The visible hand of government: creating markets, tax competition and regulating retail financial products

by

Isaac Alfon

Submitted to the Department of Economics
in fulfilment of the requirements for the degree of
Doctor of Philosophy

at the

UNIVERSITY COLLEGE LONDON

January 2003

Abstract

Chapter one introduces the thesis and discusses the changing economic role of government. Chapter two assesses the efficiency of the supply function mechanism used in wholesale electricity markets. It presents a specific model of symmetric duopoly with two types of predictable and substitute demands. Equilibrium is characterised. Prices are not as high as implied in the literature, which assumes that demands are not predictable.

Chapter three analyses the effect of asymmetries between countries on environmental taxes where trade is liberalised, pollution is local, capital moves freely, governments and firms behave strategically and product markets are integrated after trade liberalisation. Following trade liberalisation, non-cooperative environmental policy must address two externalities: a competition externality and a trade externality. However, setting environmental policy cooperatively eliminates the trade externality. Entry is more likely when countries are different and would be associated, for example, with the exporting country. Chapter four relaxes the assumption of fixed wages. Labour is immobile, there are no imperfections in labour markets, and labour supply is linear. Wages remain different after trade liberalisation. They are higher, for example, in the country with a lower valuation of environmental damage.

Chapter five introduces an additional perspective for retail financial products and their regulation based on consumers' ignorance about their needs. This complements the standard perspective based on information asymmetries. Consumers have different needs and the same degree of ignorance about those needs. They take a view about their needs and search the market for a product that suits the (perceived) needs. Ex-post utility in an unregulated market is a decreasing function of consumers' ignorance. This approach is used to characterise product and advice regulation. It suggests that neither of these interventions will make all consumers better off than in an unregulated equilibrium.

Acknowledgements

I have been working full time while pursuing this thesis. As a result, the demands that I have put on all those mentioned here have been greater than in usual circumstances. It is very difficult to thank everyone for meeting these extra demands so in what follows I hardly make this distinction.

First I wish to thank David Ulph, my supervisor. I considered myself lucky to have David as my teacher and I am grateful for his patience and time – typically, at the end of a working day. David has been a continuous source of ideas, encouragement, and support. He has taught me a great deal about economic theory and how to value it, and I hope to build on this during my continuing professional career. Richard Blundell supervised me during my first year at UCL and I am grateful for his time and support.

I am grateful to Nir Vulkan, a colleague and a friend, for his encouragement throughout this thesis; to Paul Plummer for useful discussions and comments on Chapter 2 and Clive Briault and Marco Ottaviani for their comments on Chapter 5.

I am also grateful to National Economics Research Associates (NERA) and the Financial Services Authority (then, the Securities and Investments Board) for providing financial support. I am particularly grateful to certain people in these organisations, Peter Andrews, Paul Johnson and Penelope Rowlatt for their encouragement and support. Neither of them nor these organisations are responsible for the views expressed here.

Finally, a big thank you is also due to the "sleeping partners" in this enterprise. My parents, Yamin and Cotta, have always gone out of their way to support and understand my interests – including pursuing this thesis – and I am always very grateful for this. Eva, my wife, has been a constant source of support throughout these years and has always believed that I could finish this thesis – even when I had doubts. My brothers, Viviana, José, Luna and Yann, my uncles, Mery, "Tite", Pilar, Jacob, Mercedes, Roberto and Blanqui, and the Edery family have also been a valuable source of support. In the last years, Benjamin and Joseph, my sons, have provided the necessary smiles to get here.

Table of Contents

List of Figures and Tables

Chapter 1

Introduction

1.1 The role of the state

Adam Smith described a competitive market system using the metaphor of the invisible hand. By extension one could refer to the role played by the state as the visible hand – its institutions make it a virtue to affect market outcomes. This gives rise to questions such as "what business affairs should be undertaken by society itself acting through its government?" and "when government does not itself directly intervene, how far should it allow individuals and corporations to conduct their own affairs as they please?". These are indeed not new questions. They are posed as such in Marshall (1920). The neo-classical answer to these questions is based on the view that a market-based system, i.e. a decentralised price mechanism, will maximise social welfare. Thus, if the role of the state is to maximise society's welfare, the state should intervene to maintain a system of property rights to support a competitive market system and also when the price mechanism fails to maximise social welfare.

One of the developments in the last 10 to 15 years has been an increased awareness of the effects on markets of state intervention and of its limits. As a result, there have been developments such as privatisation, free movement of capital and the gradual withdrawal of the welfare state that are re-shaping the role of the state in the economy. One of the effects of these development is to increase the challenge to economics – and economists – to find "those effects of visible causes which are remote or lie before the surface", Marshall (1920). The resulting challenges include dealing with alternatives to price and quantity competition; taking into account firms' knowledge about demand; sequencing decisions by economic agents in multi-stage games; assessing the effect of asymmetries between countries; modelling the effect of consumers' limited information. These challenges are not necessarily new. However, the re-shaping of the role of the state makes meeting them also relevant to issues of public policy.

This thesis meets the above challenges in three specific public policy issues and in doing so contributes to the understanding of the role of the state. These three areas are: the incentive properties of wholesale market arrangements created by unbundling an integrated electricity utility (Chapter 2), environmental taxes when capital moves freely (Chapters 3 and 4) and the regulation of retail financial products (Chapter 5).

1.2 Creating markets

The state can create markets by defining new property rights,[1] addressing information asymmetries[2] or by unbundling a public-owned utility.[3] Here, I focus on the latter aspect in the particular case of a vertically integrated electricity utility. The market uses the high-voltage transmission network to serve as the nexus between the upstream activities (generation) and the downstream activities (supply). The motivating example is England and Wales but similar markets exist in Norway, Sweden, Spain, New Zealand, Australia, and various Latin American countries.

Unbundling and creating a market to serve as a nexus between generation and supply of electricity activities gives rise to technical issues – whether the lights will go out – and to economic issues – whether appropriate incentives for efficiency are provided. The importance of getting the technical issues right should not be underestimated but will not be pursued here. The range of economic issues raised by formal unbundling is vast. Broadly speaking, the issues are whether there are more efficient alternatives to vertical unbundling and whether unbundling takes place in such a way as to introduce incentives for efficiency.

Here I will not consider alternative market structures. I just note that many of the regulatory issues that arose in the case of vertical utilities that were privatised as such are about the operation of vertical relationships.[4] Introducing incentives for efficiency is a

---

[1] For example, tradable pollution permits, Markandya (1991).

[2] Akerlof (1970).

[3] Creating markets is not necessarily restricted to the state. Firms create new markets as a result of product innovation and by addressing the information asymmetries that exist between buyer and seller and prevent a market from operating.

[4] The most noticeable example of this kind is the UK's Monopolies and Mergers Commission decision to

vast undertaking. It includes assessing the extent to which competition can play a role in each of the underlying activities[5] and then designing an appropriate incentive mechanism for each one.

In Chapter 2, I explore the adequacy of one of these incentive mechanisms: the bidding system that exists at the core of the wholesale market for electricity. For this purpose, I use a model of supply function equilibrium. An electricity generation firm will have many power plants, each one with various sets that can be operated at different levels – and costs. So it is difficult to describe the firm as offering a price (Bertrand competition) or a quantity (Cournot competition) but it is appropriate to describe the firm as offering a supply function.

In this market, firms bid every day a supply function. Existing modelling of supply function equilibrium assumes that demand is a continuum of states and that future demand is unknown. Newbery (1992) suggests that duopolists would have the ability to maintain prices well above marginal costs. It is nevertheless recognised that firms could constrain their ability to raise prices in the spot market, for example, by selling in the contract market, Green (1999).

I explore an alternative way of modelling a supply function equilibrium by taking into account the demand side in a way that is more appropriate to wholesale electricity markets. So I assume that there are two types of daily demand, high and low, demands are substitutes, and firms can predict daily demands. I present a specific model of a symmetric duopoly where demand has these characteristics. Firms choose between quadratic supply functions and are subject to capacity constraints. An equilibrium is well defined, unique and has the correct comparative statics properties. The equilibrium prices and welfare losses predicted by the model presented in this paper are substantially lower than those suggested by Newbery (1992). The results are in line with the

require the vertical separation of British Gas into two companies one undertaking supply to consumers and another one transport and storage of gas, MMC (1993).

[5] Typically, transmission and low-voltage distribution is a natural monopoly requiring some form of regulation, supply is an activity where competition can be introduced gradually and regulation may well be a temporary feature, and generation is an activity where competition can play a predominant role from the outset.

empirical evidence from Wolfram (1999), which suggests that prices are not as high as suggested by existing models.

This is a specific model that assumes that demand in the next day is known. This is in contrast with the existing literature, which assumes that demand is unknown. The model could be extended in various directions. For example, the model could be used to explore equilibrium results assuming that firms have some knowledge about demand, i.e. not as little as in the existing literature and not as much as assumed here. There is nevertheless a policy implication. The analysis brings into question the view that making a wholesale market more competitive should focus on changing the principle of a bidding system that prices electricity according to the price of the most expensive plant in use.

1.3 Environmental taxation and trade

One of the traditional roles of the state has been to correct externalities such as those created by productive activities that generate local pollution, perhaps through a tax. In a closed economy, the level of an environmental tax would depend mainly on market structure and on the damage to the environment. The reduction of barriers to trade and to the free movement of capital have affected the process of setting environmental taxes in various ways. First, they increase the effectiveness of one of firms' possible responses – relocation to another country. A firm that is located in one country, Metropolis, and sells its products in two countries, Metropolis and Capitolina, might decide to relocate to Capitolina if pollution taxes in Metropolis increase significantly. Second, they introduce the possibility of governments behaving strategically to attract firms into their country and, therefore, using environmental taxation as a substitute for trade policy. These result in a possible 'race to the bottom' in tax setting between governments so that an equilibrium will be characterised by taxes that are below the level of the first-best tax (marginal damage to the environment).

In these circumstances, it is appropriate to model governments' tax setting behaviour as a multi-stage game involving firms and governments. There is an extensive body of literature on governments' rent seeking behaviour in the context of environmental policy, see Sturm (2002) and Ulph (1997). There are, however, various aspects relevant to this possible 'race to the bottom' that – to the best of our knowledge – are not simultaneously addressed in the literature. First, product markets are integrated after the abolition of trade barriers. (This means that firms perceive both countries as part of the same market and do not make separate decisions for each country.) Second, if firms can relocate then market structure should be endogenous. Third, decisions must be sequenced in a way that reflect the agent's commitment and it is unclear that the state can make a credible commitment to a level of environmental tax. Fourth, reality is characterised by differences between countries. And, finally, the effect on wages needs to be considered.

Ulph (1995) addresses the first three of these aspects. In particular, Ulph assumes that governments cannot make a credible commitment to a level of environmental tax before firms decide to enter the market but can commit to a level of environmental tax before firms make their output decisions. The main result is that, in a symmetric equilibrium, after the abolition of trade barriers, environmental tax will be above the first-best tax or marginal damage to the local environment – so no evidence of race to the bottom – and that new entry into this integrated product market is unlikely.

Chapter 3 addresses the fourth issue by relaxing the standard assumption that countries are identical. It also explores alternative assumptions about the extent to which governments take into account the effect of their environmental policies on the neighbouring country – non-cooperative and cooperative tax setting behaviour.

The analysis shows that, when governments set the environmental tax in a non-cooperative form, there are two externalities that create a wedge between the environmental tax and marginal damage. One of the sources of externalities is related to the level of output, which, other things being equal, depends on the intensity of competition. I refer to this as the competition externality. The other externality is related to the trade surplus (so a trade deficit reduces the externality). I refer to this as the trade externality. Consequently, the environmental tax would be above the first best tax in the exporting country and most probably in the importing country as well. Furthermore, environmental policy in the exporting country will be tougher than in the importing country. I also find that setting taxes cooperatively eliminates the trade externality so that the environmental tax only needs to cover the competition externality.

I then use some simple simulations to show the effect of abolishing trade barriers and of the resulting environmental policy on entry. I find that in a non-cooperative tax setting, entry is more likely in an asymmetric equilibrium. Entry would be associated with the exporting country or with the small country. Cooperation in tax setting results in lower taxes in the exporting country and increases further the likelihood of entry.

The main aim of Chapter 4 is to test the robustness of the conclusion in Chapter 3 by relaxing the assumption that wages are fixed. I then use the framework to explore the effects of trade liberalisation on wages when there are environmental taxes and capital is mobile. For these purposes, I extend the model in Chapter 3 in a very simple way. I assume that labour is immobile, that firms are wage takers and that there are no imperfections in the labour market. The latter means that wages will adjust with no impediments and that there is no involuntary unemployment. (Otherwise, it may be as if wages remained fixed as discussed in Chapter 3.) I extend the model in the simplest possible way and I assume that demand for the product and the labour supply are separable and that the labour supply is linear.

Based on these assumptions, the analysis suggests that the effect of abolishing trade barriers is to reduce the equilibrium level of environmental taxes. The analysis also suggests that wages will remain different after integration because of government intervention (the environmental tax). Another effect of abolishing trade barriers is to increase equilibrium wages. The simulations suggest that wages will be higher in the country with a lower valuation of environmental damage or the country with a smaller population.

These results depend on the assumptions made. One feature of an alternative utility function, Stone-Geary, is that by a suitable transformation, it results in a labour supply function that becomes vertical when the number of hours supplied is close to a pre-determined maximum number of hours available for work. I explore how the equilibrium will be characterised in these circumstances assuming that there is a fixed supply of labour. In this case, environmental taxes will be lower after the abolition of trade barriers, but the output remains the same and there is no new entry.

1.4 Retail financial regulation

One of the effects of the gradual withdrawal of the welfare state is the increasing importance of the market for retail financial products for social welfare. However, government intervention in retail financial services regulation has traditionally been analysed with almost no regard to economic considerations.[6] Consequently, policy decisions may have been unduly influenced by considerations about the status quo rather than by any formal analysis of the underlying market failures and the alternatives to address them. So the UK developed a seemingly unique system of advice regulation[7] whereas other European countries have relied more on product regulation.

This is gradually changing. In the UK, the current government has introduced elements of product regulation with a view to encourage a wider proportion of the population to make their own financial provision. The introduction of voluntary elements of product regulation for savings accounts, unit trusts, insurance, mortgages (CAT standards – for charges, access and terms) and for personal pensions (stakeholder pensions) and the recommendations of Sandler's review of savings, Sandler (2002). In other European countries, pressure for change is building up as a result of the financial pressure from continuing reliance on public retirement systems.[8] Finally, the lack of a co-ordinated approach for advice regulation is regarded as an obstacle for a single market for retail

---

[6] For example, the review that lead to the enactment of the 1986 UK's Financial Services Act states: "In assessing the optimum degree of regulation I have not attempted any sort of cost-benefit analysis, partly because I am not competent to undertake it and partly because I am sceptical about its practicability. [...] It may be that the most efficient market is that which is wholly free from regulation but it is unlikely that such market would afford protection to investors which anyone today would regard as adequate. One has to make a value judgement on the relative weight to be attached to market freedom and to investor protection. My judgement, as I have said, is that regulation in the interest of the latter should be no greater than is necessary to protect reasonable people from being made fools of", Gower (1985)

[7] FSA (2000) provides a comparison of the cost of regulatory institutions in various jurisdictions (Australia, Canada, Ireland, France, Germany, Hong Kong, Singapore, Sweden, UK and USA). It shows that the UK's Financial Services Authority spent during 1999/2000 £87 million in the regulation and supervision of financial advice, advisors and marketing of retail financial products. This is in contrast with all the other jurisdiction that did not spent a comparable part of their resources on this category – even if one allows for the possibility that in certain cases these costs are aggregated with the cost of regulating and monitoring credit institutions.

[8] This will become very significant as population ages in the first half of this century. Figures for the European Commission suggest that on current levels of benefits, expenditures on public pensions as a percentage of GDP is expected to increase by more than 40% between 2000 and 2050 in seven European countries and in three of them by more than 70%, Merrill Lynch (2001).

financial services and European regulators are starting to coordinate their policies, see, for example, FESCO (2001).

These developments are bringing the underlying market failures into sharper focus. Broadly speaking, they arise from consumers' inability to monitor long-term relationship and their limited information. Consequently, consumers face two distinctive types of risks. First, consumers' inability to monitor long-term relationships suggests that consumers are exposed to the insolvency of the product provider. Second, consumers' limited information has two aspects: asymmetric information about products and limited information about their needs. They both result in the purchase of products that do not suit consumers' needs and distort firms' pricing decisions. The first type of risk is dealt with, in part, by setting capital requirements and is not addressed here. Chapter 5 explores an aspect of consumers' limited information.

The standard economic approach to analyse consumers' limited information is based on models where consumers understand their needs and there is asymmetric information about products. Chapter 5 introduces an additional perspective, which is based on consumers' ignorance about their needs. This does not mean that the search process typically associated with asymmetric information becomes irrelevant. It means that there is a further step before product search where consumers' needs are established.

This approach is consistent with the regulators' understanding of financial services. Gravelle (1994) also adopts a similar approach to explore a specific aspect of the market for advice – the incentive properties of various forms of remunerating financial advisors. The recent literature on the economics of retail financial services regulation such as Llewellyn (1999), Goodhart et al. (1998), Spencer (2000) suggests that there is more than just information asymmetries between buyers and sellers in retail financial services. They suggest that there is also an issue of credence goods. The latter addresses cases where consumers know that they have a problem, for example, an appliance that is not functioning, but do not really know the type of problem. Thus, an expert is required to make the right diagnosis of the problem. The literature focuses on inefficient treatment and overcharging. The model developed in Chapter 5 differs from the credence goods literature in various ways. The main difference is that I put most of the emphasis in defining the consumer's problem and that, as a starting point, I have assumed marginal cost pricing and effectively assumed away firms' strategic behaviour.

In Chapter 5, I have developed a simple approach to understand the effect of consumers' ignorance about their needs with the following features. First, consumers' needs are eventually revealed, though it is then too late to do anything about it. Second, consumers have different needs so that different types of consumers make different types of error. Third, consumers are not completely ignorant about their needs. The underlying degree of ignorance is assumed to be the same across the population. Consumers' information about their needs is correct in the sense that in terms of a range of beliefs about their needs, this range includes the true needs. Finally, I assume that consumers incur a direct utility loss if they make insufficient provision. If consumers provide in excess, the cost is the financial cost of paying for something that is not really needed.

In an unregulated market, consumers develop a view about their needs, search the market (at a cost) for that product and buy it. I characterise consumers' (average) ex-post utility in an unregulated market, which is a decreasing function of consumers' ignorance about their needs. I then use this approach to explore the effects of two polar approaches to address the welfare losses that arise from consumers' ignorance about their needs: product regulation and advice regulation. Broadly speaking, product regulation means controlling the supply whereas advice regulation means regulating through a mechanism that affects the demand for the product.

As a starting point, I adopt a very simple approach to product regulation and I assume that the regulator has the powers to specify the details of the regulated product sold to all consumers. The regulator observes the distribution of consumers' true types and chooses a product that maximises total welfare. I assume that in this case, consumers do not incur a search cost. I then characterise a consumer's ex-post utility under advice regulation. An adviser does not observe a consumer's needs. He can learn what the consumer knows and assess more accurately the consumer's needs. In this case, the search cost is the cost of advice. Now the regulator's role is to set the quality of advice that maximises total welfare.

The main result from comparing the resulting levels of ex-post utility in these three cases is that different types of consumers will be better off under different regulatory regimes. This result is broadly consistent with changes in the UK, which are effectively leading to a regulatory regime that combines product regulation and advice regulation.

This work is far from being a complete model of retail financial services and the model or the approach developed here could be extended in various ways. For example, a model of advice that combines the assessment of consumers' needs and search could be built. The main result of this analysis also suggests that another possible extension of this model will be exploring the effect of the decisions to combine product regulation and advice regulation.

References

Akerlof, G. (1970), "The market for 'lemons': quality uncertainty and the market mechanism", *Quarterly Journal of Economics*, vol. 84, pp. 488 – 500

Financial Services Authority (2000), "*Annual report*", London

Forum of European Securities Commission (2001), "*Standards and rules for harmonising core conduct of business rules for investment products*", Consultative Paper, February (available from http://www.europefesco.org)

Gower, L. C. B. (1985), "*Review of investor protection*", HMSO, London

Goodhart, C. A. E., P. Hartmmann, D. Llewellyn, L. Rojas-Suarez, L. and S. Weisbrod (1998), *Financial regulation. Why, how and where now?*, Routledge, London

Gravelle, H. (1994), "Remunerating information providers: commission versus fees in life assurance", *Journal of Risk and Insurance*, vol. 61, pp. 425 – 457

Green, R. (1999), "The electricity contract market in England and Wales", *Journal of Industrial Economics*, vol. 42, pp. 107 – 124

Llewellyn, D. (1999), "*The economic rationale for financial regulation*", FSA Occasional Paper 1, April

Marshall, A. (1920), *Principles of economics*, eighth edition, MacMillan, London

Markandya, A. (1991), "Global warming: the economics of tradeable permits" in D. Pearce (ed.), *Blueprint 2: greening the world economy*, Earthscan Publications, London

Merrill Lynch (2001), "*Pension barometer for Europe*", January

Monopolies and Mergers Commission (1993), "*British Gas*", HMSO, London

Newbery, D. M. (1992), "*Capacity-constrained supply function equilibria: competition and entry in the electricity spot market*", Department of Applied Economics, University of Cambridge, Mimeo, June

Sandler, R. (2002), "*Medium and long-term retail savings in the UK*", The Stationery Office, London (available from http://www.hm-treasury.gov.uk/)

Spencer, P. D. (2000), *The structure and regulation of financial markets*, Oxford University Press, Oxford

Sturm, D. (2002), "*Trade and the environment: a survey of the literature*", Mimeo, University of Munich (available from http://www.vwl.uni-muenchen.de/ls_marin/sturm/)

Ulph, A (1997), "Environmental policy and international trade", in C. Carraro and D. Siniscalco (eds.), *New directions in the economic theory of the environment*, Cambridge University Press, Cambridge

Ulph, D. (1995), "*Environmental taxation and entry: a symmetric equilibrium*", University College London, Mimeo

Wolfram, C. (1999), "Measuring duopoly power in the British electricity spot market", *American Economic Review*, vol. 89, pp. 805 – 826

Chapter 2

Creating a wholesale electricity market

2.1 Introduction

The British government restructured the electricity industry in England and Wales prior to its privatisation from 1990. The industry was restructured vertically, e.g. generation was separated from the high-voltage transmission network, and horizontally, e.g. three electricity generation companies were formed.

One of the features of the restructuring was the introduction of competition in the generation of electricity – mainly between two fossil fuels generating companies.[1] They compete in a centralised spot market by offering (to generate) electricity from their various generation sets at different prices. So it is as if they offered a supply function. This form of competition generalises the standard price and quantity competition in the literature. Thus one could say that Bertrand's model of price competition implies a horizontal supply function and that Cournot's model of quantity competition implies a vertical supply function. Similar markets exist in Norway, Sweden, Spain, New Zealand, Australia and various Latin American countries.

Consumers in this market tend to be large users and firms that sell electricity to retail consumers so this market is referred to as the wholesale electricity market. The intention of these arrangements was to produce a price that approximates those set to maximise social welfare – see below. However, there have been concerns that the likely strategic behaviour of generators was not fully taken into account and that the price of electricity in the wholesale market was therefore unnecessarily high. Thus, there were a number of inquiries initiated by the UK electricity regulator, see for example, Offer (1991) and Offer (1992). The cumulative effect of these reviews was the decision in 1994 by the electricity regulator to impose a two-year cap on the average prices in this market.

[1] The third generating company owned the nuclear plants and because of their cost structure, it is unlikely to compete on prices with the two others companies.

Competition with supply functions has also been the subject of academic research. Klemperer and Meyer (1989) – hereafter KM – developed the concept of equilibrium in supply functions in a seminal article. Their approach has been applied to the behaviour of generating companies, for example, Bolle (1992), Newbery and Green (1992), Newbery (1992), and Green (1999a).[2] One of the main conclusions of these papers is that duopolists have the ability to drive prices above marginal costs. However, empirical analysis by Wolfram (1999) shows that prices "while higher than marginal costs are not nearly as high as most theoretical models predict". In her analysis, Wolfram suggests three possible explanations – regulatory constraints, the threat of entry and the contract market. The purpose of this chapter is to explore an alternative explanation: whether duopolists competing in supply functions do indeed have the ability to drive prices above marginal costs as suggested in the literature.

