence:

\[
d \ln q_i = \frac{(\alpha/\sigma)}{-\left(1 + \frac{\omega \theta}{\eta - \theta}\right)/\sigma \eta} d \ln \tilde{p}^A
= \frac{\alpha \eta}{1 + \frac{\omega \theta}{\eta - \theta}} d \ln \tilde{p}^A \tag{A.40}
\]
Appendix B

Parameter effects from Chapter 4

Partial derivatives showing effects on price transmission magnitude and speed

The directional effects of the parameters on adjustment speed and magnitude are given, respectively, by the partial derivatives of the adjustment speed parameter in equation (4.4) and of the elasticity of price transmission expression, \( \tau = \frac{\frac{d \ln p^*}{d \ln p^*}}{\frac{d \ln p^*}{d \ln p^*}} = \frac{\alpha}{1 + \frac{\eta}{\theta}} \). The signs of the partial derivatives below are derived assuming \( \eta > \theta \), which is also a sufficient condition for stability of the model (see Appendix A).

First, it is straightforward to see that increasing the adjustment cost parameter \( \delta \) makes the eigenvalue smaller in magnitude (less negative), and thus slows adjustment to steady state. This parameter has no influence on the magnitude of price transmission.

In contrast, the agricultural input cost share \( \alpha \) has no effect on the speed of price transmission but largely determines the magnitude:

\[
\frac{\partial \tau}{\partial \alpha} = \frac{1}{1 + \frac{\alpha}{\eta - \theta}} > 0
\]

\( \text{(B.1)} \)

A higher discount rate \( r \) has no effect on the magnitude of price transmission, but decreases the speed:

\[
\frac{\partial \lambda}{\partial r} = -\frac{\lambda}{2\lambda + r} < 0
\]

\( \text{(B.2)} \)
APPENDIX B. PARAMETER EFFECTS FROM CHAPTER 4

The influence of the market power parameter on adjustment speed is shown through the partial derivative:

\[
\frac{\partial \lambda}{\partial \theta} = \frac{1}{2\delta \eta (2\lambda + r)} \left( 1 + \frac{\omega \eta}{(\eta - \theta)^2} \right) > 0 \quad (B.3)
\]

Assuming that the parameters fall inside the stable range (\(\eta > \theta\) is sufficient), the derivative is positive, so that adjustment is faster in less competitive industries. Market power also reduces the magnitude of transmission, according to the derivative:

\[
\frac{\partial \tau}{\partial \theta} = \frac{\alpha}{(1 + \frac{\omega}{\eta - \theta})^2} \frac{\omega \eta}{(\eta - \theta)^2} < 0 \quad (B.4)
\]

which is negative for all parameter values (except \(\omega = 0\)).

Higher demand elasticity causes slower adjustment, according to the derivative:

\[
\frac{\partial \lambda}{\partial \eta} = -\frac{1}{2\delta \eta (2\lambda + r)} \left( 1 + \frac{\omega \eta}{(\eta - \theta)^2} \right) < 0 \quad (B.5)
\]

More elastic demand also increases the magnitude of price transmission:

\[
\frac{\partial \tau}{\partial \eta} = \frac{\alpha}{(1 + \frac{\omega}{\eta - \theta})^2} \frac{\omega \theta}{(\eta - \theta)^2} > 0 \quad (B.6)
\]

Finally, the eigenvalue derivative with respect to \(\omega\) is:

\[
\frac{\partial \lambda}{\partial \omega} = -\frac{1}{2\delta \eta (2\lambda + r)} \frac{1}{(\eta - \theta)} > 0 \quad (B.7)
\]

which is positive. The magnitude of price transmission is also affected by the curvature of demand as shown by the derivative:

\[
\frac{\partial \tau}{\partial \omega} = -\frac{\alpha}{(1 + \frac{\omega}{\eta - \theta})^2} \frac{\omega}{\eta - \theta} < 0 \quad (B.8)
\]

When the demand curve is flatter (closer to a linear curve, with \(\omega = 1 + \eta\)) price transmission is faster, but with a smaller long-run magnitude, than when demand is more convex (close to constant-elasticity, \(\omega = 0\)).
Appendix C

Solving the dynamic game of Chapter 5

Derivation of steady state condition (general dynamic game)

The steady state equilibrium for the differential game is found via the use of the Pontryagin necessary conditions. When closed-loop strategies are played, the current-value Hamiltonian, adjoint equations, and maximum condition for player $i$ are (with the time arguments omitted for brevity):

$$H^i = p(Q) q_i - C_i(q_i) - A_i(\dot{q}_i) + \lambda^i_i \dot{q}_i + \lambda^i_j \dot{q}_j$$  \hspace{1cm} (C.1)

$$\dot{\lambda}^i_i = r \lambda^i_i - p'(Q) q_i - p(Q) + C'_i(q_i) - \lambda^i_j \frac{\partial q_j}{\partial q_i}$$  \hspace{1cm} (C.2)

$$\dot{\lambda}^i_j = r \lambda^i_j - p'(Q) q_i - \lambda^i_j \frac{\partial q_i}{\partial q_j}$$  \hspace{1cm} (C.3)

$$A'_i(\dot{q}_i) = \lambda^i_i$$  \hspace{1cm} (C.4)