The approach adopted here is as follows. Previous literature applies KM framework, which assumes that demand is a continuum of states and could take any value between two known values. However, electricity demand is characterised by a different form of variability of demand. It varies during the day in a predictable fashion and is characterised by the cross elasticity of the demand between different periods of the day. The latter results from consumers' ability to re-arrange their consumption pattern to minimise the total cost of electricity. This model of supply function equilibrium incorporates these effects.

Thus, the motivation of this chapter is the observation of the new arrangements that characterise the wholesale electricity market, the creation of two dominant players and the concerns about the price efficiency of these arrangements. The contribution made by the chapter is, however, theoretical – a model of supply function competition that incorporates information characteristics and the effect of demand interdependency.

---

[2] An alternative approach to analyse the behaviour in this market is that of Von-der-Fehr and Harbord (1992). They emphasise that bids submitted by electricity companies constitute a step function rather than a smooth and differentiable function, as assumed by the supply function approach. However, the algorithm that calculates the pool prices automatically converts the price bids into a monotonic schedule, Green (1992). It is therefore appropriate to address the determination of the equilibrium based on smooth and differentiable functions.

Section 2.2 describes briefly the electricity market in England and Wales and what makes a supply function relevant to study the decision making process of a firm. Section 2.3 surveys the literature on supply function equilibrium. Section 2.4 presents an alternative modelling approach to supply function competition. Section 2.5 discusses the equilibrium results. Section 2.6 qualifies the result from the model. Finally, section 2.7 concludes and derives policy implications.

2.2 A framework for supply functions

This section presents a framework for supply functions. First, I describe the arrangements of the electricity industry in England and Wales since 1990.[3] Next, I consider the conditions in which a supply function may be useful to analyse firms' decisions.

2.2.1 Market structure: the electricity market in England and Wales

The new structure of the electricity market was established on April 1990. Before that date the industry was, in practice, a vertically integrated utility managed by the Central Electricity Generation Board (CEGB), see Figure 2.1. The structure was broken down vertically and horizontally. Generation, transmission, distribution and supply activities were formally separated, see Figure 2.2.

The generation assets of the CEGB were vested into three companies: Nuclear Electric took over the nuclear power stations and National Power and PowerGen took over the non-nuclear stations. The high-voltage transmission system was vested into the National Grid Company. This also acts as a system manager and is in charge of ensuring the orderly despatch of plants. The distribution and supply assets[4] were vested into 12 regional electricity companies that fitted in the existing electricity area boards. The relationship

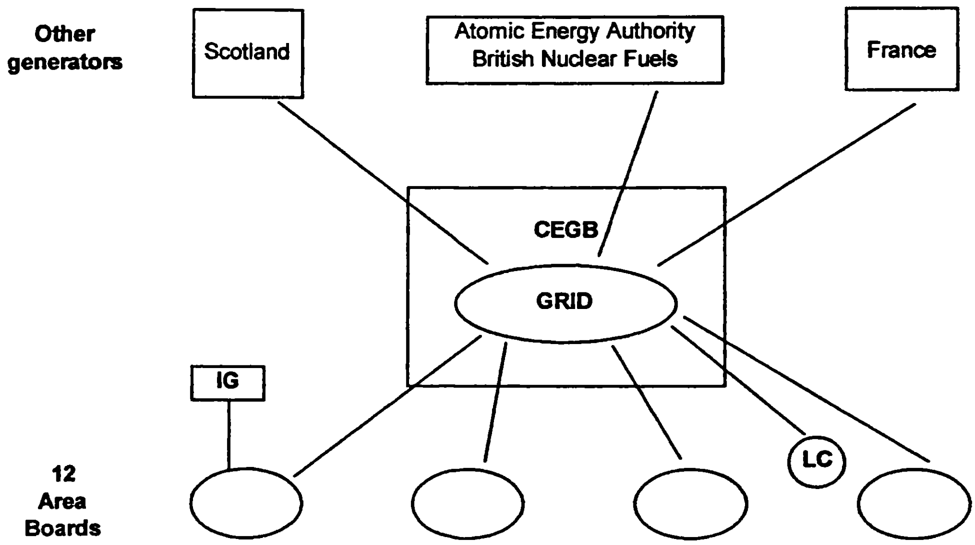
---

[3] Much has been written about the restructuring of the electricity industry in England and Wales prior to its privatisation. Green (1991) provides a detail description of the arrangements in the wholesale markets, Vickers and Yarrow (1991) review the changes introduced to transmission and distribution and Posner (1993) looks at the implications for primary fuels and provides a very useful summary of the changes. The arrangements introduced in Scotland and Northern Ireland are different. Scotland maintained a vertically integrated industry. The arrangements in Northern Ireland included the separation of generation and the introduction of a regime conducive to supply competition but without introducing a wholesale market.

[4] Distribution is the delivery of electricity over medium and low voltage networks to final consumers. Supply consists of purchasing electricity in the wholesale market and selling it to consumers.

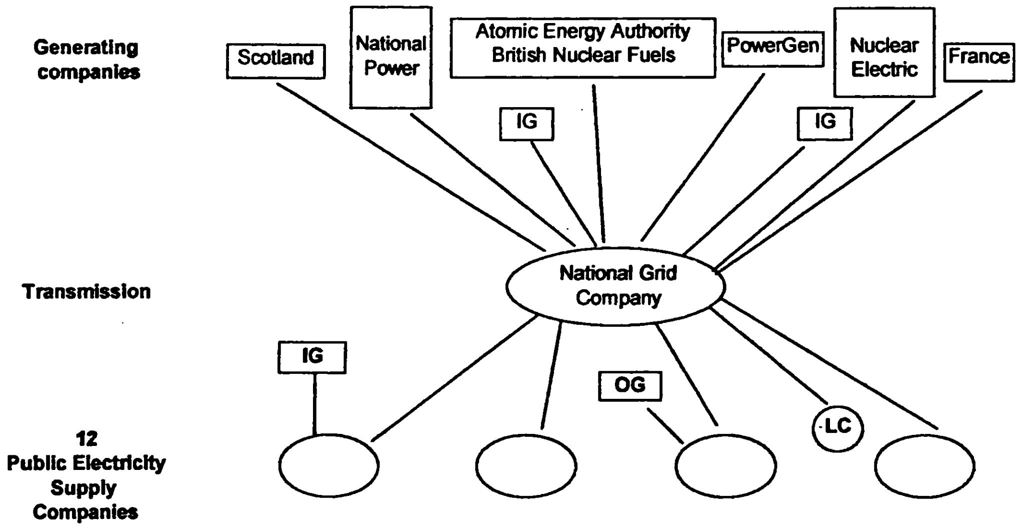
between these new entities is governed by contracts, by market rules, or by a combination of both. In late 1990, the regional electricity companies were privatised and in early 1991 the two fossil fuel generation companies were privatised.

Figure 2.1: The industry before 31 March 1990



Notes: CEGB - Central Electricity Generation Board
IG - independent generator
LC - large customer

Figure 2.2: The industry after vesting on 1 April 1990



Notes: National Power, PowerGen and Nuclear Electric - Successor Companies to CEGB
IG - independent generator
LC - large customer
OG - own generation

An independent electricity regulator was appointed to oversee the industry (then, the Office of Electricity Regulation, Offer). Transmission and distribution were regarded as natural monopolies and were subject to a regime of price-cap regulation. Supply was regarded as a temporary monopoly and was subject to a regime of price-cap regulation that has been phased out gradually. Given the number of power stations and the size of the system, competition in generation was regarded as feasible. An unregulated wholesale electricity market was established.[5] There are two elements to this market: a centralised spot market (also denoted the pool market) and a contract market. The next paragraphs provide a very brief description of the pool market and highlight the similarities of the arrangements in the spot market with the supply function framework.

Every day each power generation company decides what capacity is available and submits a bid to the system manager indicating its willingness to supply certain levels of output at specific prices from each power station during the next 24 hours. The system manager orders the bids in ascending order, and combines them with a demand forecast. An equilibrium price or system marginal price (SMP) for each half-hour of the following day is then determined. There is also a payment to all capacity available for generation whether used or not. This is a function of the balance between demand (as forecast) and capacity available, and of the value attached to unserved demand. In 1990, the regulator set the latter at £2 per kWh with an inflation adjustment. So when the equilibrium price is high because capacity is tight there is also a separate payment for capacity.

In addition, the system manager must buy additional power to offset deviations during the day from the pre-determined schedule and because of transmission system constraints. These costs are passed through to consumers in the pool market ("uplift"). The relationship between these magnitudes is summarised in Table 2.1.

Green (1991) shows that the approach in the pool input price is a first approximation to the prices set to maximise welfare when demand is subject to stochastic shocks and various technologies are used to meet demand. It is, however, an open question whether (and to what extent) this approach will indeed result in prices that maximise welfare when

[5] This changed with the introduction by the electricity regulator of a two-year cap on the average pool price until February 1996.

duopolists are bidding to generate electricity – perhaps a separate capacity payment is not required because the system marginal price increases sufficiently when capacity is short.

Table 2.1: Price relationships in the pool market

Pool input price	System marginal price + capacity payment	Price paid for unit of electricity scheduled to be generated. Announced with despatch orders of the day
Pool output price	Pool input price + uplift	Price paid by participants in the wholesale electricity market

A contract market overlies the spot market. The difference between the contract market and the pool or spot market is that trade is carried out through bilateral contracts, typically, contracts for differences. These contracts are financial instruments and involve no physical transactions. They are normally designed as option contracts against the pool input price and aim to hedge the parties in spot market transactions against price variability. They can be designed in many possible ways. A portfolio of contracts for differences may include call options, which set a ceiling to the spot price, and call-put options, which have the effect of setting a fixed price for spot price transactions.

The industry was privatised with a substantial number of contracts in place. The effect of contracts for differences is an aspect of the wholesale electricity market that has been widely researched, for example, Powell (1992), Lucas and Taylor (1993) and Green (1999a). These contracts pre-commit power generation companies to supply a certain amount of electricity. So, if a large proportion of a firm's capacity is sold in contracts, the bidding strategy in the pool market is affected and there will be a downward pressure on prices in the spot market. Green (1999a) models the effect of contracts on bidding strategies and shows that forward sales constraint generators' ability to price above marginal costs. Helm and Powell (1992) provide an empirical analysis of the relation between the level of contract cover and prices in the spot market. Their analysis identified a change in the long-term relationship between price and quantities in the market following the termination of the first tranche of contracts.

Another important part of the arrangements introduced, which is sometimes overlooked as bearing no direct relationship with the wholesale electricity market, is the gradual opening up of supply activities to competition. The reasoning underlying this view might be related to the fact that supply costs represent a small share of end-users price, about 15%, whereas generation costs represent about 60%. However, competition in supply allows an increasing number of electricity consumers to select the least cost supplier.[6] The main form of competition in supply is to depart from average prices and to offer prices that are closely related to pool prices. This makes the demand for electricity more price sensitive and enhances the substitution effects during different periods of the day.

Summing up, the bid process used to determine the system marginal price resembles the supply function framework described in the literature (see next sub-section). The same bid, however, is also used to determine other elements affecting revenue such as capacity payments and uplift. This increases the role that strategic considerations might have in the determination of a firm's bidding strategy.

2.2.2 A supply function framework

I turn now to consider the conditions in which a supply function framework may be useful to model firms' pricing decisions. A supply function is some functional relationship between price and quantity that reflects the behavioural assumptions of the firm. In the case of a duopoly, each firm selects a supply function so that it maximises its profits taking into account the response of its rival. Thus, in a supply function equilibrium the dimension of choice of the firm is not one (price or quantity) but infinite.

The key concept underlying a supply function is commitment. This could be either a commitment to a course of action prior to the resolution of some uncertainty or a commitment for a number of periods of time. KM suggested that in the presence of some form of uncertainty, say about demand, firms may prefer to commit themselves to a supply function that is independent of the uncertainty rather than to a specific price or quantity. Turnbull (1983) and Robson (1981) also investigated the existence of equilibrium in linear supply functions as the means of verifying the consistency of conjectures in equilibrium. However, the innovation in KM's approach is to regard the selection of the form of the supply function as part of the profit maximising process.

[6] Initially, only consumers with an average maximum demand of at least 1 MW were allowed to select a supplier other than its regional electricity company. The threshold was reduced in April 1994 to 100 kW and it was abolished by 1998.

Supply functions are also useful to analyse firms' pricing decision when they have to set in advance production plans for a number of periods of time. For example, firms may have to specify in advance a production plan for a number of predictable realisations of demand during a day. This case is similar to the one addressed in KM's paper since firms could also identify the optimal price-quantity choice for each realisation of the demand. A supply function gives firms a mechanism to incorporate all these points. This interpretation underlies the application of the supply function framework to the wholesale electricity market by Newbery and Green (1992), Bolle (1992) and Newbery (1992). However, power generation companies are not bidding against a random demand as assumed in KM. They are bidding against a demand forecast that they can also predict. Thus, I could regard demand as known for the purpose of bid determination.

A Nash equilibrium in supply functions requires that firms have no ex-post regrets about the supply function offered before the uncertainty was cleared. The equilibrium is determined by assuming that a firm could identify its optimal choice if it could assume away the uncertainty and then determine its optimal choice, taking into account the optimal choice of its rival. A supply function is then selected so that it contains all the optimal price-quantity combinations.

The alternative to setting a quantity or a price with reference to the expected realisation of the demand will not be a Nash equilibrium – optimal ex-post, once the uncertainty is cleared. In this case, firms will have to adjust either the price (if a quantity was announced) or the quantity (if a price was announced). A supply function is superior insofar as it provides an optimal adjustment to the uncertainty. With no uncertainty, however, there is only one profit maximising price-quantity combination. Any supply function containing that combination will be an equilibrium supply function so there will be a multiplicity of equilibria.

An auctioneer (or system manager) was not an explicit element of KM supply function framework. This is so because in their view the supply function will be implicit in the organisational arrangements of the firm and because firms do not quote a fixed price regardless of demand conditions. However, when a time dimension is involved an auctioneer may be necessary to compute the equilibrium prices and to ensure that once prices are announced trade takes place at those prices. An important element of the institutional arrangements is then the set of rules that the auctioneer uses to determine the equilibrium price. As suggested by Bolle (1992), those rules could be such that the auctioneer behaves like the "Walrasian" auctioneer or such that he behaves strategically. In the first case, the auctioneer will accept any supply function. In the second case, he could operate according to some pre-determined rules. In the case of the wholesale electricity markets, the auctioneer has two main roles. First, he prepares the demand forecast based on a known methodology and on historic data (adjusted on the light of weather forecast and other specific factors).[7] Second, the auctioneer orders the bid in ascending order and sets the system marginal price as the price of the most expensive unit in the system. This suggests that a further aim of building a model of supply function competition is to verify the consistency of these rules with the equilibrium result, which assumes implicitly the existence of an auctioneer.

2.3 Models of supply function equilibrium

This section surveys various models of supply function equilibrium in order to set the scene for the model developed in the next section. I start reviewing KM's model and what is common among these models.

In these studies, demand is a function of price and is subject to an exogenous random shock, θ, that shifts it horizontally, $Q = D(p, \theta)$. Firms know that the random shock that affects the demand is a continuum of states and distributed on $[\theta_0, \theta_1]$, and that $D_{p\theta} = 0$. It is also assumed that $D_p < 0$ and that $D_\theta > 0$. In addition, firms have identical cost functions with strictly positive marginal costs.

Before the uncertainty is realised, each firm announces its supply function, $S(p)$. Once the uncertainty is cleared, an equilibrium price can be determined and production takes place. A Nash equilibrium in supply functions is a pair of supply functions such that for each firm the supply function selected is the best response given the supply function selected by its rival.

[7] The demand outturn may still differ from the auctioneer's forecast. If so, there may be additional payments to power generation companies, part of the "uplift".

The supply function is the unknown so profit maximisation leads to a differential equation. This introduces two separate but related issues. The first issue is identifying the solution – solving the differential equation. The solution will always include an arbitrary constant of integration, which implies that solving the differential equation only shows that there may exist a multiplicity of equilibria. The second issue is therefore identifying an equilibrium – a value for the constant of integration. Ideally, the equilibrium selection mechanism will lead to a unique value for the constant of integration so that a unique equilibrium in supply functions can be shown to exist. In what follows I survey models of supply function equilibrium from KM, Bolle (1992) and Newbery (1992) and show how these issues have been dealt with.

We start by outlining what is common to all these models. Given the supply function of firm j, firm i serves the residual demand. Thus, firm i maximises the following (unrestricted) profit function with respect to prices

$$\Pi_i = p\big(D(p,\theta)-S_j(p)\big)-C\big(D(p,\theta)-S_j(p)\big) \tag{1}$$

The first-order condition is

$$D(p,\theta)-S_j(p)+[p-C'(D(p,\theta)-S_j(p))]\cdot(D_p(p)-S'_j(p))=0 \tag{2}$$

The optimal supply function is derived by solving the first-order condition for $S'_j(p)$. For a symmetric duopoly, the equilibrium supply function is a function, $S(p)$, that solves the following differential equation

$$S'(p)=\frac{S(p)}{p-C'\big(S(p)\big)}+D_p(p) \tag{3}$$

Note that the equilibrium supply function will be independent of the random shock and that will be valid for any level of demand.

Local second-order conditions for profit maximisation are satisfied if the slope of the supply function is positive in the relevant range (p. 1254 in KM). With unbounded support of the uncertainty, that is for θ on $[0,\infty]$, global second-order conditions and the requirement of a unique equilibrium price are satisfied for any realisation of the uncertainty. Otherwise, I need to verify the existence of a unique equilibrium.

Consider now the first issue – solving the differential equation. An analytical solution to the differential equation in 3 above cannot be easily identified because it is not necessarily an exact differential equation.[8] KM adopt a more general approach to solve this differential equation by transforming it into a system of two autonomous differential equations. So equation 3 is re-arranged as follows

$$\frac{dS}{dt} = S(p) + D_p(p)(p - C'(S(p))) \tag{4}$$

$$\frac{dp}{dt} = p - C'(S(p)) \tag{5}$$

Note that t is not necessarily time, it can be any arbitrary operator. The system is solved by using the eigenvalues method and making the following assumptions about marginal costs and demand

$$C'(S(p)) = cq \tag{6}$$

$$D(p;\theta) = \theta - \varepsilon\, p \tag{7}$$

where c and ε are positive numbers. In these circumstances, a solution is well specified. There are two positive and unequal eigenvalues, λ_1 and λ_2 such that $\lambda_1>1$ and $0<\lambda_2<1$. The general solution of this system of equations is

$$\begin{pmatrix} S \\ p \end{pmatrix} = A_1 e^{\lambda_1 t}\begin{pmatrix} v_1 \\ w_1 \end{pmatrix} + A_2 e^{\lambda_2 t}\begin{pmatrix} v_2 \\ w_2 \end{pmatrix} \tag{8}$$

where A_1 and A_2 are arbitrary constants of integration and v_i and w_i are the associated eigenvectors.

Consider now the second issue – the equilibrium selection mechanism. KM characterise a supply function equilibrium by identifying the limit cases of $S'(p)$ in equation 3. Since the condition for a supply function to be a candidate for an equilibrium is a positive slope, it follows that the two limit cases for $S'(p)$ are zero and infinite.

Setting $S'_j(p)$ to zero, suggests that firm i assumes that its actions create no response from firm j. This is similar to the assumption made by a duopolist that uses quantities as its strategic variable and takes as given the output of its rival, as in Cournot's model. Thus, I refer to the supply function derived by setting $S'_j(p)$ to zero as the Cournot supply function. In fact, the equilibrium price and quantity obtained by tendering this supply

[8] This is a differential equation that is the total differential of a function in the same variables.

function are the same that will be obtained for a given realisation of the uncertainty by setting quantities under the standard Cournot approach. The second limit case is a slope of infinite. Equation 3 suggests that, in this case, the supply function will then be the marginal cost schedule. These two supply functions set the limits for the equilibrium supply function (a solution to equation 3); the Cournot supply function from above and the competitive supply function from below. Furthermore, the two limit cases trace through the origin and so will the equilibrium supply function (claims 1 and 2 in KM).

The constants of integration in equation 8 are therefore selected so that the resulting supply function is consistent with the features of the equilibrium supply function identified above. Namely, that quantity, S, and price, p, tend to zero when $t \to -\infty$ and that S / p has a positive limit when $t \to \infty$. Given λ_1 and λ_2, this suggests that A_1 must be zero. The second constant of integration, A_2, is different from zero but a value for it is not necessary since it disappears when one substitutes to obtain S, quantity, as a function of p. The resulting equilibrium supply function is linear with a positive slope and with no intercept.

Bolle (1992) studies the effect that alternative assumptions about demand will have on the equilibrium solution. He considers three alternative variations under the assumption that marginal cost is zero.

In the first two variations, Bolle assumes that end-users' demand is independent of the price paid in the wholesale market. That is end-users pay a price that may be some average of the prices prevailing at different times of the day. Hence, the demand in the wholesale market is vertical. He shows that in these circumstances there is a multiplicity of equilibria characterised by a decreasing supply function and that on average monopoly prices will result. This result supports the importance of introducing supply competition and, in general terms, of increasing the number of consumers paying pool related prices.

The third variation of the game relates to KM where end-users pay the wholesale prices (plus any additional costs incurred). On the assumption that marginal costs are zero, the differential equation representing the first-order condition identified earlier (equation 3) can then be written as

$$S'(p) = \frac{S(p)}{p} + D_p(p) \tag{9}$$

Assuming also that demand is linear in prices, as in equation 7 above, equation 9 will be an exact differential equation. One can then obtain the following equilibrium supply function

$$S(p) = Z\,p - \varepsilon\,p \ln p \tag{10}$$

where Z is an arbitrary constant of integration. This solution presents two essential features. First, a unique equilibrium does not exist – I could identify a supply function equilibrium for each possible value of Z. Second, for a given constant of integration, the supply function is not always an increasing function of prices. It increases up to a point and then decreases. Hence, for any given constant of integration, there may be more than one equilibrium solution in the range of positive price and quantities.

Newbery (1992) focuses on the application of the supply function equilibrium to the electricity spot market in England and Wales. He also assumes that marginal costs are zero and that demand is linear. His contribution is to devise an equilibrium selection criterion that addresses the issues identified by Bolle while explicitly introducing the role of capacity constraints. This is achieved by imposing a very specific form of behaviour. The equilibrium selection mechanism is based on the choice of a constant of integration with reference to the behaviour at full capacity output. Newbery assumes that firms will take the rivals' output as given when they are capacity constrained. Assuming that demand is linear, the Cournot schedule can be derived from equation 9 by setting $S'(p)$ to zero and solving it for $S(p_C) = K$. Newbery then gets that

$$K = p_c\,\varepsilon \tag{11}$$

The price for full capacity output is then K/ε. This pair of price and quantity is then substituted into the equilibrium supply function in equation 10. Solving for Z, he obtains

$$Z = \varepsilon\left[1 + \ln\frac{K}{\varepsilon}\right] \tag{12}$$

This expression determines a value of the constant of integration, Z, such that the supply function will intersect the Cournot schedule at full capacity output. This is then substituted into the supply function, equation 10, to yield the following equilibrium supply function

$$S(p) = \varepsilon\,p\left[1 + \ln\frac{K}{\varepsilon\,p}\right] \text{ for } p \le \frac{K}{\varepsilon} \tag{13}$$

This supply function explicitly incorporates capacity constraints and is an increasing function of p. Note that the supply function equilibrium is independent of the (uncertain) level of demand.

The main conclusions from the literature are the following. First, a supply function equilibrium may be well determined as shown in KM but, in practice, it may be difficult to identify the precise form of the equilibrium supply function. Second, the equilibrium supply function will be independent of the level of demand – this ensures that the supply function is ex-post optimal and valid for any level of demand. Third, the equilibrium supply function is an increasing function of prices, which suggests that the rules of the pool market auctioneer are broadly consistent with profit maximising behaviour.[9] Finally, these contributions emphasise firms' ability to drive prices substantially above marginal costs.

2.4 Equilibrium in supply functions when demands are interdependent

In this section, I develop an alternative model of supply function equilibrium that provides an alternative way to address the multiplicity of equilibria identified in the previous literature.

2.4.1 Model's assumptions

The model of supply function equilibrium presented here is based on similar assumptions about firms (e.g. zero marginal costs) and auctioneer as previous models. The model is based, however, on different assumptions about the demand aimed at capturing the features of electricity markets.

Suppose that demand is not a continuum of states as assumed in the previous literature and that it can be separated into two types of demand. So there is one period with high demand (period 1) and another period with low demand (period 2). Suppose also that the level of demand changes between these two periods in a predictable way. In addition, the quantity consumed in a given period depends on the price in that period and on the price prevailing

[9] Concerns have, nevertheless, been expressed about the calculation of prices in the pool market. One of the main concerns seems to be related to the definition of periods with excess capacity and its effect on the system marginal price, Lucas and Taylor (1994).

in the other period. This is particularly relevant to electricity consumption where demand during the day can be shifted in response to prices. Assuming that demand is linear in each price, I can represent demand in period *i* as

$$D_i(p_i;p_j)=a_i-\varepsilon p_i+\mu p_j \quad i\neq j;\ i,j=1,2 \tag{14}$$

where ε is a positive number that denotes the own price response, μ is the cross price response and $a_1>a_2$. Demands in periods 1 and 2 are imperfect substitutes – μ is a positive number. In addition, I assume that the size of the own price response is always larger than the cross price response, $\varepsilon>\mu$.

2.4.2 The model

There are two firms denoted as *M* and *N*. Each firm, say *M*, selects a price for each period such that it maximises its own profits

$$\pi_M=p_1S_M(p_1)+p_2S_M(p_2) \tag{15}$$

subject to demand being satisfied in each period

$$D_i(p_i;p_j)=S_M(p_i)+S_N(p_i) \quad i\neq j;\ i,j=1,2 \tag{16}$$

I focus on a Nash equilibrium so firm *M* selects its supply function taking into account the supply function selected by firm *N*. Firm *M* therefore maximises the following profit function

$$\pi_M=p_1\left[D_1(p_1;p_2)-S_N(p_1)\right]+p_2\left[D_2(p_2;p_1)-S_N(p_2)\right] \tag{17}$$

First-order conditions for profit maximisation require that for each period

$$\frac{\partial \pi_M}{\partial p_i}=\left[D_i(p_1;p_2)-S_N(p_i)\right]-p_i\left[\varepsilon+S'_N(p_i)\right]+\mu p_j=0 \tag{18}$$

Note that if I set μ to zero, this first-order condition is very similar to the one that results from KM approach – see equation 2. The only difference results from the different assumption about costs. I can rewrite each of the first-order conditions as

$$S_M(p_i)-p_i\left[\varepsilon+S'_N(p_i)\right]+\mu p_j=0 \tag{19}$$

I substitute in the above the following which is based on the equilibrium condition

$$S_M(p_i)+S_N(p_i)-a_i=-\varepsilon p_i+\mu p_j \tag{20}$$

This constitutes a step change with respect to KM approach. Consequently, the first-order condition, and the equilibrium supply function, will depend on the level of demand, a_i. This is now feasible because the level of demand is known unlike in KM approach. I can then rewrite the first-order condition as

$$2S_M(p_i)-p_iS'_N(p_i)+S_N(p_i)-a_i=0 \quad (21)$$

A Nash equilibrium in this model is a supply function and a pair of market clearing prices, $\hat{p}_1$ and $\hat{p}_2$, such that they satisfy for any supply function announced by the other firm (firm *N*) the following

$$2S_M(\hat{p}_1)-\hat{p}_1S'_N(\hat{p}_1)+S_N(\hat{p}_1)-a_1=0 \quad (22)$$

$$2S_M(\hat{p}_2)-\hat{p}_2S'_N(\hat{p}_2)+S_N(\hat{p}_2)-a_2=0 \quad (23)$$

Assuming that firms *M* and *N* are symmetric, I can rewrite the above equations as

$$3S(\hat{p}_1)-\hat{p}_1S'(\hat{p}_1)-a_1=0 \quad (24)$$

$$3S(\hat{p}_2)-\hat{p}_2S'(\hat{p}_2)-a_2=0 \quad (25)$$

These equations (or equations 22 and 23) do not constitute a system of two differential equations. They are not satisfied for any price as in KM approach. These equations show the relationship that equilibrium output and prices must satisfy for each period given a supply function. I cannot therefore solve them as differential equations.

I consider quadratic supply functions because by a suitable choice of parameters they can be consistent with the evidence of bidding in the pool (Helm and Powell (1992)) and with decreasing marginal returns to scale that characterise electricity generation. So I assume a quadratic supply function of the following form

$$S(p_i)=Ap_i+Bp_i^2+C \quad (26)$$

The (quadratic) supply function above reaches a maximum when *p* equals *-A/2B*. Equilibrium prices should be positive so in equilibrium either *A* or *B* must be positive. The parameter *C* must be negative so that the price for any level of output larger than zero is positive. In addition, I assume that firms will set the intercept so that they bid their full capacity output at the maximum price identified above. That is, *C* satisfies

$$K=A\left(\frac{-A}{2B}\right)+B\left(\frac{-A}{2B}\right)^2+C \quad (27)$$

After re-arranging I obtain that

$$C=K+\frac{A^2}{4B} \quad (28)$$

I substitute now the supply function in equation 26 into the first-order conditions (equations 24 and 25) and into the equilibrium conditions (equation 16). An equilibrium is then characterised by the simultaneous solution of the following five equations

$$2Ap_i + Bp_i^2 + 3C = a_i \tag{29}$$

$$a_i - \varepsilon\, p_i + \mu\, p_j = 2\left(Ap_i + Bp_i^2 + C\right) \tag{30}$$

for the two periods, $i = 1, 2$ and equation 28. There are therefore five equations and five parameters (A, B, C, p_i and p_j) and an equilibrium may exist.

Theorem 2 in Annex 2.1 shows that local second-order conditions for profit maximisation are satisfied by any quadratic supply function that solves 29 and 30, that has a positive slope over the relevant range and that it is not too concave (B is small in absolute terms).

2.4.3 Identifying the equilibrium solution

I proceed now to solve the above system of equations to identify an equilibrium solution. I start by simplifying it. Substituting 29 into 30, I get

$$\varepsilon\, p_i - \mu\, p_j + Bp_i^2 - C = 0 \tag{31}$$

This equation for i equals 1 is subtracted from the same equation for i equals 2. I obtain then

$$(\varepsilon + \mu) + B(p_1 + p_2) = 0 \tag{32}$$

This expression suggests that in equilibrium B must be negative. This is so because ε and μ are positive by assumption and because equilibrium prices are positive. It also suggests that in equilibrium A must be positive, otherwise the supply function will reach its maximum at a negative price.

I subtract now from equation 29 for $i = 2$ the same equation for $i = 1$. Combining the result with equation 32 yields

$$[2A - (\varepsilon + \mu)](p_1 - p_2) = (a_1 - a_2) \tag{33}$$

The equilibrium is now defined by the following five equations

$$2Ap_2 + Bp_2^2 + 3C = a_2 \tag{34}$$

$$\varepsilon\, p_1 - \mu\, p_2 + Bp_1^2 - C = 0 \tag{35}$$

$$(\varepsilon + \mu) + B(p_1 + p_2) = 0 \tag{36}$$

$$[2A - (\varepsilon + \mu)](p_1 - p_2) = (a_1 - a_2) \tag{37}$$

$$C = K + \frac{A^2}{4B} \tag{38}$$

To simplify further these equations I introduce the following change of variables. I set $A = x(\varepsilon + \mu)$; $B = -y(\varepsilon + \mu)^2$; and $q_i = p_i(\varepsilon + \mu)$, where x, y, q_i are positive. The system of equations above then becomes

$$2xq_2 - yq_2^2 + 3C = a_2 \tag{39}$$

$$\frac{\varepsilon q_1 - \mu q_2}{\varepsilon + \mu} - yq_1^2 - C = 0 \tag{40}$$

$$q_1 + q_2 = \frac{1}{y} \tag{41}$$

$$q_1 - q_2 = \frac{(a_1 - a_2)}{2x - 1} \tag{42}$$

$$C = K - \frac{x^2}{4y} \tag{43}$$

I combine now equations 41 and 42 to get an expression for the equilibrium prices. I obtain that

$$q_1 = \frac{1}{2y} + \frac{a_1 - a_2}{2(2x-1)} \tag{44}$$

$$q_2 = \frac{1}{2y} - \frac{a_1 - a_2}{2(2x-1)} \tag{45}$$

Result 1: (a) equilibrium prices are well defined for any given supply function (any values of x and y other $x = ½$); and (b) equilibrium prices in the high demand period (period 1) will be higher than in the low demand period.

Assuming that x is greater than ½, the above expressions suggest that non-negative equilibrium prices require that

$$y\frac{a_1 - a_2}{2x - 1} \leq 1 \tag{46}$$

I now substitute the above expressions for q_1 and q_2 into equations 39 and 40 to obtain the equilibrium values for x and y. I obtain then the following expressions

$$4x - 1 - 2y(a_1 + a_2) + 12Cy - y^2\left(\frac{a_1 - a_2}{2x - 1}\right)^2 = 0 \tag{47}$$

$$\frac{\varepsilon - 3\mu}{\varepsilon + \mu} - 4Cy - y^2\left(\frac{a_1 - a_2}{2x - 1}\right)^2 = 0 \tag{48}$$

Let us denote in the above equation

$$H \equiv \sqrt{\frac{\varepsilon - 3\mu}{\varepsilon + \mu}} \tag{49}$$

where $\varepsilon > 3\mu$. Note that where demands are substitutes, μ is positive, H is smaller than 1.

I substitute C from equation 43 into 47 and 48 and obtain

$$y^2\left(\frac{a_1 - a_2}{2x-1}\right)^2 + y\left[2(a_1 + a_2) - 12K\right] + 3x^2 - 4x + 1 = 0 \tag{50}$$

$$y^2\left(\frac{a_1 - a_2}{2x-1}\right)^2 + 4Ky - x^2 - H^2 = 0 \tag{51}$$

I need to solve now the two equations above for x and y to identify an equilibrium solution. Unfortunately, these equations cannot be solved analytically. I, therefore, need to characterise the equilibrium solution and to solve the equations by numerical methods.

By combining 50 and 51, I can obtain the following expression for y as a function of x

$$y = \frac{-(2x-1)^2 - H^2}{2(a_1 + a_2) - 16K} \tag{52}$$

This relationship must be satisfied by any equilibrium supply function (a pair of x and y). Note that if I allow no demand substitution ($\mu = 0$ and therefore $H = 1$), y and (hence B) will still be well defined.

Since x and y are positive by definition the denominator of the expression in equation 52 must be negative. This, in turn, implies that the level of capacity necessary for an equilibrium is

$$K > \frac{1}{4}\frac{a_1 + a_2}{2} \tag{53}$$

In a competitive equilibrium, each firm sets prices equal to marginal cost – zero by assumption. Thus, the equilibrium output in each period will be a_1 and a_2 respectively and $(a_1 + a_2) / 2$ will be the average output.

Result 2: a necessary level of capacity for an equilibrium to exist is that each firm has a capacity of at least a quarter of the average output under competitive conditions.

2.4.4 Further characterisation of the equilibrium solution

Annex 2.2 characterises in more detail the equilibrium solution. Here I summarise the main results.

First, I find that there are two possible equilibria (Result A1). Second, I refine further the requirement of capacity for an equilibrium to exist identified earlier (Result 2 above). In equilibrium, each firm will have capacity to supply, at least, half of the average output in a competitive equilibrium (Result A2). Both equilibria satisfy this requirement. Third, I select an equilibrium so that the equilibrium prices are not greater than the price for full capacity output. Only one of the two equilibria will satisfy this requirement (Result A3) and this equilibrium is unique (Result A4).

2.5 Discussion

This section discusses the equilibrium results of the model presented in the previous section. To that effect, first I study, using numerical methods, the comparative statics of the equilibrium solution. Second, I verify whether the necessary level of capacity identified in the previous section appears to be satisfied in practice. Third, I compare the equilibrium results of the model with alternative equilibria and quantify the size of the static welfare losses implied by a duopoly in supply function. Finally, I compare the equilibrium results of this model with those from Newbery's model.

2.5.1 Comparative statics

The system of equations that characterises the equilibrium solution has no analytical solution. So, the comparative statics of the equilibrium solution are illustrated by numerical methods – solving equations 50 and 51. The results of these simulations are summarised in Figure 2.3.

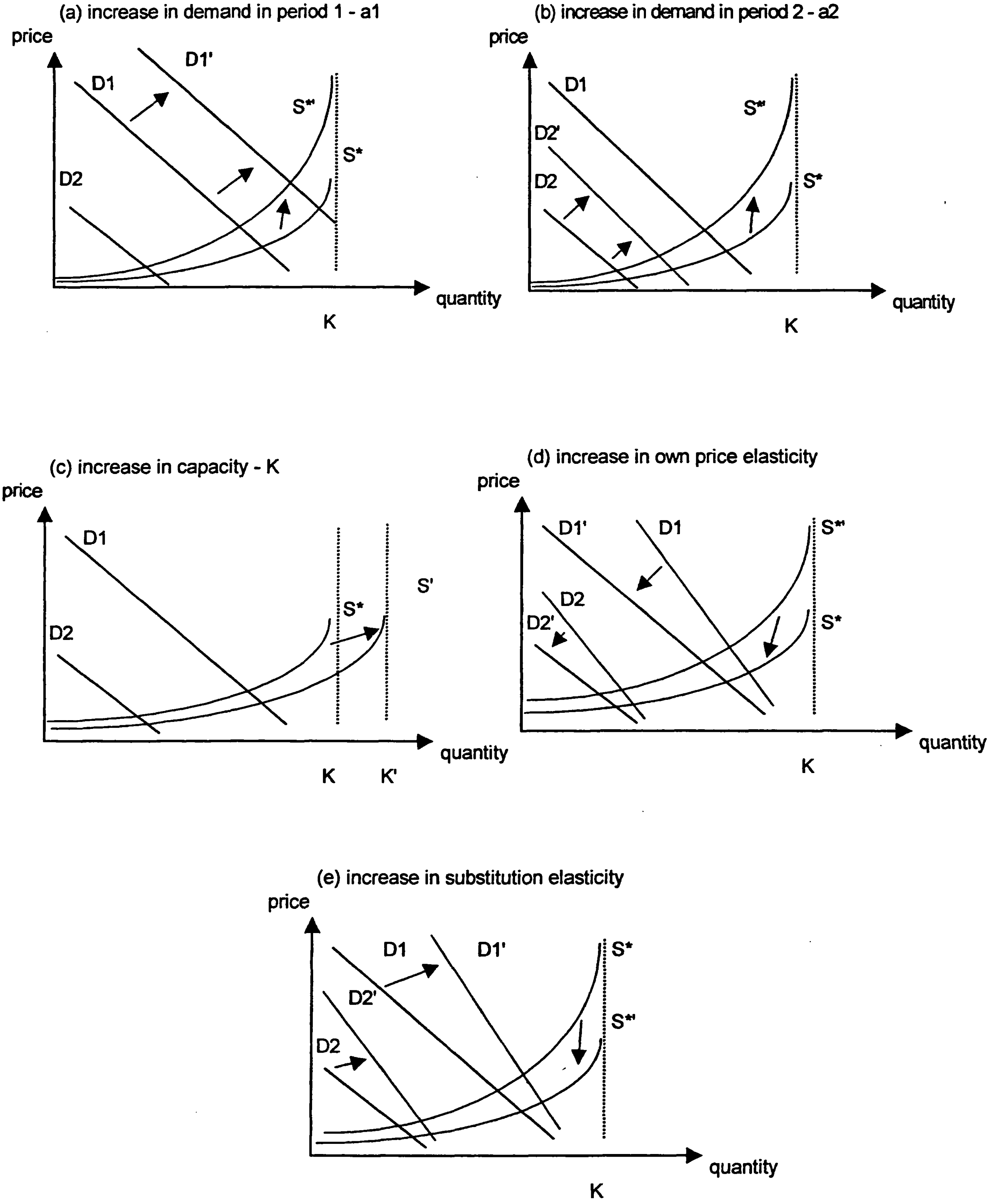
Panel (a) of Figure 2.3 illustrates the effect on the equilibrium of an increase (reduction) in the level of demand in period 1. This increase (reduction) leads to a shift of the equilibrium supply function to the left (right). The new equilibrium will be characterised by higher (lower) prices and quantities in period 1. The simulations suggest that the overall effect on period 2 is unclear, prices will be higher but the effect on quantities is ambiguous and depends on demand elasticities. An increase (reduction) in demand in period 2 has a similar effect – see panel (b). The equilibrium supply function shifts to the left (right). The new equilibrium will be characterised by higher (lower) prices and quantities in period 2. The effect in period 1 is to increase equilibrium prices and to reduce the equilibrium output. The numerical simulations also suggest that generation companies prefer an increase in peak demand to an increase in off-peak demand. Their profits will increase more as a result of an increase in demand in period 1 (peak) than as a result of an increase in demand in period 2 (off-peak).

A higher (lower) capacity shifts the equilibrium supply function to the right and leads to lower (higher) prices in both periods as shown in panel (c). The equilibrium profits will be lower (higher).

An increase (reduction) in the own-price elasticity affects the demand in both periods and makes them (steeper) flatter as shown in panel (d). The equilibrium supply function shifts to the right (left). As a result, the equilibrium prices and quantities are lower (higher) in both periods.

The effect of an increase (reduction) in the substitution elasticity is to make the quantity demanded in each period larger (smaller) for any given price. Both demands will be steeper (flatter) as shown in panel (e). The equilibrium supply function will shift to the right (left). Prices in period 1 will be lower (higher) and prices in period 2 will be higher (lower) as consumption is shifted to period 2. Equilibrium output increases slightly in both periods albeit for different reasons. The change in the equilibrium output is determined by two factors: the effect of changes in equilibrium prices and the effect of changes in substitution elasticity. In period 2, the dominant effect is the change in substitution elasticity. In period 1, the dominant effect is lower prices. Despite that, the resulting increase in total output is very small.

Figure 2.3: Comparative statics



The results from this model about changes in substitution elasticity can be compared with those from an experiment on time-of-use electricity pricing conducted through 1989/90, Henley and Pearson (1994). Households were given a vector of time-of-use prices; consumption levels were then recorded and compared with the consumption of similar households on normal tariffs. The result of the experiment could be compared with the effect on the equilibrium level of consumption that changing the substitution elasticity from zero to a positive number would have in our model. Henley and Pearson found that time-of-use pricing did not change significantly the average weekly consumption of electricity. It affected the pattern of consumption of electricity: consumption was shifted towards periods with low prices. The results from our model are broadly consistent; in both cases the total level of consumption is hardly affected when the substitution elasticity increases.

This discussion suggests that generation companies may not be willing to support measures aiming to even the demand profile since they will always profit more from an increase in peak demand than from an increase in off-peak demand. They will regard similarly measures aiming to increase the ability to shift demands between periods since their profits will be adversely affected. However, these incentive properties need to be kept in perspective since the ability to promote these measures lies outside the scope of generation companies. For example, the ability to shift demands between periods is mainly determined by the extent and the effectiveness of supply competition, which is also determined by other factors.

2.5.2 Necessary level of capacity

I identified the level of capacity necessary for an equilibrium in supply functions to exist (Result A2). I verify whether this condition appears to be satisfied in the case of England and Wales at the time of the restructuring of the electricity supply industry. Equation A45 in Annex 2.2 referred to the level of capacity of a firm. I could then verify that the total capacity available in the market would be

$$K' > \frac{a_1 + a_2}{2} \tag{54}$$

where K' denotes the total level of capacity available (equal to $2K$).

In a competitive equilibrium, each firm sets prices equal to marginal cost – zero by assumption. Thus, the equilibrium output in each period will be a_1 and a_2 respectively and $(a_1 + a_2) / 2$ will be the average output. The CEGB used to set prices based on a merit order that reflected marginal costs. I use, therefore, data from the last financial year of the CEGB (all the figures in this section have been taken from the prospectus of the offer of shares of National Power and PowerGen).

Electricity supplied during 1989/90 was 242,275 GWh or about 664 GWh per day on average. That is, the total capacity should be such that at least 332 GWh could have been generated. Net capacity by the end of 1989/90 was 58,471 MW, which will imply a potential daily output of 1,403 GWh.[10] This suggests that the requirement identified here for a symmetric duopoly equilibrium will have been satisfied.

2.5.3 Comparison with alternative equilibria

I compare the equilibrium result of the model developed here with alternative equilibria. I start therefore by characterising a competitive supply function equilibrium and a "Cournot" supply function equilibrium.

Consider a competitive supply function equilibrium. Each firm on its own is not able to exert any influence on the market price. Thus, the best it can do is to bid its marginal cost. Note that in this case, I am able to identify unambiguously the form of the optimal supply function: the marginal cost schedule. For example, when marginal costs are constant, the supply function will be horizontal up to the capacity constraint, from there on the supply function becomes vertical. Given the assumptions of the model, the equilibrium price will be zero and the firm's output $a_i / 2$. When the firm is capacity constrained, the equilibrium price is determined by demand.

In a Cournot equilibrium, each firm maximises its profit assuming that its decisions will not affect the decisions of its rival. I can derive the equilibrium solution by assuming that the slope of the rival's supply function is zero in equation 2 or 18 and solving the equilibrium. As indicated in KM (page 1258), the same equilibrium price and quantity can be derived by assuming that the firm sets output in period 1 and 2 taking as given the output of its rival.

---

[10] This is derived by multiplying the net capacity, 58,471 MW, by 24 (hours) and dividing the result by 1,000 (1,000MWh = 1GWh).

The equilibrium price is then

$$p_i = \frac{1}{3}\frac{\varepsilon a_i + \mu a_j}{\varepsilon^2 - \mu^2} \tag{55}$$

The equilibrium output is $a_i / 3$ when the firm is not capacity constrained. The main feature of this equilibrium is that the equilibrium quantities are independent of the demand in other periods and of the own price elasticity.

As indicated before, the equilibrium price and quantity that result from the supply function model lie between those of the competitive and the Cournot equilibrium. Table 2.2 illustrates this with the base case used in the comparative statics. It also provides an indication of the size of the price distortion introduced by competition in supply functions developed here. It suggests that the distortion will be relatively small given the large scope for increasing prices up to the Cournot level. The equilibrium prices and quantities of the supply function model are much closer to the competitive equilibrium than to the Cournot equilibrium. This is broadly in line with the empirical analysis of pool prices in Wolfram (1999). This concludes that prices are not nearly as high as the standard Cournot model or the supply function competition based on KM's framework predict.

Table 2.2: Comparison of different equilibria

Assumptions: $K = 16$ (firm's capacity), $a_1 = 12$ (intercept of peak-time demand), $a_2 = 2$ (intercept of off-peak time demand), $\varepsilon = 0.2$ (own price elasticity), $\mu = 0.05$ (cross demand elasticity) and zero marginal costs.

	Competitive equilibrium	Supply function equilibrium	Cournot equilibrium
Price			
Period 1	0	3.0	22.2
Period 2	0	0.5	8.9
Quantity per firm			
Period 1	6	5.7	4.0
Period 2	1	1.0	0.7

I have also estimated the reduction in consumers' surplus implied by the supply function equilibrium model developed here and compared it with the reduction implied by the Cournot equilibrium. The results are presented in Table 2.3 in money terms and as a percentage of the consumers' surplus in a competitive equilibrium.[11] The size of the total reduction in consumers' surplus is different in both equilibria. In our model of supply function equilibrium, losses in consumers' surplus constitute about 10% of the surplus of the competitive equilibrium, which compares with about 60% in the case of Cournot equilibrium. Furthermore, the dead-weight losses implied by this supply function equilibrium are minimal. Most of the reduction in consumers' surplus constitutes a transfer of resources to the producers, whereas in the case of Cournot competition the dead-weight loss constitutes about one-fifth of the reduction in consumers' surplus.

Table 2.3: Welfare losses - reduction in consumers' surplus

Assumptions: $K = 16$ (firm's capacity), $a_1 = 12$ (intercept of peak-time demand), $a_2 = 2$ (intercept of off-peak time demand), $\varepsilon = 0.2$ (own price elasticity), $\mu = 0.05$ (cross demand elasticity) and zero marginal costs.