In steady state, the rate of change of the state variable $q_i$ and of the adjoint variables $\lambda^i_j$ and $\lambda^i_i$ must be 0. With this substitution, the following steady state markup of price over cost can be derived (see Dockner (1992, p385)):

$$p^*(Q^*) \left( 1 + \frac{S_i}{\eta} \left( 1 + \frac{\partial q_i}{r - \frac{\partial q_i}{\partial q_j}} \right) \right) = C'_i(q^*_i)$$  \hspace{1cm} (C.5)
**Derivation of feedback rules (linear-quadratic game)**

The value functions \( V_i(q_i, q_j), i \neq j \) take the state variables as arguments, and must satisfy the Bellman equations (for \( i = 1, 2 \)):

\[
rV^i(q_i, q_j) = \max_{q_i} \{ (a - q_i - q_j)q_i - c_iq_i - \frac{b}{2}q_i^2 - \frac{k}{2}q_i^2 + V^i_{q_i}(q_i, q_j) \dot{q}_i + V^i_{q_j}(q_i, q_j) \dot{q}_j \}
\]  

(C.6)

By maximising the right hand side, the optimal strategies \( \dot{q}_i^*(q_i, q_j) \) are shown to depend on the value function as follows:

\[
\dot{q}_i^*(q_i, q_j) = \frac{1}{k} V^i_{q_i}(q_i, q_j)
\]  

(C.7)

Guessing a linear-quadratic form for \( V^i(q_i, q_j) \), and solving the resulting partial differential equation system (see Dockner (1992, p391)), the equilibrium strategies are found to be given by a *symmetric* feedback rule (despite the asymmetry in the optimisation problems):

\[
\dot{q}_i^*(q_i, q_j) = \frac{1}{k} [\beta_i + \delta q_i(t) + \sigma q_j(t)]
\]  

(C.8)

**Derivation of adjustment path and adjustment speed measure (linear-quadratic game)**

To calculate the path by which the industry output converges to steady state, the optimal strategies in equation (5.19) are summed to give a differential equation describing the motion of the industry quantity:

\[
\dot{Q} = \frac{1}{k} [(\beta_i + \beta_j) + (\delta + \sigma) Q]
\]  

(C.9)

Setting \( \dot{Q} \) to zero in the differential equation yields the steady state output value \( Q^* \), in terms of the coefficients \( \delta, \sigma, \beta_i \) and \( \beta_j \):

\[
Q^* = -\frac{\beta_i + \beta_j}{\delta + \sigma}
\]  

(C.10)

Substituting this steady state expression into (C.9) gives a differential equation describing the change in industry quantity as a fixed proportion of
the discrepancy between the current and steady state output values:

\[ \dot{Q} = \frac{-\delta + \sigma}{k} (Q^* - Q(t)) \]  \hspace{1cm} (C.11)

Since the demand curve is linear, this can be simply rearranged into a continuous-time version of a ‘partial adjustment model’:

\[ \dot{p} = \frac{-\delta + \sigma}{k} (p^* - p(t)) \]  \hspace{1cm} (C.12)
Bibliography


[47] Fallow, Brian (2009), "Tinkering won’t fix creaking tax system", *NZ Herald*, 3 December


[77] Hendriks, Dave (2005), "Apportionment in Property Valuation: Should We Separate the Inseparable?", *Journal of Property Investment and Finance* 23(5): 455-470


[92] MacroPlan (2009), City of Monash Rating and Revenue Strategy Consultation, 20 August


[111] NZ Treasury (2009), "Other Base Broadening and Revenue Raising Ideas", Background Paper for Session 3 of VUW Tax Working Group, September


[125] Putland, Gavin R. (2008), "Why site-value rating is better, and how to implement it with no losers", Land Values Research Group, 1 September, www.lvrg.org.au


[133] Rodrik, Dani (1997), "Sense and Nonsense in the Globalization Debate", *Foreign Policy* 107, Summer


[143] Smith, Julie (1993), Taxing Popularity: The Story of Taxation in Australia, ANU Federalism Research Centre


Value Taxation: The Efficient and Equitable Source of Public Finance, Armonk NY: M.E. Sharpe


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Helm, Timothy Walter

Title:
Essays on the economics of price transmission

Date:
2012

Citation:

Persistent Link:
http://hdl.handle.net/11343/37539

File Description:
Essays on the economics of price transmission

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.