	Supply Function Equilibrium		**Cournot equilibrium**	
	Money terms	**Percent (*)**	**Money Terms**	**Percent (*)**
Transfer to producers				
Period 1	34.4	8.6%	177.8	44.4%
Period 2	0.5	2.0%	11.9	44.4%
Total	34.9	8.2%	189.6	44.4%
Dead-weight loss				
Period 1	0.9	0.2%	44.0	11.1%
Period 2	0.3	1.0%	3.0	11.1%
Total	1.1	0.3%	47.4	11.1%
Reduction in consumers' surplus				
Period 1	35.2	8.8%	222.2	55.6%
Period 2	0.8	2.9%	14.8	55.6%
Total	36.0	8.4%	237.0	55.6%

(*) Percentage of the consumers' surplus in a competitive equilibrium
Note: figures may not add up due to rounding

Table 2.4 illustrates the reduction in consumers' surplus and average prices that results from changes in the substitution elasticity, μ, for various levels of own-price elasticity, ε. The results in the table also indicate that the larger the substitution elasticity the lower the welfare losses and the lower the average prices that result from supply function competition. This confirms the view that the ability to shift demands between periods reduces the scope of duopolists to increase prices over marginal costs. The table also shows

[11] This is calculated as the area under the demand function assuming that the price in the other period is set at the competitive level.

that the larger the own price elasticity (or the flatter the demand) the larger the welfare losses, despite equilibrium prices being smaller. This is so because the flatter the demand the smaller the equilibrium output and the potential consumers' surplus.

Table 2.4: Welfare losses - reduction in consumers' surplus as a percentage of the consumers' surplus in a competitive equilibrium

Assumptions: $K = 16$ (firm's capacity), $a_1 = 12$ (intercept of peak-time demand), $a_2 = 2$ (intercept of off-peak time demand) and zero marginal costs					
	Supply function equilibrium				**Newbery (1992)**
Cross price elasticity, μ	0	0.03	0.05	0.06	0 (implicit)
Own price elasticity, $\varepsilon = 0.2$					
Period 1	11.1%	9.8%	8.8%	8.3%	24.5%
Period 2	7.2%	3.9%	2.9%	2.6%	12.0%
Total	11.0%	9.5%	8.4%	7.9%	24.1%
Average price (*)	3.0	2.8	2.6	2.6	6.8
Own price elasticity, $\varepsilon = 0.3$					
Period 1	11.1%	10.2%	9.6%	9.3%	24.5%
Period 2	7.1%	4.7%	3.7%	3.4%	12.0%
Total	11.0%	10.0%	9.3%	9.0%	24.1%
Average price (*)	2.0	1.9	1.8	1.8	4.5

(*) Weighted average of equilibrium prices in period 1 and 2. Weights used are the equilibrium quantities.

2.5.4 Comparison with Newbery's model

I have compared the results of the supply function equilibrium model developed here with the results of the model developed in Newbery (1992) and summarised in Section 2.3. The form of that equilibrium supply function (equation 13) is such that the equilibrium price cannot be solved analytically. It has been solved by numerical methods using the same base case that was used in previous tables.

I have conducted a range of simulations to illustrate the comparative statics of Newbery's model. Most of the results are not substantially different from those presented in Section 2.5.1. The main difference is associated with changes in the price elasticity, ε, and I discuss this here.

An increase in the price elasticity shifts the supply function to the right, the equilibrium prices in both periods are smaller – see Table 2.5. However, the equilibrium output remains unchanged. Theorem 2 in Annex 2.1 confirms this intuition and shows that at equilibrium $\frac{\partial S(p)}{\partial \varepsilon}$ equals to zero. This is not, however, a structural feature of KM approach. It is the result of the solution approach in Newbery (1992) and of the use of the Cournot price for full capacity output to identify a solution for the differential equation that represents the first-order conditions (equations 10 to 12 in Section 2.3). Theorem 3 in Annex 2.1 shows that only when the price for full capacity output is taken from the Cournot schedule the equilibrium output will be independent of the own price elasticity.

Table 2.5: Comparative statics of the equilibrium results from Newbery (1992)

Assumptions: K = 7 (firm's capacity), a_1 = 5 (intercept of peak-time demand), a_2 = 1 (intercept of off-peak time demand) and zero marginal costs.

Own price elasticity, ε	Prices Period 1	Prices Period 2	Quantities Period 1	Quantities Period 2
0.10	15.70	1.65	5.21	0.92
0.20	7.85	0.82	5.21	0.92
0.25	6.28	0.66	5.21	0.92
0.30	5.23	0.55	5.21	0.92
0.35	4.49	0.47	5.21	0.92

Comparing the equilibrium from Newbery's model with the supply function equilibrium developed here – Table 2.4 – shows that equilibrium prices and welfare losses are of a different order of magnitude. (Note that welfare losses predicted by Newbery's model will be unaffected by own price elasticity since the equilibrium output is independent of it.) The losses in consumers' welfare predicted by Newbery's model as well as the predicted equilibrium prices are higher than those predicted by the model developed here (even if demands are independent). The difference should be attributed to the different modelling approach. The intuition of this result is as follows. The supply function from Newbery and KM approach are valid for any possible level of demand – however high. Thus, given a capacity constraint, a high price will be necessary to choke-off the possible high demand. The equilibrium prices will be high even if that high level of demand will not materialise. However, when a supply function equilibrium takes explicit account of the information about demand, the equilibrium prices and welfare losses are likely to be substantially smaller.

2.6 Qualifications

I have developed here a very specific model of supply function equilibrium with two identical firms, where there are two types of demand (peak and off-peak), firms estimate accurately the demand and use that information in their bidding strategy. The purpose of this section is to reflect on the assumptions of the model and on how one might consider extending the model.

First, the specific assumptions of the model such as the number of firms and the demand function. The model has been developed on the assumption that demands in period 1 and 2 are substitutes. However, there may be situations where the notion of supply function equilibrium is also relevant and demands are complement – the parameter μ in the demand, equation 14, is negative. The model developed here is also relevant to that case. In this case, second-order conditions for equilibrium are also satisfied – Theorem 1 in Annex 2.1 – and the equilibrium condition is identified – Annex 2.2. Extending the model for N identical firms will also be valuable. The general result that incorporating demand information into the bids results in an equilibrium that is qualitatively different from that developed in the literature using KM's work should remain valid.

Second, a capacity constraint is also one of the parameters of the model. In reality, capacity will be determined by the interaction of market forces. One could build the choice of capacities into the model by extending it into a two-stage model where firms simultaneously choose capacity levels in the first stage, and they compete in prices in the second stage. This two-stage approach can yield interesting results as in Kreps and Scheinkman (1983). They show that (under certain conditions) the unique equilibrium outcome is the Cournot outcome. This is despite the existence of price competition in the second stage of the game. The general intuition of this result is that solutions to oligopoly games depend both on the strategic variables that firms (are assumed to) employ and on the form of the game in which these variables are employed.[12]

[12] Indeed, the results of a two stage model of capacity choice (stage one) and price competition (stage two) are sensitive to the assumptions about the rationing rule used to model behaviour when one firm is already constrained by its capacity. Using the terminology in Tirole (1988), it matters whether one assumes the "efficient-rationing rule" (residual demand is parallel to the aggregate demand) or the "proportional-rationing rule" (residual demand is proportional to the aggregate demand). Kreps and Scheinkman (1983) used the efficient-rationing rule. Davidson and Deneckere (1986) show that Cournot's equilibrium is unlikely to emerge in a similar two-stage model when one assumes the proportional-rationing rule.

Third, the model assumes that there is no demand uncertainty. In reality, firms are unlikely to know so much about the demand as assumed here and as little as assumed in the applications of KM framework. So an alternative way of extending the model is to consider the role of demand uncertainty. Here the key modelling consideration is how uncertainty matters. Two broad approaches to incorporate uncertainty are discussed here. One could assume that the intercept of the demand is a random variable. Green (1999a) considers this approach in the context of a model of supply function equilibrium based on KM framework. The paper suggests that relaxing this assumption about the demand would not affect the results on strategies or average prices unless some agents are risk averse. So one could extend the model developed here by assuming that there is some uncertainty about the demand and consumers are risk averse (more on this below). An alternative is Kay (1979) where uncertainty is such that "capacity and price must generally be determined before the level of demand is known". So the same price is charged for peak and off-peak demand. This notion does not sit well with the concept of a supply function equilibrium where prices vary with demand but one could focus on the effect of demand uncertainty on capacity choices. It seems that this could be modelled in three stages. In a first stage, firms choose capacities given some uncertainty about future demand. In a second stage, the uncertainty is realised. And in a third stage, firms bid in the spot market given the realisation of demand and capacity choices. This approach is not very different from the one used in the next chapter where entry decisions are made before the environmental tax is set and production decisions are made after the tax has been set.

Finally, forward sales in the contract market are one of the institutional features of the market – see Section 2.2.1 – but they have not been explicitly modelled. It would be useful to extend the model presented here to allow explicitly for the link between the spot market (based on the concept of supply function equilibrium) and the contract market. This matters because one would expect sales in the contract market to reduce the incentives to push up prices in a spot market, Allaz and Vila (1993). One could extend the model presented in this chapter into a multi-stage model where in a first stage firms sell in a forward market and in a second stage firms bid in the spot market using a supply function framework. Green (1999a) adopts this approach, using KM framework for the spot market. He shows that "generators may well hedge most of their output with forward contract sales", which removes much of the incentive identified using KM framework to raise prices above marginal costs. A supplement to that paper, Green (1999b), considers the case where there is uncertainty about the demand. It is assumed that the intercept of the demand – which has the same form as equation 7 – is a random variable (with known and small variance). Buyers are risk averse so the forward price exceeds the expected price in the spot market by a hedging premium. In this case, introducing uncertainty does not change the general conclusion from Green (1999a) and results in generators having an increased incentive to sell in the contract market.

2.7 Conclusion

I have developed a specific model of supply function competition, which encapsulates some features of the wholesale electricity markets not present in the previous applications of the supply function equilibrium. These features relate to the structure of demand. First, demand is discrete and varies within a day in a predictable fashion. Second, demands during a day are not independent and I have assumed that they are substitutes. Third, firms use the information about demand in their bids.

I have shown that under these conditions, an equilibrium is well determined and unique for a broad class of quadratic supply functions when capacity constraints are explicitly introduced. The equilibrium has been solved by numerical methods. The comparative statics of the model are broadly consistent with evidence of an experiment of time of day pricing, Henley and Pearson (1994). In both cases, allowing for demand substitution hardly affects the level of consumption. It shows that demand interdependency reduces the ability of firms to raise prices above marginal costs.

The equilibrium results have been compared with an alternative equilibrium based on Cournot's assumptions. The results suggest that the distortion introduced by the model of supply function competition developed here will be relatively small. This is in line with the empirical findings in Wolfram (1999) that the actual prices in the wholesale electricity market "are not nearly as high as most theoretical models predict". The equilibrium results have also been compared with equilibrium predictions based on Newbery (1992). Prices and welfare losses implied by the model developed here are substantially smaller. This is not only related to the absence of demand interdependency in Newbery (1992) but also to the different modelling approach adopted here. The results of this model question therefore the ability of generating companies to raise prices above marginal costs. This suggests that the bidding mechanism that lies at the core of the pool market arrangements may be relatively efficient since it introduces a relatively small price distortion (which is certainly smaller than what is implied by other contributors).

This conclusion should not be seen as suggesting that these firms will not enjoy "supra-normal" profits. The analysis is based on a stylised model of the spot electricity market with simplifying assumptions that could be relaxed in various directions, such as the choice of capacity and the interaction with the contract market. However, there seems to be an important policy implication that can be derived from this model. It relates to the areas that might need to be considered in any attempt to make the wholesale electricity market more competitive. This analysis suggests that these attempts should not focus on challenging the principle of a bidding system that prices electricity according to the price of the most expensive plant in use in the system. It should focus on the role of the additional payments such as capacity payments, the mechanism for the determination of periods of excess capacity and the transparency of the contract market. This is in broad terms the direction of change in countries other than Britain where similar systems already exist, Ruff (1999). It is in contrast with the changes introduced in 2001 in England and Wales, see, for example Offer (1998). They extend the system of bilateral trading to the spot market and should make the forward market in electricity more transparent and liquid. This suggests that there may be cases where the combined importance of the associated issues is such that it affects the overall direction of the policy.

Annex 2.1: Proof of theorems

Theorem 1: second-order conditions for profit maximisation are satisfied if the supply function has a positive slope within the relevant range and it is not too concave (*B* is small in absolute terms).

Proof: the derivative of the marginal profit (equation 18 in the text) in period *i* with respect to the own price is

$$\frac{\partial^2 \pi_M}{\partial p_i^2} = -2[\varepsilon + S'_N(p_i)] - p_i S''_N(p_i) \qquad \text{(A1)}$$

Assuming a quadratic supply function as in equation 26 in the text, I obtain that

$$\frac{\partial^2 \pi_M}{\partial p_i^2} = -2\varepsilon - 2(A + 2Bp_i) - 2Bp_i \qquad \text{(A2)}$$

The term in brackets is the slope of the supply function, at the equilibrium this is positive whereas *B* is negative. Thus, the expression will be negative as long as the supply function is not too concave (*B* is small in absolute terms).

The derivative of the marginal profit with respect to the price in the other period is

$$\frac{\partial^2 \pi_M}{\partial p_i \partial p_j} = 2\mu \qquad \text{(A3)}$$

Note that this expression is always positive (negative) since I assumed that demands are substitutes (complements). Thus, it only remains to be shown that

$$\frac{\partial^2 \pi_M}{\partial p_1^2}\frac{\partial^2 \pi_M}{\partial p_2^2} - \left(\frac{\partial^2 \pi_M}{\partial p_i \partial p_j}\right)^2 > 0 \qquad \text{(A4)}$$

Substituting A1 and A3 implies that

$$[2\varepsilon + 2S'_N(p_1) + p_1 S''_N(p_1)][2\varepsilon + 2S'_N(p_2) + p_2 S''_N(p_2)] - 4\mu^2 > 0 \qquad \text{(A5)}$$

This expression is always satisfied at equilibrium since the slope of the supply function is positive, $S''_N(p_i)$ is negative and $\varepsilon > \mu$. This condition will also be satisfied if demands are complement (negative μ). QED.

Theorem 2: the equilibrium output in Newbery (1992), $S(\tilde{p})$, is unaffected by changes in the slope of the demand.

<u>Proof</u>: I want to show that

$$\frac{dS(\tilde{p})}{d\varepsilon} = 0 \tag{A6}$$

for $S(p)$ as in equation 13 of the main text.

I use the envelope theorem to prove that this expression is correct. I need to show that for the equilibrium price, $\tilde{p}$, the supply function in equation 13 satisfies that

$$\frac{\partial S(p)}{\partial p}\frac{\partial p}{\partial \varepsilon} + \frac{\partial S(p)}{\partial \varepsilon} = 0 \tag{A7}$$

The first and the third term in the expression above can be derived by partial differentiation of the equilibrium supply function. This yields

$$\frac{\partial S(p)}{\partial p} = \varepsilon \ln\left(\frac{K}{\varepsilon p}\right) \tag{A8}$$

$$\frac{\partial S(p)}{\partial \varepsilon} = p \ln\left(\frac{K}{\varepsilon p}\right) \tag{A9}$$

To obtain $\frac{\partial p}{\partial \varepsilon}$, I need an expression for the equilibrium price, $\tilde{p}$. This satisfies the following equation

$$2S(\tilde{p}) = D(\tilde{p},\theta) \tag{A10}$$

I substitute in this expression the demand function and the equilibrium supply function. These are equation 7 and 13 in the text respectively. After re-arranging, I obtain that

$$K = \varepsilon\tilde{p}\, e^{\frac{\theta}{2\varepsilon\tilde{p}}} e^{-\frac{3}{2}} \tag{A11}$$

Note that an expression for $\tilde{p}$ cannot be derived from the above. So to obtain $\frac{\partial \tilde{p}}{\partial \varepsilon}$, I define the following implicit function

$$\phi(\tilde{p},\varepsilon,K,\theta) = K - \varepsilon\tilde{p}\, e^{\frac{\theta}{2\varepsilon\tilde{p}}} e^{-\frac{3}{2}} \tag{A12}$$

Using the total derivative of this implicit function and assuming that K and θ are constant, I obtain that

$$\frac{\partial \tilde{p}}{\partial \varepsilon} = -\frac{\partial\phi/\partial\varepsilon}{\partial\phi/\partial\tilde{p}} = -\frac{\tilde{p}}{\varepsilon} \tag{A13}$$

I substitute now equations A8, A9 and A13 in equation A7

$$\frac{dS(\tilde{p})}{d\varepsilon}=\left(-\frac{\tilde{p}}{\varepsilon}\right)\varepsilon\ln\left(\frac{K}{\varepsilon\tilde{p}}\right)+\tilde{p}\ln\left(\frac{K}{\varepsilon\tilde{p}}\right)=0 \tag{A14}$$

QED.

Theorem 3: (a) the supply function equilibrium in Newbery (1992) – equation 13 in the text – can be generalised for any price for full capacity output; (b) only when the price for full capacity output is the Cournot's price – K/ε – the equilibrium level of output is independent of the own price elasticity.

Proof:

(a) derivation of a generalised version of the supply function equilibrium in Newbery (1992).

I start from the solution to the differential equation (equation 10 in the text). This is

$$S(p)=Z\,p-\varepsilon\,p\ln(p) \tag{A15}$$

where Z is the arbitrary constant of integration. Let us denote p_{FC} the price for full capacity output, so $S(p_{FC})=K$. Substituting this and solving for Z yields

$$Z=\frac{K}{p_o}+\varepsilon\ln(p_{FC}) \tag{A16}$$

I substitute this back into A15 and get

$$S(p)=\frac{K\,p}{p_{FC}}+\varepsilon\,p\ln\left(\frac{p_{FC}}{p}\right)\quad p\leq p_{FC} \tag{A17}$$

Note that if I substitute K/ε for p_{FC}, I obtain equation 13 in the text.

(b) given an equilibrium price, $\tilde{p}'$, the equilibrium output, $S(\tilde{p}')$, is independent of the own price elasticity, ε, only if the price for full capacity output in A17 equals the Cournot price.

I need to show that, given the supply function in A17, in equilibrium

$$\frac{dS(p)}{d\varepsilon}=\frac{\partial S(p)}{\partial p}\frac{\partial p}{\partial\varepsilon}+\frac{\partial S(p)}{\partial\varepsilon}=0 \tag{A18}$$

when $p_{FC} = K / \varepsilon$. The first and the third term in the expression above can be derived by partial differentiation of the equilibrium supply function in equation A17. This yields

$$\frac{\partial S(p)}{\partial p} = \frac{K}{p_{FC}} + \varepsilon \ln\left(\frac{p_{FC}}{p}\right) - \varepsilon \qquad \text{(A19)}$$

$$\frac{\partial S(p)}{\partial \varepsilon} = p \ln\left(\frac{p_{FC}}{p}\right) \qquad \text{(A20)}$$

To obtain $\frac{\partial \tilde{p}}{\partial \varepsilon}$, I need an expression for the equilibrium price. An equilibrium price is a price, $\tilde{p}'$, that satisfies that

$$2S(\tilde{p}') = D(\tilde{p}', \theta) \qquad \text{(A21)}$$

I substitute in this expression the demand function – equation 7 in the text – and the equilibrium supply function – equation A17 above.

$$\tilde{p}'\left(\frac{2K}{p_{FC}} + \varepsilon\right) + 2\varepsilon\,\tilde{p}' \ln\left(\frac{p_{FC}}{\tilde{p}'}\right) = \theta \qquad \text{(A22)}$$

Note that $\tilde{p}'$ cannot be derived from this expression. To obtain $\frac{\partial \tilde{p}'}{\partial \varepsilon}$, I define the following implicit function

$$\phi(\tilde{p}', p_{FC}, \varepsilon, K, \theta) = \tilde{p}'\left(\frac{2K}{p_{FC}} + \varepsilon\right) + 2\varepsilon\,\tilde{p}' \ln\left(\frac{p_{FC}}{\tilde{p}'}\right) - \theta = 0 \qquad \text{(A23)}$$

Using the total derivative of the implicit function defined above and assuming that K and θ are constant, I obtain that

$$\frac{\partial \tilde{p}'}{\partial \varepsilon} = -\frac{\partial \phi / \partial \varepsilon}{\partial \phi / \partial \tilde{p}'} = -\frac{\tilde{p}'\left(1 + \ln\left(\frac{p_{FC}}{\tilde{p}'}\right)\right)}{\frac{2K}{p_{FC}} - \varepsilon + 2\varepsilon \ln\left(\frac{p_{FC}}{\tilde{p}'}\right)} \qquad \text{(A24)}$$

If I substitute K / ε for p_{FC} in equations A19, A20 and A24. I obtain then the same derivatives that I obtained in Theorem 2. Thus, the equilibrium level of output will be independent of the own price elasticity only if the price for full capacity is the Cournot price. QED.

Annex 2.2: Further characterisation of the equilibrium solution

To characterise further the equilibrium solution, we start with equations 50 and 51 in the main text. Consider the following substitution of variables

$$u = 2x - 1 \tag{A25}$$

$$v = 2y\frac{a_1 - a_2}{2x - 1} \tag{A26}$$

$$\alpha = \frac{a_1 + a_2}{a_1 - a_2} \tag{A27}$$

$$\phi = \frac{8K}{a_1 - a_2} \tag{A28}$$

Using these expressions, I can rewrite equations 50 and 51 in the text as follows

$$v^2 + uv(4\alpha - 3\phi) + 4u^2 - (u+1)^2 = 0 \tag{A29}$$

$$v^2 + \phi vu - (u+1)^2 - 4H^2 = 0 \tag{A30}$$

An equilibrium is a solution to equations A29 and A30 for positive values of u and v.

This change of variable permits us to re-arrange the equations as explicit forms of u and v that can be analysed to characterise the equilibrium solution.

From equation A29, I can derive an expression for v^2 - $(u + 1)^2$ which I then substitute in equation A30. After re-arranging, I obtain

$$v = \frac{1}{\phi - \alpha}\left(u + \frac{H^2}{u}\right) \tag{A31}$$

Using equations A27 and A28, it can be shown that the requirement that $(\phi - \alpha)$ be positive leads to the capacity requirement identified in equation 53 in the text. This is always satisfied in the equilibrium. Thus, v will be positive if u is positive.

I derive now u as an explicit function of v. From equation A31, I can derive the following expression

$$vu(\phi - \alpha) = u^2 + H^2 \tag{A32}$$

This is then substituted in equation A30. After re-arranging, I get the following expression for u as function of v

$$u = \frac{v^2 - \left(1 + 3H^2\right)}{2 - v\alpha} \qquad \text{(A33)}$$

The function has an asymptote for v equal to $2 / \alpha$ (smaller than 2 since α is greater than 1). In addition, the requirement of positive equilibrium prices (see equations 46 in the text and A26 here) implies that this function will be defined only for v smaller than 2. The value of u, however, will not be unambiguously positive even when v is positive (and smaller than 2). I need therefore to identify the conditions for u to be positive. This requires that either the numerator and the denominator are positive in which case $v > \sqrt{(1+3H^2)}$ and $v < 2 / \alpha$ or that both are negative in which case $v < \sqrt{(1+3H^2)}$ and $v > 2 / \alpha$.

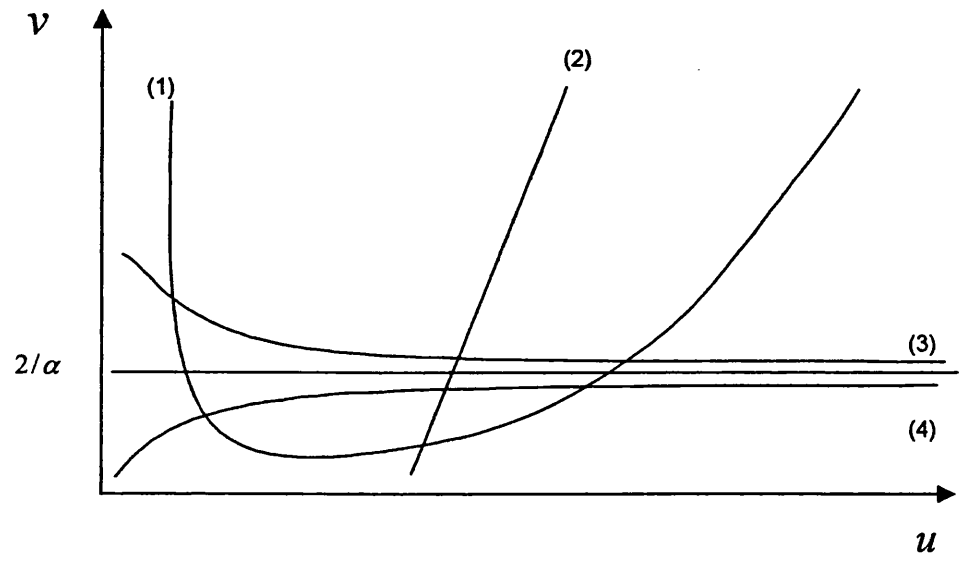
To determine which of the cases above is relevant and characterise the equilibrium I need to establish whether $\sqrt{(1+3H^2)}$ is greater or smaller than $2/\alpha$. Unfortunately, there is a degree of overlap between these two expressions depending on the underlying parameters.[13] This suggests that there is not a unique way to plot equation A33 and that I need to consider both cases above. If $\sqrt{(1+3H^2)} > 2 / \alpha$, equation A33 will converge to $2 / \alpha$ from above. In the alternative case, it will converge from below. Figure A2.1 shows the plot of equations A31 and A33 on the two assumptions about $\sqrt{(1+3H^2)}$ and $2 / \alpha$.

Result A1: regardless of the prevailing assumption about the parameters, if an equilibrium exists there will be two possible equilibria for positive values of u and v.

I need to identify, first, under which conditions an equilibrium will exist. In these circumstances if an equilibrium exists, there will be two feasible equilibria and both will satisfy the requirement of positive prices. Second, I need to establish a selection mechanism between the two feasible equilibria.

[13] Note that if $H \to 1$ (i.e. $\mu \to 0$) then $\sqrt{1+3H^2} \to 2$. If $H \to 0$ (i.e. $\varepsilon \to 3\mu$) then $\sqrt{1+3H^2} \to 1$. The other boundary, $2 / \alpha$, can be written using equation A27 as $2 (1 - \tau) / (1 + \tau)$ where τ is a number between 0 and 1 such that $a_2 = \tau a_1$. Thus, if $\tau \to 1$ then $2 / \alpha \to 0$ and if $\tau \to 0$ then $2 / \alpha \to 2$.

Figure A2.1: Equilibrium solution



Notes: (1) represents equation A31.
(2) represents equation A47 – the restriction that prices should be smaller than the price for full capacity output. This is satisfied in the area to the right of the line.
(3) and (4) equation A33. Note that they cut the vertical axis at v = √(1+3H^2). They converge to the asymptote 2 / α.

Figure A2.1 suggests that the condition for the existence of an equilibrium is that the minimum of the function denoted by equation A31 is sufficiently low to intersect with A33. The minimum of the function in equation A31 occurs when u is equal to H; the value of the function at that point is therefore

$$v_{\min} = \frac{2H}{\phi - \alpha} \qquad \text{(A34)}$$

The overlap between 2 / α and √(1+3H^2) implies that the analysis of whether the minimum is "sufficiently low" needs to be carried out with respect to the two possibilities.

Consider first the case where

$$\sqrt{1+3H^2} > \frac{2}{\alpha} \qquad \text{(A35)}$$

I need to show that the minimum, $v_{\min}$, is below the asymptote 2 / α. That is

$$\frac{2H}{\phi - \alpha} < \frac{2}{\alpha} \qquad \text{(A36)}$$

Re-arranging the above with the help of equations A27 and A28, I obtain that for an equilibrium to exist the following must be satisfied

$$\frac{8K}{a_1+a_2}-1>H \tag{A37}$$

Equation 53 in the text identified the necessary level of capacity for an equilibrium, I can therefore express any equilibrium level of capacity as

$$K\equiv\frac{a_1+a_2}{8}(1+\delta) \tag{A38}$$

where δ denotes the excess of capacity with respect to the requirement of equation 53 in the text. I can then substitute this expression into equation A37. Thus, given equation A35 the excess capacity (with respect to the necessary condition in equation 53 in the text) for an equilibrium to exist is

$$\delta>H \tag{A39}$$

Consider now the case where

$$\sqrt{1+3H^2}<\frac{2}{\alpha} \tag{A40}$$

In this case, equation A33 will converge to $2/\alpha$ from below. I need to show then that the minimum point, v_{min}, of equation A31 is below $\sqrt{(1+3H^2)}$. The latter is the lowest value that I could input to equation A33. That is,

$$\frac{2H}{\phi-\alpha}<\sqrt{1+3H^2} \tag{A41}$$

Re-arranging the above using equations A27 and A28, I get the following equilibrium requirement

$$\frac{8K-(a_1+a_2)}{2(a_1-a_2)}>\sqrt{\frac{H^2}{1+3H^2}} \tag{A42}$$

I substitute the expression in equation A38 here. I obtain that the necessary excess of capacity for an equilibrium to exist given equation A40 is

$$\delta>\frac{2}{\alpha}\sqrt{\frac{H^2}{1+3H^2}} \tag{A43}$$

Note that the first term in the right hand side of this expression is smaller than 1 (α is greater than 1) and so is the second one (see footnote 13). The right hand side is therefore smaller than 1.

I have obtained a condition for an equilibrium to exist, A39 and A43, given each of the alternative configurations of the parameters, A35 and A40. These conditions have been expressed in terms of the excess capacity with respect to the necessary capacity requirement identified in Result 2 in the main text. The terms identified in the right hand side of both conditions are both smaller than 1. So a sufficient condition for an equilibrium to exist will be

$$\delta \geq 1 \tag{A44}$$

Combining this with equation A38, I can refine the capacity requirement that I identified previously, equation A53. I obtain that each firm needs to hold a level of capacity such that

$$K \geq \frac{1}{2}\frac{a_1 + a_2}{2} \tag{A45}$$

Result A2: a sufficient condition for an equilibrium to exist is that each firm holds capacity to supply, at least, half of the average demand in a competitive market. (This supersedes Result 2 in the main text.)

If any of the above conditions are satisfied I will have two possible equilibria with positive equilibrium prices. I select between these equilibria, for example, by verifying that equilibrium prices are not greater than the price for full capacity output, *-B/2A*. I verify this for period 1 equilibrium price, equation 44 in the text. This implies that the following must be satisfied

$$\frac{1}{2y} + \frac{a_1 - a_2}{2(2x-1)} < \frac{x}{2y(\varepsilon + \mu)} \tag{A46}$$

Using equations A25 and A26, I can re-arrange this expression and get

$$v < -\left(\frac{1}{\varepsilon + \mu} + 2\right) + \frac{u}{(\varepsilon + \mu)} \tag{A47}$$

This equation has also been plotted in Figure A2.1 as an equality. It suggests that the requirement that equilibrium prices should be smaller than the price for full capacity output is only satisfied in the area to the right of the line.

Result A3: only one of the two possible equilibria constitutes an equilibrium.

Finally, I need to show now that the equilibrium is always unique: there is always one equilibrium located on the right of the line denoted by equation A47 in Figure A2.1. Given the u-shape of equation A31 it is enough to show that the intersection of equation A47 with the u-axis will be to the right of the value for which equation A31 has a local minimum. This function has a minimum at $u = H$. Equation A47 intersects the u-axis at $1 + 2(\varepsilon + \mu)$.

Result A4: a unique equilibrium is always identified when demands are substitutes (H smaller than 1).

I have modelled the supply function equilibrium with interdependent demands on the assumption that μ is positive (demands are partial substitutes). The same logic that I have developed in this sub-section can be used to characterise an equilibrium if demands are assumed to be complements, μ is negative (H is then greater than 1). The second-order conditions for profit maximisation will not be affected if demands are complements (see Theorem 1 in Annex 2.1). The only possibility is then $\sqrt{(1+3H^2)} > 2 / \alpha$.[14] I obtain that the necessary condition for an equilibrium is

$$\delta > H \tag{A48}$$

However, the uniqueness of the equilibrium will not be guaranteed any longer.

[14] If $H > 1$ then $\sqrt{1+3H^2} > 2$ and $2 / \alpha < 2$ because $\alpha > 1$.

References

Allaz, B. and J. -L. Vila (1993), "Cournot competition, futures markets and efficiency", *Journal of Economic Theory*, vol. 59, pp. 1 – 16

Bolle, F. (1992), "Supply function equilibrium and the danger of tacit collusion: the case of spot markets for electricity", *Energy Economics*, vol. 14, pp. 94 – 102

Davidson, C. and R. Deneckere (1986), "Long-run competition in capacity, short-run competition in price and the Cournot model", *Rand Journal of Economics*, vol. 17, pp. 404 – 415

Green, R. (1991), "Reshaping the CEGB: electricity privatisation in the UK", *Utilities Policy*, March, pp. 245 – 254

Green, R. (1999a), "The electricity contract market in England and Wales", *Journal of Industrial Economics*, vol. 42, pp. 107 – 124

Green, R. (1999b), "Supplementary materials to 'The Electricity contract market in England and Wales'", available from the editorial web site of the *Journal of Industrial Economics*, www.Stern.nyu.edu/~jindec/

Helm, D. and A. Powell (1992), "Pool prices, contracts, and regulation in the British electricity supply industry", *Fiscal Studies*, vol. 13, pp. 89 – 105

Henley, A. and J. Pearson (1994), "Time-of-use electricity pricing: evidence from a British experiment", *Economic Letters*, vol. 45, pp. 421 – 426

Kay, J. (1979), "Uncertainty, congestion and peak load pricing", *Review of Economic Studies*, vol. 46, pp. 601 – 611

Klemperer, P. D. and M. A. Meyer (1989), "Supply function equilibria in oligopoly under uncertainty", *Econometrica*, vol. 57, pp. 1243 – 1277

Kreps D. M. and J. A. Scheinkman (1983), "Quantity pre-commitment and Bertrand competition yields Cournot outcomes", *Bell Journal of Economics*, vol. 14, pp. 326 – 337

Lucas, N. and P. Taylor (1993), "Characterising generators behaviour: bidding strategies in the pool. A game theory analysis", *Utilities Policy*, April, pp. 129 – 135

Lucas, N. and P. Taylor (1994), "Structural deficiencies in electricity pricing in the pool", *Utilities Policy*, January, pp. 72 – 76

Newbery, D. M. (1992), "*Capacity-constrained supply function equilibria: competition and entry in the electricity spot market*", Department of Applied Economics, University of Cambridge, Mimeo, June

Newbery, D. M. and R. Green (1992), "Competition in the British electricity spot market", *Journal of Political Economy*, vol. 100, pp. 929 – 953

Offer (1991), "*Report on pool price inquiry*", December

Offer (1992), "*Review of pool prices*", December

Offer (1998), "*Review of electricity trading arrangements: proposals*", July

Posner, M. V. (1993), "Electricity and the primary fuels: technology, market structure and prices", *National Institute Economic Review*, August, pp. 64 – 86

Powell, A. (1992), "*Trading forward in an imperfect market: the case of electricity in Great Britain*", Queen Mary and Westfield College, Paper no. 261, May

Robson, A. J. (1981), "Implicit oligopolist collusion is destroyed by uncertainty", *Economics Letters*, vol. 7, pp. 144 – 148

Ruff, L. (1999), "*New electricity trading arrangements for England and Wales: a comment by NERA on OFGEM's proposals*", NERA Energy Regulation Brief, August

Tirole, J. (1988), *The theory of industrial organisation*, MIT Press, Cambridge, Massachusetts

Turnbull, S. J. (1983), "Choosing duopoly solutions by consistent conjectures and by uncertainty", *Economics Letters*, vol. 13, pp. 253 – 258

Vickers, J. and G. Yarrow (1991), "The British electricity experiment", *Economic Policy*, vol. 12, pp. 187 – 232

Von-der-Fehr, N. H. M. and D. Harbord (1992), "*Spot market competition in the UK electricity market*", Department of Economics, University of Oslo, Memorandum no. 9

Wolfram, C. (1999), "Measuring duopoly power in the British electricity spot market", *American Economic Review*, vol. 89, pp. 805 – 826

Chapter 3

Environmental taxation and trade: an asymmetric equilibrium

3.1 Introduction

Trade between countries is being liberalised. This is normally done through trade agreements such as GATT and NAFTA that abolish tariffs on imports and prohibit the introduction of further tariffs. Countries, however, can still affect trade because they retain the ability to tax local firms in respect of the damage that they may cause to the local environment.[1]

There are two concerns about using environmental tax policy in a context of liberalised international trade. First, governments may find it advantageous to use environmental policies to give local firms a competitive edge against its foreign competitors by setting lower environmental taxes – rent-seeking behaviour. If so, environmental taxes may be below first-best taxes (marginal damage), resulting in emissions being greater than what they will be in a first-best tax equilibrium. (This is because polluters do not pay the full cost imposed on society by their actions so that they have an incentive to pollute more from society's point of view.) Second, firms can locate in response to governments' decisions on environmental taxation where capital moves freely. Thus, environmental policy could affect industry structure. Setting up production facilities in a country, however, is not cost-free: firms will incur costs – sunk costs – which will not be recovered if they relocate production to another country. Thus, governments could also use environmental policy to attract firms to locate in their country.[2]

[1] Ideally, firms will be taxed according to the marginal damage that they impose on the environment. This level of tax is typically referred to as the first-best tax. In addition, governments also have the ability to tax firms in respect of pollution that crosses national borders and affect the global environment (global warming), other country's environment (acid rain) or both. Modelling pollution that crosses national boundaries raises different issues. It will be mentioned as a possible extension of the model in Section 3.7.

[2] It is worth noting at the outset that market structure might also change in response to environmental policies through other mechanisms such as relocation, shifting production from a plant at one location to plants in other locations and exiting the industry without starting up at another location.

The combined effect of these two aspects could be a race to the bottom in environmental taxes. I use this term to refer to levels of environmental taxes that are below the level of a first-best tax (marginal damage).[3] So, for example, I will not refer to a race to the bottom where trade liberalisation results in lower environmental taxes that remain above the first-best tax.

The perception that a race to the bottom in environmental taxes is likely to follow trade liberalisation may result in one or both of the following policy recommendations. First, countries should harmonise their environmental policies. Second, countries that impose tighter environmental regulation should be allowed to impose tariffs on imports from countries with more lax environmental policies (see, for example, Daly (1993)).

There is a voluminous literature on the relationship between trade and environment. For example, a recent survey, Sturm (2002), groups the literature around seven main issues: the properties of optimal environmental policies in open economies with local pollution; the properties of such policies when pollution extends across national borders; the consequences of trade liberalisation; the political economy of the trade and environment nexus; the effect of trade on (renewable and non-renewable) natural resources; the implications of trade in hazardous waste; and the location of firms and production. The volume of this literature can be easily gauged from that survey. It focuses on the first four issues and it includes about 120 references, which are presented as not being an exhaustive list of references.

This chapter focuses on a particular aspect: the consequences of trade liberalisation for environmental taxation – the third issue in Sturm's survey – using the framework of strategic environmental policy, also surveyed in Ulph (1997). The latter is an area that developed in the 1990s inspired by work on strategic trade policy, e.g. Brander and Spencer (1985). More specifically, the purpose of the chapter is to explore the effect of trade liberalisation on environmental taxation and industry entry when both firms and governments act strategically. As noted in Wilson (1996), the issue is not whether a low

---

[3] As Wilson (1996) says, a more accurate term will be "race towards the bottom". Another term used in the literature to refer to this outcome is "ecological dumping", see Section 2 of Rauscher (1994) for a discussion. The parallel here is dumping in commodity markets, which is associated with prices below marginal costs.

environmental tax will attract more entry. Jeppesen et al. (2002) survey a wide range of empirical studies of the relationship between new firm location and environmental regulation. They find some regularities such as studies undertaken in recent years finding that capital flows respond to heterogeneous environmental regulation. The issue is whether entry will not harm the country because firms pay an effective compensation (environmental tax) for the damages to the environment that they cause. Wilson (1996) also surveys the theoretical literature on this potential race. He concludes that the existence of a race to the bottom is very dependent on the assumptions about the type of market imperfections and about the taxes and subsidies that are available.

In this chapter, I consider a specific case where pollution is local and there are two different countries both consuming and producing. For this purpose I extend a model, Ulph (1995), to explore the effect that asymmetries between countries and an alternative form of government tax setting behaviour has on environmental taxation and entry. The range of possible asymmetries between countries that one could consider is, however, enormous. It includes technology, set-up costs, labour costs, environmental damage, size, and others. Here asymmetries are restricted to labour costs, damage to the environment and size of the countries (population). These asymmetries complicate the calculations and limit the extent to which inferences about the extent of entry can be made based on analytical results. Thus, I supplement these with some simple simulations of the equilibrium results.

This chapter is structured as follows. Section 3.2 puts this work in the context of the existing literature. Section 3.3 discusses the assumptions of the model. Section 3.4 analyses tax setting behaviour in an asymmetric context. Section 3.5 and 3.6 characterise the extent of new entry. Section 3.7 considers some possible extensions of the model. The last section summarises and concludes.

3.2 Relevant literature

As I mentioned in the introduction, there is a voluminous literature on the relationship between trade and the environment. The purpose of this section is to relate this chapter's contribution to the existing literature. I focus in a situation where there is local pollution and the government has no other policy instrument such as trade policy or emissions standards to control this externality.[4] For this purpose, we start by discussing the relationship between trade and environmental taxation when markets are competitive. We then discuss the assumptions that are typically made in the literature about the main features of the model used in this chapter.

Ulph (1997) and Sturm (2002) summarise the results of earlier work in the literature when product markets are competitive. In that case, the benchmark is a small country with only a pollution distortion where the best the government can do is to set an environmental tax that equals marginal damage. The situation is different for a large economy where its environmental policy will affect the world price of the relevant good. There is, however, no generalised case for a race to the bottom in environmental taxes. For example, it has been shown that the exporter of a pollution intensive good may want to impose an environmental tax that is above marginal damage, Rauscher (1994) and Krutilla (1991).

In recent years, a strand of the literature that explores environmental policy when product markets are not competitive and governments act strategically – strategic environmental policy – has developed. This literature has been inspired by work on strategic trade policy such as Brander and Spencer (1985). The motivation of that literature is that a country will find it advantageous to capture a large share of an international market characterised by oligopolistic competition. If so, export subsidies can be used to shift profits home.[5]

[4] Section 3.7 explores the issue of pollution that extends beyond national boundaries. A separate strand of the literature considers the choice of an environmental policy instrument. Typically, this is a choice between an emission standard and a tax. Ulph (1997) suggests that the results are ambiguous. For example, consider the extreme case where there is no possibility of abatement. In this case, a government using standards is effectively committing firms to a level of output. As the output is fixed, there is no incentive for the rival government to engage in strategic behaviour. In this case, the dominant strategy for the other government is to use standards. (If both governments set taxes, there will be strategic behaviour. And, if one government uses standards and the other a tax, the one using standards will set them consistent with becoming a Stackelberg leader.) However, this conclusion is not robust. It can be shown that where abatement is possible and marginal abatement costs are linear, the dominant strategy for the other government is to use an environmental tax.

[5] As noted in Barrett (1994), there is a difference between strategic trade policy and strategic environmental policy. An environmental tax below marginal damage offers an implicit subsidy that is costly to society because the subsidy worsens domestic pollution. In the case of models of strategic trade policy, the export subsidies that are the subject of strategic behaviour are not costly to society. They simply transfer resources from the domestic government to domestic firms.

One of the issues discussed in strategic environmental policy is whether in equilibrium environmental policy is "tough" or "weak". In the case of an environmental tax, tough (weak) means a tax above (below) the marginal damage to the environment. The basic model used to explore strategic environmental policy assumes that there are two (or various) producing countries selling in another country (or international market); industry structure is fixed; typically, there is one firm in each country; firms set outputs in a Cournot fashion; and governments can only use one policy instrument, Ulph (1997). The main conclusion of this literature is that when pollution is local and governments set an environmental tax, environmental policy is weak. This provides support to environmentalists' concerns about a race to the bottom. However, this result is dependent upon the model's assumptions. For example, Barrett (1994) shows that if firms compete in prices, environmental policy will be tough.

We review below some of the literature relevant to the specific features of the model used in this chapter – local consumption and product market integration, government's commitment, free movement of capital, asymmetries between countries and alternative approach to government's environmental policy.

There is consumption of the relevant good in both countries. This is also a feature of models of strategic trade policy, e.g. Brander and Krugman (1983) but not of all of them, e.g. Brander and Spencer (1985). And as suggested above, it is not a feature of the standard model used to explore strategic environmental policy. One rationale for this assumption is that all the output is exported. As noted in Neary and Leahy (2000), another rationale for this assumption is that "some output is domestically consumed, that home and foreign markets are integrated and that domestic consumer surplus does not enter the social welfare function". Broadly speaking, if there is local consumption and the consumers' surplus enters the social welfare function, a government's incentive to engage in profit shifting is reduced. Ulph (1997) suggests that if the government has one policy instrument to address both local pollution and imperfect competition, this may lead the government to choose an environmental tax below the level of the first-best tax.

Product markets are integrated after the abolition of trade barriers. This is in contrast with models of strategic trade policy such as Brander and Krugman (1983) where markets remain segmented after the abolition of trade barriers. This means that each firm perceives each country as a separate market and makes distinct decisions for each one. The main effect of assuming market integration is that there is an increase in the number of firms with the consequent pressure on prices and hence profits. This reduces the extent to which governments may be able to use environmental tax to shift profits from foreign firms.

When governments (or firms) act strategically, one of the issues is the extent to which they can commit to a course of action as part of their strategy.[6] Typically, models of strategic trade policy assume that the government commits itself to a level of subsidy before firms choose their outputs, e.g. Brander and Spencer (1985). The assumption about a government's commitment matters. Eaton and Grossman (1986) show that when all the assumptions from Brander and Spencer are maintained and firms are assumed to compete on prices rather than quantities, it is optimal to tax exports. However, Carmichael (1987) then shows that if the government cannot pre-commit to the level of subsidy and firms set prices before government sets the subsidy, it is optimal to set an export subsidy.[7] A similar issue arises in the case of environmental policy where a government's limited ability to commit to a course of action may be relevant to firm's investment in abatement technology and industry entry.[8] Petrakis and Xepapadeas (1999) explore the effect on abatement effort of a government's inability to commit to an environmental tax. The main result is that firms will spend more on abatement technology because the government's inability to commit to an environmental tax. Thus, the environmental tax will be lower than the first-best tax. Welfare is also lower because the higher consumer surplus (from increased output) and the lower level of pollution are not sufficient to compensate for the increased expenditure in abatement. Ulph (1995)

[6] More generally, commitment is usually modelled by making actions sequential rather than simultaneous. For example, in the case of Stackelberg competition, the "leader" commits itself to a level of output by choosing it ahead of its rivals whereas in the standard Cournot model all firms choose their outputs simultaneously. Commitment is also used in other areas of industrial organisation, for example, characterising business strategy towards new entry (the "animal spirits" taxonomy, summarised in Chapter 8 of Tirole (1988)). Spencer and Brander (1992) provide a more general treatment of the role of commitment. The paper explores the trade-off between the strategic value of commitment and the value of retaining flexibility. They show that if uncertainty is sufficiently large firms might prefer to forego the value of pre-commitment and wait for the uncertainty to be resolved.

[7] Carmichael (1991) provides an empirical analysis of the behaviour of the US export credit bank, which suggests that the sequence of actions of firms and the government might matter.

[8] In this case, a reasonable sequence of decisions will be as follows: firms pre-commit (invest in abatement technology or enter the market), the government sets the environmental tax and firms make their output decisions.

also investigates the effect of commitment on the level of the environmental tax and industry entry where capital moves freely between countries and there is no trade. The analysis shows that a government's inability to commit to a level of tax increases the equilibrium level of tax above the first-best tax.

One of the main features of discussions about globalisation is the implication of the free movement of capital. As a result, firms can locate in what may appear to them as the most convenient place. Markusen et al. (1993) present a simple model with two countries, segregated markets, local pollution and imperfect competition where firms choose the number and location of their plants. They explore how a given industry structure may change when there are no environmental taxes and one country introduces a unilateral tax. They also illustrate the implication of setting environmental taxes assuming that the industry structure will not change. They show that once the industry structure changes, the resulting social welfare could be below the level that prevailed before the introduction of the environmental tax. Markusen et al. (1995) extend the previous paper to consider the case where two governments compete in terms of environmental policy to attract a firm with increasing returns to scale at the plant level. The firm has to decide whether to maintain plants in both regions, one region or shut down. They show that if the marginal damage to the environment is not large, governments will compete by undercutting each other's pollution taxes though this will not necessarily result in a race to the bottom as defined in Section 3.1. Another paper that explores issues associated with market structure is Ulph (1994). It extends Markusen et al. (1993 and 1995) to consider the effect of government's commitment to a tax rebate (before the environmental tax has been set) to prevent relocation to other countries.[9] The analysis suggests that reducing firms' ability to adjust to environmental policy may be undesirable. Relocating can be the means of gaining market share and a policy of rebates will limit that. Furthermore, it may encourage inward investment, which harms domestic producers and simply transfers tax revenues abroad.

Symmetry is not a representative characteristic of the world. Reality is characterised by asymmetries between countries and between firms. These issues have been analysed in the context of the strategic trade literature. For example, Neary (1994) considers an

[9] This, however, may be seen as an example of a partial commitment to a level of tax.

international duopoly to explore which sector of the economy should be supported. He shows that strategic trade policy should support those sectors in which the home country has a cost advantage over foreign firms, i.e. a comparative advantage in profit shifting. Leahy and Montagna (2001) characterise optimal trade policy where there are cost differences between domestic firms within a sector. They show that the effect of asymmetry between firms may depend on the government's cost of funds. For example, if the cost of funds is sufficiently low, the low cost firm will receive higher export subsidies. Similar issues arise in the context of strategic environmental policy where asymmetries are the main basis for environmentalists' concerns. Furthermore, understanding the effect of asymmetries on environmental policies is also important given the perception that a race to the bottom in environmental tax implies one country setting lower environmental taxes. Katsoulacos et al. (1996) explore the effect of asymmetries between countries in the standard setting for strategic environmental policy described earlier – n firms located in n countries serving an international market and no local consumption. The paper shows that in a non-cooperative equilibrium the two countries adopt environmental taxes that are below the level of the first-best tax. The effect of asymmetries is that the country with lower firm's output will follow a tougher environmental policy than the other country. It is also shown that under the model's assumptions, this will be the country with higher unit labour costs or higher damage to the environment. Conrad (1997) obtains similar results with a model of two firms located in different countries that trade between them when the product markets remain segregated.

Finally, governments do not necessarily set environmental policy assuming that their actions do not affect the neighbouring countries (non cooperative tax setting behaviour). It will therefore be useful to explore the effect of an alternative assumption about governments' tax setting behaviour in the context of asymmetries described above. A government may take into account the effect that environmental policy could have on its neighbouring country. One could therefore envisage a situation where governments harmonise the tax setting *process* and set environmental taxes cooperatively. (This is different to harmonising environmental taxes.) Katsoulacos et al. (1996) consider the possibility that government set environmental taxes cooperatively when countries are asymmetric. The paper shows that the environmental tax will be above the first-best tax when countries set taxes cooperatively. However, Hoel (1997) shows that when industry is endogenous, it is not obvious that a government environmental policy will be more or less strict under non-cooperation than under cooperation.

3.3 The model

As indicated earlier, the analysis in this chapter is based upon an extension of the model in Ulph (1995). This section describes the assumptions in Ulph (1995), the results obtained in that paper for a symmetric equilibrium and the additional assumptions necessary to extend the paper to consider asymmetry between countries and an alternative form of government tax setting behaviour.

3.3.1 Overview

Widgets are produced and consumed in two countries and this represents a small proportion of their output so that a partial equilibrium analysis is an adequate representation of the reality. Production of widgets involves the emission of local pollution. The policy instrument available to the government to control pollution is an environmental tax on output.

Both governments and firms behave strategically at two levels. Each government behaves strategically towards firms by committing to an environmental tax before firms make their output decisions. A government also behaves strategically towards the other government and it assumes that its decision about an environmental tax will not affect the other government's decision about its own environmental tax. Firms behave strategically in the entry game where they commit themselves to produce in a country before the government sets its environmental tax. They also behave strategically in the output market where they assume that their output decisions will not affect their rivals' decisions (Cournot's behaviour). This is therefore a game with the following sequence of moves: firms decide to enter (stage 0), governments set environmental taxes (stage 1), firms decide whether to be operative (stage 2), and finally, firms set their outputs in a Cournot fashion (stage 3).

3.3.2 Consumers

Consumers in both countries have identical tastes. Their preferences can be represented by the following utility function

$$u(x) = ax - \frac{x^2}{2} + z \qquad (1)$$

where x is the number of widgets consumed by a consumer and z is the consumer's expenditure on other goods. An individual's demand for widgets is then

$$x = a - p \qquad (2)$$

where p is the price of a widget.

Consider an autarky equilibrium where the two countries are separated by an infinite trade barrier. The relevant demand for widgets in each country is independent of the demand in the other country. So for a country with m_i consumers, the demand is the aggregate inverse-demand function where each of the consumers is represented by equation 2.[10] Multiplying both sides by m_i and re-arranging I get,

$$p = a - \frac{X_i}{m_i} \qquad (3)$$

where $X_i = m_i x$. Given an equilibrium price, the consumers' surplus in this market will be

$$\frac{X_i^2}{2m_i} \qquad (4)$$

Consider now the abolition of trade barriers. I assume that there are no transport costs and that product markets are integrated. This means that all firms have the same ability to serve consumers regardless of their location and that all consumers face identical prices. An individual's consumption will be therefore the same in both countries as per equation 2 above. Total consumption in each country, however, will differ if the number of consumers is different. The aggregate inverse demand function is derived by multiplying both sides of an individual's demand function by $m_1 + m_2$. After re-arranging, I obtain

$$p = a - \frac{X_1 + X_2}{m_1 + m_2} \qquad (5)$$

Given an equilibrium price, total consumers' surplus in the integrated market will be now

[10] This is the standard assumption in the literature. It ignores the distribution effects of trade liberalisation and environmental taxation.

$$\frac{(X_1 + X_2)^2}{2(m_1 + m_2)} \tag{6}$$

3.3.3 Firms

Foreign residents own local firms. However, they are dispersed in third countries so that no one has any significant power to affect trade between these countries. There is free movement of capital between the two countries. Entry is only subject to an irrecoverable set up cost, $S > 0$, which firms must incur before they start producing. Thereafter, they also incur fixed costs in each period, $F > 0$, in which they are operative. Technology is owned by foreign residents so that it is the same in both countries. Thus, the set up cost and the fixed cost will be the same in both countries.

Here I will assume that fixed costs are zero to focus on new entry. Positive fixed costs enhance the effect of integration by allowing re-distributing production between the two countries. For example, firms that enter the market in one country before the abolition of trade barriers may shut down their plants if profits do not exceed fixed costs and this may increase the opportunities for entry in the other country. (See Markusen et al. (1993, 1995) for an analysis of plant relocation.)

Unit labour costs in country i are constant, c_i, but may differ between countries. This may be the result of differences in wages or in labour productivity. Production involves the emission of local pollution. The government sets a tax, t_i, per unit of output to control emissions. Thus, the total unit cost of production is $c_i + t_i$.

3.3.4 Government

The government's objective function is the maximisation of society's welfare. The policy instrument available to the government to achieve this objective is an environmental tax.

The assumptions made above determine the form of a government's objective function. Thus, it excludes firms' profits since these revert to the owners of the capital who are residents of third countries. Furthermore, the proceeds of environmental taxes will not be a transfer of resources within society and without welfare implications. They represent the part of firms' profits that stays in the country thereby increasing society's welfare and should be therefore included in the welfare function.

The other elements of society's welfare are the consumers' surplus, which increases the country's welfare, and the damage to the environment, which reduces it. I assume that environmental damage is also a source of asymmetries between countries because it reflects the ability of emissions to generate damage to the local environment and individuals' willingness to pay for a cleaner environment, which are idiosyncratic.[11]

Consider first the case where countries are separated by an infinite trade barrier – autarky equilibrium. In this case, production equals consumption, and both will be denoted by X_i. I can therefore write society's welfare function for country i as

$$W_i = \frac{X_i^2}{2m_i} + t_i X_i - \frac{d_i}{2} X_i^2 \tag{7}$$

where d_i represents the damage (per unit of output) made to the local environment and X_i represents the amount produced and consumed.

Consider now society's welfare function when trade barriers are abolished and product markets integrate. There is now a distinction between a country's total consumption, X_i, and a country's total production, Y_i. The relevant demand function for widgets is that in equation 5 above and the consumers' surplus is that in equation 6. However, a government will be interested in share of consumers' surplus that accrues to consumers in its country and not about the total. For country 1, this is

$$\frac{m_1}{(m_1 + m_2)} \frac{(X_1 + X_2)^2}{2(m_1 + m_2)} \tag{8}$$

In the post integration equilibrium, consumers in both countries will face the same prices for widgets and will therefore consume the same amount. Thus, it can be shown that the consumers' surplus of residents in country 1 from equation 8 will be identical to the consumers' surplus from equation 4. The consumers' surplus in equation 8 conveys, however, the concept of equilibrium in an integrated market. The other elements of the

[11] An alternative is to assume that damage to the environment depends exclusively on technology, which is identical in both countries, so that damage to the environment is the same in both countries.

government's objective function are not affected by integration. Thus, society's welfare function post integration is

$$W_i = \frac{m_i(X_1 + X_2)^2}{2(m_1 + m_2)^2} + t_i Y_i - \frac{d_i}{2} Y_i^2 \tag{9}$$

3.3.5 The symmetric equilibrium

An equilibrium is obtained by solving the model by backwards induction; first output and prices are determined in the market stage, then the environmental tax is set and finally the number of firms is determined. Ulph (1995) uses the objective function in equation 7 to explore the extent of new entry after the removal of trade barriers when governments and firms behave strategically and the two countries are identical – symmetric equilibrium. Two equilibria are determined: an autarky equilibrium and a post integration equilibrium, i.e. the four-stage game is played twice. In the autarky equilibrium, the environmental tax and the industry structure (number of firms) are determined. The latter is determined by assuming that capital moves freely between countries and that there is free entry into the industry (subject to a sunk cost). The game is then played when trade barriers are abolished and a new equilibrium is determined: the relevant industry structure is that from the autarky equilibrium and a new environmental tax is determined. Additional firms will enter if profits to existing firms exceed sunk costs.

In the autarky equilibrium, product demand and government decision making are independent of the other country. A government's behaviour is characterised by the following first-order condition

$$(t_1 - d_1 X_1)mN = X_1 \tag{10}$$

The marginal damage, d_1X_1, is the first-best tax. The difference between the actual tax, t_1, and the marginal damage in the left-hand side of equation 10 represents the distortion introduced by the strategic behaviour of firms and governments. The equilibrium level of tax is

$$t_1 = d_1 X_1 \left(1 + \frac{1}{d_1 m_1 N_1}\right) \tag{11}$$

and is above the first-best tax.

The number of firms is determined with reference to the assumption of free movement of capital and the requirement that profits will be sufficient to cover sunk costs, S. The resulting number of firms will then be

$$N_1^a = \frac{(a-c_1)\sqrt{\frac{m_1}{S}} - 2}{1 + d_1 m_1} \geq 1 \tag{12}$$

This gives the following condition for the existence of the autarky equilibrium

$$(a-c_1)\sqrt{m_1 / S} > 3 + d_1 m_1 \tag{13}$$

Consider now the abolition of trade barriers. If product markets are integrated, there are two main direct changes affecting firms. First, the number of firms competing in the market increases. This means that, other things being equal, profits will be smaller. Second, total demand increases. This means that, other things being equal, profits will increase. The extent of additional entry in the post integration equilibrium will depend on the balance between these two effects. For example, if the resulting number of firms in the autarky equilibrium is large then there may be little entry after the abolition of trade barriers. The discussion here assumes that this is not necessarily the case. Ulph (1995) explores the case where each government sets environmental taxes to maximise society's welfare taking as given the other government's behaviour (non-cooperative equilibrium). The results of the market stage are summarised in Annex 3.1. The tax equilibrium is now characterised by two first-order conditions. For country 1, this is

$$(t_1 - d_1 Y_1)\frac{(m_1 + m_2)N_1^a(N_2^a + 1)}{(N_1^a + N_2^a + 1)} = Y_1 - X_1 \frac{N_1^a}{(N_1^a + N_2^a + 1)} \tag{14}$$

In a symmetric equilibrium, population is the same in both countries ($m_1 = m_2$), damage to the environment is the same ($d_1 = d_2$), and labour costs are the same ($c_1 = c_2$). If so, the autarky equilibrium in the two countries will be identical resulting in the same market structure, $N_1^a = N_2^a$. Furthermore, the post integration equilibrium will be characterised by no trade because production equals consumption in each country and between countries.

The equilibrium level of the environmental tax after integration is then

$$t = dY\left(1 + \frac{1}{2dmN^a}\right) \tag{15}$$

Comparing this with the environmental tax in the autarky equilibrium, equation 11, suggests that, other things being equal, abolishing trade restrictions reduces the environmental tax. However, it remains above the level of the first-best tax. Thus, there is no evidence of a race to the bottom in environmental taxes in a symmetric equilibrium.

There is an infinite supply of capital so that entry of new firms in the post integration equilibrium is based upon a comparison of profits against sunk costs, given the existing number of firms in the autarky equilibrium. Substituting the equilibrium tax, equation 15, and the number of firms in the autarky equilibrium, equation 12, into the equilibrium profit function (see Annex 3.1), results in the following entry condition

$$(a-c)\sqrt{m/S} < 2+\sqrt{2} \tag{16}$$

Conditions 13 and 16 together require that the condition for entry post integration in a symmetric equilibrium is

$$3+d\ m < (a-c)\sqrt{m/S} < 2+\sqrt{2} \tag{17}$$

This is a very stringent condition and suggests that new entry post integration is unlikely to occur. This suggests that of the two direct effects on profits, the increase in the number of firms is likely to dominate so that industry structure is determined by the industry structure in the autarky equilibrium. The next sections focus on tax setting and on the entry decision and characterise the asymmetric equilibrium under different assumptions about governments' behaviour after the abolition of trade barriers.

3.4 The tax equilibrium

This section characterises the tax equilibrium (stage 3) after trade barriers are abolished, when countries are different and under alternative assumptions about governments tax setting behaviour. The analysis of firms' output decisions in Ulph (1995) is unaffected and the results are summarised in Annex 3.1.

3.4.1 Non-cooperative equilibrium

In an integrated market, social welfare is represented by equation 9. It depends on the total level of consumption and output in both countries, which in turn depends on the environmental tax in each country (see Annex 3.1). In a non-cooperative equilibrium, each government ignores the effect that its tax decision has on the other country's welfare and output. Thus, each government sets its own environmental tax taking the other country's tax as given. Welfare is therefore maximised by setting the environmental tax so that the partial-total derivative[12] of the country's welfare function with respect to its own taxes, the first-order condition, is zero. For country 1, this is

$$\left.\frac{dW_1}{dt_1}\right|_{dt_2=0} = \frac{\partial W_1}{\partial(X_1+X_2)}\frac{\partial(X_1+X_2)}{\partial t_1} + \frac{\partial W_1}{\partial Y_1}\frac{\partial Y_1}{\partial t_1} + \frac{\partial W_1}{\partial t_1} = 0 \tag{18}$$

I treat total consumption, $X_1 + X_2$, as one variable because the markets are integrated and it is the total consumption that dictates the equilibrium price in the market, equation 5. An equilibrium is a pair t_1 and t_2 that solves this equation and a similar one for country 2. Using the welfare function in equation 9, the first-order condition above can be written

$$\frac{m_1(X_1+X_2)}{(m_1+m_2)^2}\frac{\partial(X_1+X_2)}{\partial t_1} + (t_1 - d_1Y_1)\frac{\partial Y_1}{\partial t_1} + Y_1 = 0 \tag{19}$$

The market game (stage 3) where outputs are set, is unaffected by the change in the government's objective function. The remaining derivatives in the above expression can be derived from the equilibrium results of stage 3 in Ulph (1995), see Annex 3.1. Substituting these into the first-order condition and re-arranging, I obtain

$$-\frac{m_1(X_1+X_2)}{(m_1+m_2)}\frac{N_1}{(N_1+N_2+1)} - (t_1 - d_1Y_1)\frac{(m_1+m_2)N_1(N_2+1)}{(N_1+N_2+1)} + Y_1 = 0 \tag{20}$$

The first term in the above equation can be simplified because in equilibrium individual's consumption will be the same in both countries. I can then re-arrange the above as

$$(t_1 - d_1Y_1)\frac{(m_1+m_2)N_1(N_2+1)}{(N_1+N_2+1)} = Y_1 - X_1\frac{N_1}{(N_1+N_2+1)} \tag{21}$$

This is the first-order condition in equation 14 above, which was obtained using the alternative government's objective function defined in equation 7. This suggests that modelling the government's tax setting behaviour using a more general form of the consumer's surplus as in equation 9 above does not affect the equilibrium results. This is because, in equilibrium, individual's consumption (and hence consumer's surplus) is the same in both countries. Re-arranging this expression, I obtain

[12] I use this terminology to distinguish between the first-order condition, $\left.\frac{dW_1}{dt_1}\right|_{dt_2=0}$ and one of its elements, $\frac{\partial W_1}{\partial t_1}$.

$$(t_1 - d_1 Y_1) = \frac{1}{(m_1 + m_2)} \left(\frac{Y_1}{N_1} + \frac{(Y_1 - X_1)}{(1 + N_2)} \right) \tag{22}$$

The right hand side suggests that there are two sources of externalities that create a wedge between taxes and marginal damage (or explain the toughness of environmental policy): the total level of production and the trade surplus. (A trade deficit reduces the wedge between taxes and marginal damage – more on this below.) Other things being equal, the level of production is larger when the intensity of competition is stronger, so I refer to the first source as a competition externality. The trade surplus (or deficit) results from international trade so I refer to the second one as the trade externality. The competition externality is inversely proportional to the size of the industry in its own country whereas the trade externality is inversely proportional to the size of the industry in the other country.

Equation 22 also suggests that environmental policy will be different in the two countries. In the exporting country, where production exceeds consumption, the environmental tax will be above the first-best tax. In the importing country, the environmental tax will be above the first-best tax if the average output per firm is greater than the average import from firms in the other country. Given that in both countries there was an industry in the autarky equilibrium, it is likely that the amount imported in the post integration equilibrium will not exceed the amount produced locally. If so, the environmental tax will not exceed the first-best tax in the importing country. This suggests the following results.

Result 1: environmental tax will be above first-best tax in the exporting country and most probably in the importing country.

Result 2: environmental policy in the exporting country will be tougher than in the importing country.

Result 3: other things being equal, environmental policy will be tougher in the country with the largest output per firm.

Result 4: other things being equal, the larger the market that results post integration the more lax the environmental policy in both countries.

Equation 22 also helps us understand the symmetric equilibrium further. One of the results of the assumption that countries are identical is that although trade barriers are abolished, the equilibrium is characterised by no trade. In the context of equation 22, it suggests that the trade externality disappears and that environmental policy will be determined by the competition externality. This also helps to explain the similarity between the form of the equilibrium tax in the autarky equilibrium (equation 11) and in the post integration equilibrium when the two countries are identical (equation 15). In both cases, there is only the competition externality that results from firms' non-cooperative behaviour because there is no trade – either because of trade barriers or because the countries are identical.

3.4.2 Cooperative equilibrium

After the abolition of trade barriers, each government assumes that its tax decision affects its neighbouring country's consumption and welfare. The two governments coordinate therefore their environmental policies – *process* harmonisation – and maximise the sum of their welfare, $W_1 + W_2$, as defined in equation 9 above. Governments' cooperative tax setting behaviour is best seen as part of a wider scheme between governments such as the European Union where offsetting payments will be taken care of separately and they can be excluded from the model. If so, the equilibrium will be a pair of taxes t_1 and t_2 that solves the following two first-order conditions

$$\left.\frac{dW_1}{dt_1}\right|_{dt_2=0} + \left.\frac{dW_2}{dt_1}\right|_{dt_2=0} = 0 \tag{23a}$$

$$\left.\frac{dW_1}{dt_2}\right|_{dt_1=0} + \left.\frac{dW_2}{dt_2}\right|_{dt_1=0} = 0 \tag{23b}$$

This is not the only possible approach to model a cooperative equilibrium. A more general modelling approach will be a bargaining equilibrium where maximisation of joint welfare and tax setting (the "split" of the welfare between the countries) are modelled separately. Conrad (1997) suggests in a footnote an alternative form of a cooperative equilibrium where participation constraints are added. This means that in a cooperative equilibrium, the welfare of each country must be higher than in the non-cooperative equilibrium.

Using the social welfare function in equation 9, the first-order condition in equation 23a can be written as

$$\frac{\partial W_1}{\partial(X_1+X_2)}\frac{\partial(X_1+X_2)}{\partial t_1}+\frac{\partial W_1}{\partial Y_1}\frac{\partial Y_1}{\partial t_1}+\frac{\partial W_1}{\partial t_1}+\frac{\partial W_2}{\partial(X_1+X_2)}\frac{\partial(X_1+X_2)}{\partial t_1}+\frac{\partial W_2}{\partial Y_2}\frac{\partial Y_2}{\partial t_1}=0 \tag{24}$$

Re-arranging yields

$$\frac{X_1+X_2}{m_1+m_2}\frac{\partial(X_1+X_2)}{\partial t_1}+(t_1-d_1Y_1)\frac{\partial Y_1}{\partial t_1}+(t_2-d_2Y_2)\frac{\partial Y_2}{\partial t_1}+Y_1=0 \tag{25}$$

The market game (stage 3) remains unaffected by the assumption about the government's tax setting behaviour and the derivatives in equation 25 can be obtained from Annex 3.1. Substituting these in the above expression yields

$$-\frac{(X_1+X_2)N_1}{(N_1+N_2+1)}-(t_1-d_1Y_1)\frac{(m_1+m_2)N_1(N_2+1)}{(N_1+N_2+1)}+(t_2-d_2Y_2)\frac{(m_1+m_2)N_1N_2}{(N_1+N_2+1)}+Y_1=0 \tag{26}$$

In any equilibrium, total consumption in the two countries equals total output, that is $X_1+X_2=Y_1+Y_2$. Substituting this and re-arranging gives

$$(t_1-d_1Y_1)N_1(N_2+1)-(t_2-d_2Y_2)N_1N_2=\frac{-Y_2N_1+Y_1(N_2+1)}{(m_1+m_2)} \tag{27}$$

A similar expression is also obtained from (23b)

$$(t_2-d_2Y_2)N_2(N_1+1)-(t_1-d_1Y_1)N_1N_2=\frac{-Y_1N_2+Y_2(N_1+1)}{(m_1+m_2)} \tag{28}$$

I want to characterise the environmental policy in each country in terms of the wedge between the actual tax and the first-best tax (marginal damage) – as in the previous section. Thus, I solve equations 27 and 28 for $t_i-d_iY_i$. For country 1, this yields

$$(t_1-d_1Y_1)=\frac{Y_1}{(m_1+m_2)N_1} \tag{29}$$

The right hand side of this expression is positive so the environmental tax will be above the first-best tax. Equation 29 also suggests the following about the wedge between tax and marginal damage. First, a cooperative tax setting behaviour eliminates the trade externality identified in the non-cooperative equilibrium, equation 22. So the effect of a cooperative tax setting behaviour is similar to the effect of a tariff that eliminates the trade externality that arises when governments set environmental taxes non-cooperatively. Second, the competition externality itself is not affected by the assumption about the government's tax setting behaviour because it results from firms' non-cooperative output setting. So Results 3 and 4 derived for the non-cooperative equilibrium also apply here. Finally, it is worth noting that the characterisation of the environmental policy is similar to that resulting when countries are identical and governments set taxes non-cooperatively, see equation 15. The intuition is simple: in both cases, there is no trade externality. In the case of cooperative tax setting behaviour, the trade externality has been internalised and when countries are identical the equilibrium is characterised by no trade and so there is no trade externality.

Result 5: when governments set environmental policies cooperatively, the environmental tax will be above first-best tax in both countries.

Result 6: cooperation between governments eliminates the trade externality identified in the non-cooperative equilibrium and can be regarded as a substitute for a trade policy.

Result 7: other things being equal, environmental policy in the exporting (importing) country will be more relaxed (tougher) than in the non-cooperative equilibrium.

3.5 Entry in an asymmetric equilibrium

Ulph (1995) shows that entry is unlikely following the abolition of trade barriers when governments set taxes non-cooperatively and the two countries are identical. I want to test the robustness of this result by considering a situation where countries are not identical and then by considering a situation where countries set taxes cooperatively. I have therefore two hypotheses about new entry that I want to verify. First, I want to verify whether entry conditions in an asymmetric equilibrium will be more relaxed than what the symmetric equilibrium may suggest. Second, I want to verify whether entry will be more likely when governments set taxes cooperatively.

3.5.1 Non-cooperative equilibrium

Capital is mobile so that there will be new entry after integration if profits exceed sunk costs. The relevant number of firms is the number of firms determined in the autarky equilibrium on the assumption of free movement of capital, equation 12. The entry condition is therefore derived by substituting the equilibrium tax level in the expression of profits from the market game (see Annex 3.1).

The equilibrium tax level is derived from the relevant first-order condition in the previous section, equation 21 above. This can be re-written as

$$t_1 \alpha_1 = (1 + d_1 \alpha_1) Y_1 - \beta_1 X_1 \tag{30}$$

where

$$\beta_1 = \frac{N_1}{N_1 + N_2 + 1} > 0 \tag{31}$$

$$\alpha_1 = \beta_1 (1 + N_2)(m_1 + m_2) > 0 \quad \alpha_1 > \beta_1 \tag{32}$$

Substituting the expressions in the Annex for Y_1 and X_1 in 30, I get

$$\begin{aligned} t_1 \alpha_1 = {} & (1 + d_1 \alpha_1)(m_1 + m_2) \beta_1 \left[a + N_2 (c_2 + t_2) - (N_2 + 1)(c_1 + t_1)\right] \\ & - \frac{\beta_1 m_1}{(N_1 + N_2 + 1)} \left[a(N_1 + N_2) - N_1 (c_1 + t_1) - N_2 (c_2 + t_2)\right] \end{aligned} \tag{33}$$

This can be re-arranged with the help of equations 31 and 32 as

$$t_1 = a D_1 - c_1 E_1 + c_2 F_1 + t_2 F_1 \tag{34}$$

where

$$D_1 = \frac{(1 + d_1 \alpha_1)(m_1 + m_2) - m_1 (\beta_1 + \beta_2)}{(1 + N_2)(2 + d_1 \alpha_1)(m_1 + m_2) - m_1 \beta_1} \tag{35}$$

$$E_1 = \frac{(1 + d_1 \alpha_1)(m_1 + m_2)(1 + N_2) - m_1 \beta_1}{(1 + N_2)(2 + d_1 \alpha_1)(m_1 + m_2) - m_1 \beta_1} \tag{36}$$

$$F_1 = \frac{(1 + d_1 \alpha_1)(m_1 + m_2) N_2 + m_1 \beta_2}{(1 + N_2)(2 + d_1 \alpha_1)(m_1 + m_2) - m_1 \beta_1} \tag{37}$$

Similarly, for country 2, I obtain

$$t_2 = a D_2 - c_2 E_2 + c_1 F_2 + t_1 F_2 \tag{38}$$

Combining equations 34 and 38, I get

$$t_1 = a \frac{D_1 + F_1 D_2}{1 - F_1 F_2} - c_1 \frac{E_1 - F_1 F_2}{1 - F_1 F_2} + c_2 \frac{F_1 (1 - E_2)}{1 - F_1 F_2} \tag{39}$$

$$t_2 = a\frac{D_2 + F_2 D_1}{1 - F_1 F_2} - c_2\frac{E_2 - F_1 F_2}{1 - F_1 F_2} + c_l\frac{F_2(1 - E_1)}{1 - F_1 F_2} \tag{40}$$

These are then substituted in the expression for profits (see Annex 3.1). Given the number of firms, profits in the post integration equilibrium are

$$\pi_1(N_1, N_2) = \frac{(m_1 + m_2)}{(N_1 + N_2 + 1)^2}\left[a\left(1 + N_2\frac{D_2 + F_2 D_1}{1 - F_1 F_2} - (N_2 + 1)\frac{D_1 + F_1 D_2}{1 - F_1 F_2}\right) + \right.$$
$$\left. + c_2\frac{(1 - E_2)(N_2 - (N_2 + 1)F_1)}{1 - F_1 F_2} - c_1\frac{(1 - E_1)((N_2 + 1) - N_2 F_2)}{1 - F_1 F_2}\right]^2 \tag{41}$$

This expression is positive so that the firms that entered the market before the removal of trade barriers will remain active in the post integration equilibrium. Capital moves freely so that any profits in excess of sunk costs will generate new entry. Thus, entry post integration will take place if

$$\pi_1(N_1^a, N_2^a) > S \tag{42}$$

where N_i^a is the number of firms in country i in the autarky equilibrium (equation 12 above). It is very difficult to make any inferences about the extent of new entry that may result from integration using equation 41. This expression cannot be simplified in any significant form by considering simple forms of asymmetries such that only unit labour costs differ. The number of firms in the autarky equilibrium remains different in each country and this complicates the calculations. I have therefore carried out some simulations to characterise the asymmetric equilibrium and the extent of new entry. These are discussed in the next section.

3.5.2 Cooperative equilibrium

I proceed likewise with the results of the cooperative equilibrium of Section 3.4.2. The first-order condition, equation 29, can also be re-written as

$$t_1 \alpha_1 = (1 + d_1 \alpha_1) Y_1 \tag{43}$$

where

$$\alpha_1 = N_1(m_1 + m_2) \tag{44}$$

Substituting for Y_1 in equation 43 above using the expressions in Annex 3.1, I get

$$t_1 \alpha_1 = (1 + d_1 \alpha_1)\frac{m_1 + m_2}{N_1 + N_2 + 1} N_1\left[a + N_2(c_2 + t_2) - (N_2 + 1)(c_1 + t_1)\right] \tag{45}$$

This can be re-arranged with the help of equation 44 as

$$t_1 = a D_1 - c_1 E_1 + c_2 F_1 + t_2 F_1 \tag{46}$$

where

$$D_1 = \frac{(1 + d_1 \alpha_1)}{(N_1 + N_2 + 1) + (1 + N_2)(1 + d_1 \alpha_1)} \tag{47}$$

$$E_1 = \frac{(1 + N_2)(1 + d_1 \alpha_1)}{(N_1 + N_2 + 1) + (1 + N_2)(1 + d_1 \alpha_1)} \tag{48}$$

$$F_1 = \frac{N_2(1 + d_1 \alpha_1)}{(N_1 + N_2 + 1) + (1 + N_2)(1 + d_1 \alpha_1)} \tag{49}$$

Similarly, for country 2, I obtain

$$t_2 = a D_2 - c_2 E_2 + c_1 F_2 + t_1 F_2 \tag{50}$$

These can be in the expression for profits from the market stage (see Annex 3.1). The result is similar to the one obtained for the non-cooperative equilibrium in the section immediately above and I will therefore explore the extent of new entry using simulations.

3.6 Simulation of equiiibrium results and new entry

In this section, I use a simulation model to explore two hypotheses about the asymmetric equilibrium. First, I want to show that entry post integration is more likely than in a symmetric equilibrium. For this purpose, I want to show that for each symmetric equilibrium with new entry, there are a number of asymmetric equilibria where entry will also take place. I also want to explore whether the entry condition identified for the symmetric equilibrium tells us something about the extent of new entry when applied to each country separately. Second, I want to compare entry in the non-cooperative equilibrium with entry in the cooperative equilibrium.

The simulation model works as follows. First, the autarky equilibrium is solved and the number of firms from equation 12 is determined. The model then solves the equilibrium post-integration given the number of firms in the autarky equilibrium to verify whether entry will take place. In the non-cooperative equilibrium, the equilibrium level of taxes is determined by equations 30 to 32 and profits are determined by equation 41. In the cooperative equilibrium, the level of taxes is determined by equations 43 and 44.

The base case is a symmetric equilibrium where the condition for new entry identified in equation 17 is satisfied. The data is presented in Table A3.1 in Annex 3.2 where I have defined four such scenarios A, B, C, D. They all have identical parameters except for unit labour costs that decrease between the level in scenario A and the level in scenario D. I explore the effect of asymmetries in unit labour costs, marginal damage to the environment and population size. Tables A3.2 to A3.4 contain the post-integration results of the non-cooperative equilibrium and tables A3.5 to A3.7 the results of the cooperative equilibrium.

3.6.1 Different unit labour costs

Integration increases the size of the market for all firms. Differences in unit labour costs suggest that the country with lower unit labour costs will increase its share of the integrated market (at the expense of the other country) and that in equilibrium this country will be a net exporter. Equation 22 suggests that in this country, environmental policy will be tougher than the first-best. The overall impact that this will have on entry is unclear as equation 41 suggests.

The unit labour costs in country 1 and 2 are defined as

$$c_1 = c\,(1+H) \qquad (51)$$

$$c_2 = c\,(1-H) \qquad (52)$$

where c is the unit labour cost in the base case (see Table A3.1) and H is either 10% or 20% or 50%. Note that the difference in unit labour costs between the two countries is $2H$. However, the average marginal cost of the two countries remains unchanged.

Table A3.2 presents the results of the non-cooperative equilibrium. Each of the panels a., b., and c. refers to one of the above values of H and each column reports the relevant results for one of the four scenarios in Table A3.1. The equilibrium results show no indication of a race to the bottom in tax setting. As suggested, environmental taxes after integration in the country with low unit labour costs are higher than in the other country. However, total unit costs, $c_i + t_i$, will still be lower in the country with low unit labour costs and it is this country that exports widgets. It is worth noting that lower unit labour costs in both countries, i.e. moving from scenario A to D in Table A3.2, results in lower environmental taxes in the exporting country and higher environmental taxes in the importing country. The intuition of the result is as follows: lower unit labour costs increase consumption levels and hence production. Equation 22 suggests that this change will elicit different responses from each country. The exporting country will wish to capture a bigger share of the global market and to reduce environmental taxes. Tax setting behaviour in the importing country is not be affected by these considerations so environmental taxes increase.

Table A3.2 also summarises the extent of new entry in the non-cooperative equilibrium. The table refers first to the condition for autarky equilibrium. I want to verify that the entry condition in equation 13 is satisfied when the two countries are still separated by trade barriers. If this condition is not satisfied, there is no industry in the autarky equilibrium and I do not need to consider the extent of entry post integration. This is the case of scenario A, when I consider an increase in costs by 10% over the level of the base case.

The table also shows whether the entry condition identified in the post integration equilibrium for two identical countries, equation 16, is satisfied now by each country given the number of firms in the autarky equilibrium. For example, when H is 10% this condition will be satisfied in scenarios B, C and D for both countries. Finally, I calculate the profits in this equilibrium and compare them with sunk costs. When H is 10%, profits exceed sunk costs in both countries in scenarios B and C and in the country with low unit costs in scenario D. This suggests that when the difference in labour costs between countries increases, H increases, entry will shift towards the country with lower unit labour costs. In general terms, I find that there will be entry if at least one of the countries satisfies the entry condition for the symmetric equilibrium; this country will not be necessarily the one in which entry takes place. It suggests that equation 16 may be a necessary condition for entry to take place in the asymmetric equilibrium.

Result 8: in a non-cooperative equilibrium, entry is more likely in the country with low unit costs - the exporting country.

Table A3.5 contains the result of the cooperative equilibrium. It also indicates whether the autarky condition (equation 13) and the entry condition (equation 16) are satisfied in a symmetric equilibrium. These are not affected by the assumption about tax setting behaviour and are included here for illustration. Comparing with the non-cooperative equilibrium, I find that environmental taxes will be even lower in the exporting country and even higher in the importing country. The country with low unit costs remains as a net exporter and profits will be higher in that country. Thus, the likelihood of new entry in the exporting country increases. The results about environmental taxes are in line with the conclusion derived about the toughness of environmental policy based on the first-order conditions (equations 22 and 29). The results also suggest that the difference in environmental policies in the two countries will be smaller than in the non-cooperative equilibrium.

Result 9: governments' cooperation results in lower taxes in the country with low unit labour costs, the exporting country, and increases further the likelihood of entry in that country.

3.6.2 Different marginal damage

The welfare function suggests that, other things being equal, the larger the damage to the local environment the higher the environmental tax. This suggests that the environmental tax should be higher in the country with larger marginal damage to the environment. It is, however, unclear, how governments' rent seeking behaviour will affect the extent of entry in these circumstances.

The marginal damage to the local environment in country 1 and 2 are defined as

$$d_1 = d\,(1+H) \tag{53}$$

$$d_2 = d\,(1-H) \tag{54}$$

where d is the marginal damage in the base case (see Table A3.1) and H is 10%, 20% or 50%. Thus, the average marginal emission in the two countries is unaffected.

As one will expect, the results of the non-cooperative equilibrium in Table A3.3 show that the country with large marginal damage to the local environment imposes a higher tax than the country with a small marginal damage. The country with large marginal damage imports from the other. However, as suggested by equation 22, environmental policy is tougher in the exporting country, where marginal damage is small. This result seems to be independent of the degree of competition in the product market. Rauscher

(1994) shows that in the case of perfect competition in product markets, a country that is a net exporter of a pollution intensive good will want to pursue environmental policies that are tougher than the first-best.

Lower unit labour costs results in a lower environmental tax and a more lax environmental policy in the post integration equilibrium in both countries. This is so because lowering unit labour costs increases the size of the industry in the autarky equilibrium. Equation 22 suggests that the increase in the industry size is the dominant effect in determining the toughness of environmental policy.

Table A3.3 also shows that as the difference between the marginal damage in the two countries increases (i.e. moving from panel a. to c.), profits will concentrate in the country with lower valuation of the environment, the exporting country. This can be seen as a form of "*nimbysm*" (not in my backyard): high environmental damage affects the flow of new entry. The results are similar to those presented in Table A3.2, profits exceed sunk costs where the condition for entry in the symmetric equilibrium, equation 16, is satisfied by at least one of the countries. Equation 16 will also be a necessary condition for entry to take place.

Table A3.6 contains the result of the cooperative equilibrium. In equilibrium, the country where marginal damage is small remains a net exporter and the difference in environmental policies between the two countries is smaller. Profits in the exporting country, and the consequent likelihood of entry, will be higher than in the non-cooperative equilibrium.

Result 10: in a non-cooperative tax setting context, entry is more likely in the country where marginal damage to the environment is small, the exporting country. Governments' cooperation will increase this further.

3.6.3 Different population

The size of the population affects the industry structure in the autarky equilibrium. From equation 12, I obtain that $\frac{\partial N^a}{\partial m} > 0$ if the product of marginal damage and population is smaller than one, $d\ m < 1$. This condition is always satisfied when the entry condition, equation 17, is satisfied. So, other things being equal, the country with a large population will have a larger industry. Thus competition will be different in the two countries; it will be tougher in the country with a large population.

The autarky equilibrium is very sensitive to the size of the country. For example, given the parameters in Table A3.1, reducing the level of the population could not be considered: the condition for the existence of equilibrium will not be satisfied. This is so because changing the population affects the right hand side of the autarky equilibrium condition, equation 13, proportionally whereas the effect on the left-hand side comes in a square root. I have therefore considered discrete increases in the size of the population for these simulations of the equilibrium results.

Table A3.4 contains the results of the non-cooperative equilibrium. A comparison of the three sets of simulations (panels a., b. and c.) confirms that the environmental tax is very sensitive to the population size. It suggests that the larger the integrated market the lower the taxes (in line with Result 4 above). Taxes will be higher in the large country.

The direction of trade depends on the interaction of population size with unit labour costs. The results suggest that the small country exports when the unit labour costs are low such as in scenario D and the large country exports when unit labour costs are high such as in scenario A. Profits, however, will be higher in the small country. Table A3.4 also summarises the extent of new entry and suggests that profits in excess of sunk costs – new entry – will be more likely in the small country (though this also depends on the actual size of both countries). The results suggest that the conditions for entry identified in the symmetric equilibrium, equation 16, may be a necessary condition for entry.

Table A3.7 contains the results of the cooperative equilibrium. The direction of trade remains unaffected and that profits in the small (large) country will be higher (smaller) than in the non-cooperative equilibrium. New entry will therefore be more likely than in the cooperative equilibrium.

Result 11: in a non-cooperative tax setting context, entry is more likely in the small country. Governments' cooperation will increase this further.

3.7 Qualifications

In this chapter, I have extended the model in Ulph (1995) to consider asymmetries between the two countries and the effect of an alternative approach to tax setting behaviour. The model could be extended further (or in alternative ways) to explore the effect of different functional forms or a larger number of countries. In the next chapter, I explore the effect of relaxing the assumption that unit labour costs are fixed. Here, I want to discuss three additional extensions of the model that relate to important assumptions of the model.

First, the model assumed that removing trade barriers results in the integration of product markets so that the same price prevails in both countries. Alternatively, product markets may remain segmented where each firm perceives each country as a separate market and makes distinct quantity decisions for each one. (The implicit assumption is then that re-selling is not possible.) This is the approach adopted in Brander and Spencer (1985) and Brander and Krugman (1983). It is also the standard assumption in the analysis of strategic trade policy, Ulph (1997). Suppose that x_i denotes the output for domestic consumption from a firm in country i and that y_i refers to the output of the same firm for consumption in country j (country i export). X_i and Y_i refer to the output over all firms. If there are two countries producing and consuming and product markets are segregated, we can rewrite the profit function for a firm in country 1 as

$$\Pi_1 = x_1 p(X_1 + Y_2) + y_1 p(Y_1 + X_2) - (c_1 + t_1)(x_1 + y_1) \quad (55)$$

One could then solve stage 3 of the model in the usual way and substitute the results to obtain the new environmental tax.

Second, the model assumed that the pollution is domestic. An alternative case is that pollution extends beyond national borders.[13] Extending the model for transboundary pollution affects the country's welfare function: the environmental damage in each country will be affected by the total amount produced between the two countries and the extent to which pollution travels across borders. One could model transboundary pollution in such a way that it nests the approach adopted here for local pollution. Denote by Y_1 the polluting output from country 1. Supposing that damage to the

[13] Worth noting that environmental damage can result from flows of pollution or stocks of pollution. Here we focus on the former.

environment remains proportional to a measure of output, we could model the polluting production in country 1 as

$$Y_1 + \beta\, Y_2 \qquad (56)$$

where β is a parameter representing the proportion of output from country 2 that pollutes country 1. So where pollution is local β is zero. If the transboundary pollution becomes weaker as it travels such as acid rain then $\beta < 1$. If the transboundary pollution is global then $\beta = 1$. We can then substitute equation 56 into the welfare functions (equations 7 and 9) to characterise the environmental tax and explore how environmental policy is affected when the type of pollution changes. One interesting aspect of cases where $\beta > 0$ is that environmental policy is affected by the other country's production even when there is no international trade.

Kennedy (1994) adopts the approach in equation 56 above and provides a useful taxonomy of the effect of strategic behaviour when there are no trade barriers. The paper identifies two distortions: a rent capture and a pollution shifting distortion. The rent capture distortion is similar in spirit to the export subsidy argument in the strategic trade policy literature. It lowers the environmental tax to gain an increasing share of exports. The pollution shifting effect increases the environmental tax. The intuition is simple. If there is free trade in goods, a government can reduce domestic pollution with a lower adverse impact on domestic consumption. When pollution is transboundary, the second effect weakens and it disappears when pollution is global. Thus, it is shown that the net effect on a symmetric equilibrium is to reduce the environmental tax. The model used in Kennedy's paper is different from that used in this chapter. For example, there is no entry, product markets remain segregated and local residents own firms. The strength of the arguments suggests that this will be a useful extension of the model developed in this chapter.

Third, the model assumes that residents in a third country own local firms so that profits go abroad and do not affect society's welfare. Welzel (1995) explores the effect of ownership in the context of strategic trade policy, which normally assumes local ownership. He shows that when either local residents have a stake in foreign firms or foreigners have a stake in local firms or both, the optimal trade subsidy to local producers will be smaller than when ownership is purely local. The model presented here could be extended to allow for (partial) local ownership of firms. The qualitative effect on a government's tax setting will be determined by the combination of two opposing changes to society's welfare. First, the part of the profits reverting to local consumers increases society's welfare. So, other things being equal, the same level of welfare necessitates a lower environmental tax. Second, and following from the foregoing, part of the tax revenue does not contribute to society's welfare – it is a transfer of resources – so that society's welfare is lower. So, other things being equal, the same level of welfare necessitates a higher environmental tax. In the autarky equilibrium, free mobility of capital reduced profits net of sunk costs to zero,[14] so that the net effect of local ownership on welfare will be negative. This suggests that if ownership is shared between locals and foreigners, the equilibrium environmental tax will be higher and the number of firms will be smaller.[15] The reduced number of firms in the autarky equilibrium makes competition in the post integration equilibrium less intense so that profits may be higher and entry more likely when trade barriers are eliminated.

3.8 Conclusions

In this chapter, I have examined a specific aspect of the literature on trade and the environment. This is the effect of abolishing trade barriers on an environmental tax and industry entry when governments and firms behave strategically and pollution is local.

Ulph (1995) showed that in a symmetric equilibrium the environmental tax will remain above first-best tax (marginal damage) and that new entry is unlikely following the removal of trade barriers. I have extended the analysis in Ulph (1995) to focus on differences between countries and on governments' tax setting behaviour.

The main features of the model used in this chapter are summarised in Table 3.1 and are contrasted with the standard assumptions made in this area as per the survey in Ulph (1997). The first four cells in the last column represent the main assumptions in Ulph (1995) whereas the last two cells represent this chapter's contribution.

[14] It is implicitly assumed that local capital is as mobile as foreign capital and can be employed abroad.

[15] Profits in the autarky equilibrium, however, will be the same: free movement of capital ensures that these must cover sunk costs, which are unaffected.

Table 3.1: Features of the model

Assumptions	Standard assumptions	This chapter
Consumers	Located in a third country	Domestic and in the other producing country
Number of firms per country	One	Determined by free mobility of capital
Ownership of firms	Local	Third country
Industry structure	Exogenous	Endogenous but no government pre-commitment to a level of tax before entry
	Product markets are segmented	Product markets are integrated
Similarities between countries	Completely identical	Identical technology and differences in unit labour costs, marginal damage to the environment and country size
Governments' tax setting behaviour	Non-cooperative	Non-cooperative and cooperative

I find that when the two countries are not identical, the environmental tax remains above the first-best tax regardless of the assumption about government tax setting behaviour. I also characterise a country's environmental policy. I show that a non-cooperative approach to tax setting behaviour results in the environmental tax being determined by a trade externality and a competition externality. If governments set taxes cooperatively, the trade externality will be internalised.

I also explore the likelihood of new entry when countries are different using simulations. They suggest that entry is more likely in an asymmetric equilibrium and that entry will be associated with the exporting country or with the small country when countries set taxes non-cooperatively. Setting environmental taxes cooperatively increases further the likelihood of entry.

These results have been derived from a very specific model – local pollution, integrated product markets, etc. They add to the body of evidence presented in Wilson (1996) about the likelihood of a race to the bottom in environmental taxes. However, it is not possible to rule out such a race to the bottom because of the specific assumptions made.

Annex 3.1: Results of the market game in the post integration equilibrium

This annex summarises the results of the market game (stage 3) in the post integration equilibrium from Ulph (1995). It also presents the derivatives used to characterise the asymmetric equilibrium in the main text.

Assuming that firms choose outputs as Cournot's oligopolists, the equilibrium output in country i is

$$Y_i = \frac{N_i}{(N_1 + N_2 + 1)}(m_1 + m_2)[a + N_j(c_j + t_j) - (N_j + 1)(c_i + t_i)] \qquad \text{(A1)}$$

where $i, j = 1, 2$ and $i \neq j$. The following derivatives can be obtained from the previous equation

$$\frac{dY_i}{dt_i} = -(m_1 + m_2)\frac{N_i \ (N_j + 1)}{(N_1 + N_2 + 1)} \qquad \text{(A2)}$$

$$\frac{dY_i}{dt_j} = (m_1 + m_2)\frac{N_i \ N_j}{(N_1 + N_2 + 1)} \qquad \text{(A3)}$$

This suggests that, other things being equal, an increase in own tax leads to a reduction in own output. It also suggests that a similar change in the environmental tax in the other country leads to a (smaller) increase in the country's output.

Consumption in country 1 is

$$X_1 = \frac{m_1}{(N_1 + N_2 + 1)}[(N_1 + N_2)a - N_1(c_1 + t_1) - N_2(c_2 + t_2)] \qquad \text{(A4)}$$

Total consumption in the two countries is

$$X_1 + X_2 = \frac{m_1 + m_2}{(N_1 + N_2 + 1)}[N_1(a - c_1 - t_1) + N_2(a - c_2 - t_2)] \qquad \text{(A5)}$$

This suggests that an individual's consumption is a weighted average of consumption in the autarky equilibrium assuming marginal cost pricing. The following derivatives can be obtained from the above

$$\frac{\partial(X_1 + X_2)}{\partial t_1} = -\frac{(m_1 + m_2)N_1}{(N_1 + N_2 + 1)} \qquad \text{(A6)}$$

$$\frac{\partial(X_1+X_2)}{\partial t_2}=-\frac{(m_1+m_2)N_2}{(N_1+N_2+1)} \tag{A7}$$

They suggest that an increase in taxes will always reduce total consumption and that the larger the industry in that country the larger the impact of an increase in taxes in that country.

A firm's profit in country 1 is

$$\pi_1=(m_1+m_2)\left(\frac{a+N_2(c_2+t_2)-(N_2+1)(c_1+t_1)}{N_1+N_2+1}\right)^2 \tag{A8}$$

Annex 3.2: Results of simulations

Table A3.1: Base case

	Scenario A	Scenario B	Scenario C	Scenario D
Tastes, *a*	3	3	3	3
Sunk costs, *S*	2	2	2	2
Population, *m*	3	3	3	3
Marginal damage, *d*	0.02	0.02	0.02	0.02
Unit costs, *c*	0.5	0.4	0.3	0.25

Notes for tables A3.2 to A3.7:

- all the numerical results refer to the post integration equilibrium;
- "autarky equilibrium condition" refers to equation 13 in the text;
- "entry equilibrium condition" refers to equation 16.

Both conditions are satisfied in the base case scenarios defined in Table A3.1. I verify in the tables whether these conditions are satisfied once a degree of asymmetry between the countries is introduced. If the autarky condition is not satisfied then there is no industry and I do not need to consider the extent of entry post integration.

Table A3.2: Non-cooperative equilibrium: different unit labour costs

a. costs increased/reduced by 10%

Country 1; c_1= 1.1c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.640	0.644	0.649	0.651
Marginal damage, d_1Y_1	0.067	0.075	0.083	0.087
Consumption, X_1	3.645	3.990	4.329	4.497
Production, Y_1	3.370	3.766	4.157	4.351
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.124 (Yes)	2.060 (Yes)	2.006 (Yes)	1.982 (Yes)
Country 2; c_2=0.9c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.719	0.706	0.695	0.690
Marginal damage, d_2Y_2	0.078	0.084	0.090	0.093
Consumption, X_2	3.645	3.990	4.329	4.497
Production, Y_2	3.920	4.215	4.502	4.642
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	Yes	Yes	Yes
Profits (exceed sunk costs)	2.281 (Yes)	2.187 (Yes)	2.102 (Yes)	2.063 (Yes)

b. costs increased/reduced by 20%

Country 1; c_1= 1.2c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.602	0.613	0.626	0.636
Marginal damage, d_1Y_1	0.062	0.071	0.080	0.084
Consumption, X_1	3.647	3.992	4.330	4.497
Production, Y_1	3.099	3.543	3.986	4.207
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.037 (Yes)	1.992 (No)	1.956 (No)	1.941 (No)
Country 2; c_2= 0.8c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.758	0.738	0.718	0.709
Marginal damage, d_2Y_2	0.084	0.089	0.093	0.096
Consumption, X_2	3.647	3.992	4.330	4.497
Production, Y_2	4.196	4.441	4.675	4.788
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	No
Profits (exceed sunk costs)	2.351 (Yes)	2.246 (Yes)	2.148 (Yes)	2.101 (Yes)

c. costs increased/reduced by 50%

Country 1; c_1= 1.5c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.488	0.522	0.557	0.575
Marginal damage, d_1Y_1	0.046	0.058	0.070	0.076
Consumption, X_1	3.665	4.003	4.336	4.501
Production, Y_1	2.311	2.890	3.479	3.777
Autarky equilibrium condition	No	No	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	1.752 (No)	1.772 (No)	1.797 (No)	1.81 (No)
Country 2; c_2= 0.5c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.888	0.833	0.789	0.768
Marginal damage, d_2Y_2	0.100	0.102	0.104	0.105
Consumption, X_2	3.665	4.003	4.330	4.497
Production, Y_2	5.019	5.117	5.194	5.225
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	No	No	No
Profits (exceed sunk costs)	2.52 (Yes)	2.4 (Yes)	2.274 (Yes)	2.21 (Yes)

Table A3.3: Non-cooperative equilibrium: different marginal damage

a. damage increased/reduced by 10%

Country 1; d_1=1.1d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.681	0.677	0.674	0.673
Marginal damage, d_1Y_1	0.079	0.087	0.094	0.098
Consumption, X_1	3.644	3.990	4.329	4.497
Production, Y_1	3.612	3.953	4.287	4.452
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.191 (Yes)	2.110 (Yes)	2.038 (Yes)	2.006 (Yes)
Country 2; d_2=0.9d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.677	0.673	0.669	0.668
Marginal damage, d_2Y_2	0.066	0.072	0.079	0.082
Consumption, X_2	3.644	3.990	4.329	4.497
Production, Y_2	3.677	4.027	4.371	4.451
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.220 (Yes)	2.147 (Yes)	2.071 (Yes)	2.040 (Yes)

b. damage increased/reduced by 20%

Country 1; d_1=1.2d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.683	0.679	0.676	0.675
Marginal damage, d_1Y_1	0.086	0.094	0.102	0.106
Consumption, X_1	3.645	3.991	4.330	4.497
Production, Y_1	3.580	3.916	4.246	4.409
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.177 (Yes)	2.094 (Yes)	2.022 (Yes)	1.989 (No)
Country 2; d_2=0.8d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.675	0.670	0.667	0.665
Marginal damage, d_2Y_2	0.059	0.065	0.071	0.073
Consumption, X_2	3.645	3.991	4.330	4.497
Production, Y_2	3.709	4.065	4.414	4.586
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.234 (Yes)	2.156 (Yes)	2.088 (Yes)	2.057 (Yes)

c. damage increased/reduced by 50%

Country 1; d_1=1.5d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.688	0.684	0.682	0.681
Marginal damage, d_1Y_1	0.105	0.114	0.124	0.128
Consumption, X_1	3.648	3.995	4.334	4.502
Production, Y_1	3.486	3.809	4.125	4.280
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.134 (Yes)	2.048 (Yes)	1.973 (No)	1.938 (No)
Country 1; d_2=0.5d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.668	0.663	0.658	0.656
Marginal damage, d_2Y_2	0.038	0.042	0.045	0.047
Consumption, X_2	3.648	3.995	4.334	4.502
Production, Y_2	3.811	4.180	4.544	4.724
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.277 (Yes)	2.203 (Yes)	2.138 (Yes)	2.109 (Yes)

Table A3.4: Non-cooperative equilibrium: different size (population)

a. $m_1 > m_2 = m$

Country 1; m=4	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.647	0.644	0.643	0.642
Marginal damage, d_1Y_1	0.107	0.115	0.123	0.127
Consumption, X_1	5.269	5.711	6.146	6.361
Production, Y_1	5.333	5.737	6.137	6.334
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	2.010 (Yes)	1.950 (No)	1.898 (No)	1.873 (No)
Country 2; m = 3	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.628	0.628	0.628	0.629
Marginal damage, d_2Y_2	0.078	0.085	0.092	0.096
Consumption, X_2	3.952	4.283	4.610	4.771
Production, Y_2	3.888	4.257	4.619	4.798
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.152 (Yes)	2.074 (Yes)	2.006 (Yes)	1.975 (No)

b. $m_1 >> m_2 = m$

Country 1; m=5	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.628	0.626	0.626	0.626
Marginal damage, d_1Y_1	0.139	0.148	0.157	0.162
Consumption, X_1	6.921	7.459	7.989	8.251
Production, Y_1	6.933	7.398	7.858	8.086
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.906 (No)	1.858 (No)	1.814 (No)	1.794 (No)
Country 2; m=3	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.599	0.601	0.603	0.604
Marginal damage, d_2Y_2	0.083	0.091	0.098	0.102
Consumption, X_2	4.153	4.475	4.793	4.951
Production, Y_2	4.141	4.536	4.924	5.115
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.136 (Yes)	2.06 (Yes)	1.994 (No)	1.964 (No)

c. $m_1 > m_2 > m$

Country 1; m = 5	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.592	0.594	0.598	0.599
Marginal damage, d_1Y_1	0.144	0.154	0.163	0.168
Consumption, X_1	7.284	7.801	8.312	8.566
Production, Y_1	7.210	7.692	8.169	8.406
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.832 (No)	1.785 (No)	1.743 (No)	1.723 (No)
Country 2; m = 4	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.582	0.585	0.589	0.591
Marginal damage, d_2Y_2	0.118	0.127	0.136	0.140
Consumption, X_2	5.827	6.241	6.650	6.853
Production, Y_2	5.901	6.349	6.793	7.012
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.914 (No)	1.858 (No)	1.808 (No)	1.786 (No)

Table A3.5: Cooperative equilibrium: different unit labour costs

a. costs increased/reduced by 10%

Country 1; c_1= 1.1c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.650	0.652	0.654	0.656
Marginal damage, d_1Y_1	0.066	0.074	0.082	0.086
Consumption, X_1	3.648	3.992	4.330	4.497
Production, Y_1	3.308	3.712	4.113	4.314
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.046 (Yes)	2.001 (Yes)	1.964 (No)	1.948 (No)
Country 2; c_2=0.9c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.707	0.697	0.689	0.685
Marginal damage, d_2Y_2	0.080	0.085	0.091	0.094
Consumption, X_2	3.648	3.992	4.330	4.497
Production, Y_2	3.987	4.272	4.547	4.680
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.360 (Yes)	2.247 (Yes)	2.145 (Yes)	2.097 (Yes)

b. costs increased/reduced by 20%

Country 1; c_1= 1.2c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.620	0.628	0.637	0.641
Marginal damage, d_1Y_1	0.060	0.069	0.078	0.083
Consumption, X_1	3.658	3.998	4.334	4.499
Production, Y_1	2.980	3.439	3.901	4.133
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	1.855 (No)	1.876 (No)	1.873 (No)	1.873 (No)
Country 2; c_2= 0.8c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.734	0.719	0.705	0.699
Marginal damage, d_2Y_2	0.087	0.091	0.095	0.097
Consumption, X_2	3.658	3.998	4.334	4.499
Production, Y_2	4.337	4.558	4.767	4.866
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	No
Profits (exceed sunk costs)	2.511 (Yes)	2.366 (Yes)	2.233 (Yes)	2.170 (Yes)

c. costs increased/reduced by 50%

Country 1; c_1= 1.5c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.523	0.552	0.581	0.596
Marginal damage, d_1Y_1	0.041	0.053	0.066	0.072
Consumption, X_1	3.375	4.044	4.358	4.516
Production, Y_1	2.061	2.655	3.280	3.602
Autarky equilibrium condition	No	No	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	1.393 (No)	1.496 (No)	1.597 (No)	1.646 (No)
Country 2; c_2= 0.5c	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.807	0.780	0.753	0.739
Marginal damage, d_2Y_2	0.108	0.109	0.109	0.009
Consumption, X_2	3.735	4.044	4.358	4.516
Production, Y_2	5.408	5.433	5.436	5.429
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	No	No	No
Profits (exceed sunk costs)	2.927 (Yes)	2.706 (Yes)	2.491 (Yes)	2.386 (Yes)

Table A3.6: Cooperative equilibrium: different marginal damage to environment

a. damage increased/reduced by 10%

Country 1; d_1=1.1d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.682	0.678	0.676	0.674
Marginal damage, d_1Y_1	0.079	0.087	0.094	0.098
Consumption, X_1	3.644	3.990	4.329	4.497
Production, Y_1	3.604	3.944	4.277	4.441
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.182 (Yes)	2.100 (Yes)	2.028 (Yes)	1.996 (No)
Country 2; d_2=0.9d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.676	0.671	0.668	0.667
Marginal damage, d_2Y_2	0.066	0.073	0.079	0.082
Consumption, X_2	3.644	3.990	4.329	4.497
Production, Y_2	3.684	4.036	4.382	4.553
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.229 (Yes)	2.150 (Yes)	2.082 (Yes)	2.050 (Yes)

b. damage increased/reduced by 20%

Country 1; d_1=1.2d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.685	0.682	0.679	0.678
Marginal damage, d_1Y_1	0.086	0.094	0.101	0.105
Consumption, X_1	3.645	3.991	4.330	4.498
Production, Y_1	3.565	3.898	4.225	4.386
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.158 (Yes)	2.075 (Yes)	2.002 (Yes)	1.968 (No)
Country 2; d_2=0.8d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.672	0.668	0.664	0.662
Marginal damage, d_2Y_2	0.060	0.065	0.071	0.074
Consumption, X_2	3.645	3.991	4.330	4.498
Production, Y_2	3.725	4.083	4.435	4.610
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.253 (Yes)	2.176 (Yes)	2.109 (Yes)	2.078 (Yes)

c. damage increased/reduced by 50%

Country 1; d_1=1.5d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.693	0.690	0.688	0.688
Marginal damage, d_1Y_1	0.103	0.113	0.122	0.127
Consumption, X_1	3.650	3.996	4.336	4.504
Production, Y_1	3.449	3.765	4.073	4.224
Autarky equilibrium condition	No	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.089 (Yes)	2.001 (Yes)	1.923 (No)	1.888 (No)
Country 1; d_2=0.5d	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.661	0.655	0.650	0.648
Marginal damage, d_2Y_2	0.039	0.042	0.046	0.048
Consumption, X_2	3.650	3.996	4.336	4.504
Production, Y_2	3.850	4.228	4.600	4.784
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.325 (Yes)	2.253 (Yes)	2.191 (Yes)	2.163 (Yes)

Table A3.7: Cooperative equilibrium: different size (population)

a. $m_1 > m_2 = m$

Country 1; m=4	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.6446	0.6435	0.6431	0.6432
Marginal damage, d_2Y_2	0.107	0.115	0.123	0.127
Consumption, X_2	5.271	5.712	6.146	6.361
Production, Y_2	5.351	5.745	6.134	6.326
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	2.023 (Yes)	1.956 (No)	1.896 (No)	1.896 (No)
Country 2; m = 3	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.6299	0.6285	0.6280	0.6280
Marginal damage, d_2Y_2	0.077	0.085	0.092	0.096
Consumption, X_2	3.953	4.284	4.609	4.770
Production, Y_2	3.874	4.251	4.622	4.805
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.136 (Yes)	2.068 (Yes)	2.008 (Yes)	1.980 (No)

b. $m_1 >> m_2 = m$

Country 1; m = 5	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.6272	0.6283	0.630	0.6311
Marginal damage, d_1Y_1	0.139	0.148	0.156	0.161
Consumption, X_1	6.922	7.455	7.982	8.242
Production, Y_1	6.937	7.379	7.817	8.034
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.908 (No)	1.848 (No)	1.795 (No)	1.770 (No)
Country 2; m = 3	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.5992	0.6000	0.6014	0.6023
Marginal damage, d_2Y_2	0.083	0.091	0.099	0.103
Consumption, X_2	4.153	4.473	4.789	4.945
Production, Y_2	4.139	4.549	4.954	5.154
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	Yes	Yes	Yes	Yes
Profits (exceed sunk costs)	2.133 (Yes)	2.072 (Yes)	2.018 (Yes)	1.994 (No)

c. $m_1 > m_2 > m$

Country 1; m = 5	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_1	0.5937	0.5967	0.6003	0.6023
Marginal damage, d_1Y_1	0.144	0.153	0.163	0.167
Consumption, X_1	7.282	7.799	8.310	8.563
Production, Y_1	7.188	7.660	8.127	8.359
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.822 (No)	1.770 (No)	1.725 (No)	1.704 (No)
Country 2; m = 4	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**
Tax, t_2	0.581	0.5839	0.5873	0.5892
Marginal damage, d_2Y_2	0.118	0.128	0.137	0.141
Consumption, X_2	5.826	6.239	6.648	6.851
Production, Y_2	5.919	6.377	6.830	7.055
Autarky equilibrium condition	Yes	Yes	Yes	Yes
Entry equilibrium condition	No	No	No	No
Profits (exceed sunk costs)	1.926 (No)	1.874 (No)	1.828 (No)	1.807 (No)

References

Barrett, S. (1994), "Strategic environmental policy and international trade", *Journal of Public Economics*, vol. 54, pp. 325 – 338

Brander, J. and P. Krugman (1983), "A 'reciprocal dumping' model of international trade", *Journal of International Economics*, vol. 15, pp. 313 – 321

Brander, J. and B. J. Spencer (1985), "Export subsidies and international market share rivalry", *Journal of International Economics*, vol. 18, pp. 83 – 100

Carmichael, C. M. (1987), "The control of export subsidies and its welfare consequences", *Journal of International Economics*, vol. 23, pp. 1 – 19

Carmichael, C. M. (1991), "The interdependence of firms and government behaviour: Boeing and Eximbank", *Applied Economics*, vol. 23, pp. 107 – 112

Conrad, K. (1997), "Environmental tax competition – a simulation study for asymmetric countries", in S. Proost and J. B. Braden (eds.), *The Economic Theory of Environmental Policy in a Federal System,* Edward Elgar Publishers, Cheltenham

Daly, H. E. (1993), "The perils of free trade", *Scientific American*, November

Eaton, J. and G. M. Grossman (1986), "Optimal trade and industrial policy under oligopoly", *Quarterly Journal of Economics*, vol. 102, pp. 383 – 406

Hoel, M. (1997), "Environmental policy with endogenous plant location", *Scandinavian Journal of Economics*, vol. 99, pp. 241 – 259

Jeppesen, T., J. A. List, H. Folmer (2002), "Environmental regulations and new plant location decisions: evidence from a meta-analysis", *Journal of Regional Science*, vol. 42, pp. 19 – 49

Katsoulacos, Y., D. Ulph, and A. Xepapadeas (1996), "*Environmental taxes in international asymmetric oligopolies*", Fondazione Enrico Mattei, Working Paper 63.96

Kennedy, P. W. (1994), "Equilibrium pollution taxes in open economies with imperfect competition", *Journal of Environmental Economics and Management*, vol. 27, pp. 49 – 63

Krutilla, K. (1991), "Environmental regulation in an open economy", *Journal of Environmental Economics and Management*, vol. 20, pp. 127 – 142

Leahy, D. and C. Montagna (2001), "Strategic trade policy with heterogeneous costs", *Bulletin of Economic Research*, vol. 53, pp. 177 – 182

Markusen, J., E. Morey, and N. Olewiler (1993), "Environmental policy when market structure and plant location are endogenous", *Journal of Environmental Economics and Management*, vol. 24, pp. 69 – 89

Markusen, J., E. Morey, and N. Olewiler (1995), "Competition in regional environmental policies when plant locations are endogenous", *Journal of Public Economics*, vol. 56, pp. 55 – 77

Neary, J. P. (1994), "Cost asymmetries in international subsidy games: should governments help winners or losers?", *Journal of International Economics*, vol. 37, pp. 197 – 218

Neary, J. P. and D. Leahy (2000), "Strategic trade and industrial policy towards dynamic oligopolies", *Economic Journal*, vol. 110, pp. 484 – 508

Petrakis, E. and A. Xepapadeas (1999), "Does government pre-commitment promote environmental innovation?", in E. Petrakis, E. S. Sartzetakis and E. Xepapadeas (eds.), *Environmental regulation and market power. Competition, time consistency and international trade*, Edward Elgar Publishers, Cheltenham

Rauscher, M. (1994), "On ecological dumping", *Oxford Economic Papers*, vol. 46, pp. 820 – 844

Spencer, B. J. and J. A. Brander (1992), "Pre-commitment and flexibility", *European Economic Review*, vol. 36, pp. 1601 – 1626

Sturm, D. (2002), "*Trade and the environment: a survey of the literature*", Mimeo, University of Munich (available from http://www.vwl.uni-muenchen.de/ls_marin/sturm/)

Ulph, A. (1994), "*Environmental policy, plant location and government protection*", in C. Carraro (eds.), *Trade, innovation and the environment*, Kluwer Academic Publishers, Dordrecht

Ulph, A. (1997), "Environmental policy and international trade", in C. Carraro and D. Siniscalco (eds.), *New directions in the economic theory of the environment*, Cambridge University Press, Cambridge

Ulph, D. (1995), "*Environmental taxation and entry: a symmetric equilibrium*", Mimeo, University College London

Wilson, J. D. (1996), "Capital mobility and environmental standards: is there a theoretical basis for a race to the bottom?", in J. N. Bhagwati and R. E. Hudec (eds.), *Fair trade and harmonisation: pre requisite for free trade?*, vol. 1, MIT Press in cooperation with the American Society for International Law, Cambridge, Massachusetts

Welzel, P. (1995), "Strategic trade policy with internationally owned firms", *Bulletin of Economic Research*, vol. 47, pp. 221 – 232

Tirole, J. (1988), *The theory of industrial organisation*, MIT Press, Cambridge, Massachusetts

Chapter 4

Environmental taxation and trade: endogenous wages

4.1 Introduction

The liberalisation of trade between countries can have a profound impact on the countries involved and on the environment. The analysis presented in Chapter 3 about environmental taxation and trade assumes that unit labour costs are fixed. The main purpose of this Chapter is to explore the robustness of the conclusions in Chapter 3 by relaxing the assumption that unit labour costs are fixed. This is important because the effect on wages determines the distribution of gains from trade liberalisation.

The analysis is therefore based on the asymmetric version of the model in Ulph (1995) presented in Chapter 3. It is assumed that the supply of labour is immobile and that firms are wage takers in the labour market. The latter means that an additional strategic dimension into firms' decision making is not introduced. Thus, the stages and the nature of the decisions made by firms and governments remain unaffected. So, for example, governments recognise that the level of environmental tax is affected on wages but they will not try to manipulate them.

Extending the analysis in Chapter 3 raises a number of interesting issues for economic analysis that provide the context for this analysis such as the effect of trade liberalisation on the return to factors (and in particular wages) and the underlying reasons for unemployment. However, I focus on a limited number of issues. Ulph (1995) and Chapter 3 have shown that abolishing trade barriers will lower environmental taxes and I want to explore whether this result appears to be valid when wages are endogenous. Other issues that I will explore are the effect of abolishing trade barriers on equilibrium wages, on the wage differential between countries and on new entry.

The plan of this chapter is as follows. Section 4.2 sets the context for this analysis and the relevant literature. Section 4.3 presents the additional assumptions of the model while Section 4.4 develops the basic intuition about the effect of integration on wages and the likelihood of entry when the two countries are identical. Section 4.5 explores the equilibrium when the two countries are not identical and focuses on the effect of integration on wages. Section 4.6 presents the results of some simulations of the equilibrium results when the two countries are not identical. Section 4.7 qualifies the results. Section 4.8 summarises and concludes.

4.2 Relevant literature

As indicated earlier, extending the analysis in Chapter 3 to relax the assumption that unit labour costs are fixed gives rise to a number of different issues covered in different strands of the economics literature. The purpose of this section is to review them briefly to set the context for the analysis.

The relationship between wages and trade liberalisation has been widely explored in the literature of international trade. Different models offer, however, different predictions about the effect of trade liberalisation on wages, see, for example, Krugman and Obstfeld (2002). First, the standard Hecksher-Olin model with two countries, two goods, two factors and different endowments suggests that factors' return will be equalised when trade barriers are abolished. Samuelson (1949) also shows, in a more general context, that the integration of product markets that may follow the abolition of trade barriers results, in practice, in the integration of labour markets. Second, the standard specific-factor model[1] with similar assumptions about two countries suggests that wages (the return of the non-specific or mobile factor) will be different in the two countries and that the differential could increase as a result of trade liberalisation. Venables (1997) explores two cases of the relationship between trade liberalisation and factor mobility: the immigration from Europe to the US in the late 19th century and the European Union

[1] The standard Hecksher-Olin model assumes that the two factors are perfectly mobile between sectors within each country. The "specific-factor" model relaxes that assumption. It is then assumed that one of the factors, typically capital, is specific to each sector (hence the name "specific-factor" model) and does not move between sectors.

trade policy towards Eastern European countries. He finds that each model can explain one of these two phenomena.[2]

Another relevant issue is the underlying reasons for unemployment. Nickell (1990) suggests that imperfect competition in the relevant product market will not affect the nature of unemployment if there are no further distortions in the labour market.[3] So if there is unemployment it remains voluntary, i.e. people who are looking for work at wages higher than those on offer. The model presented in the next section assumes that there are no distortions in the labour market so that the unemployment is voluntary and that adjustments are instantaneous.

Imperfections in the labour markets suggest that there will be involuntary unemployment. This is defined as a "situation where the unemployed worker is willing to work for less than the wage received by an equally skilled employed worker and yet no job offers are forthcoming", Shapiro and Stiglitz (1984). Ashenfelter (1978) provides a succinct analysis of involuntary unemployment that discusses the underlying choices, the implications for unemployment benefit and the importance of understanding the reasons for involuntary unemployment. These reasons are crucial as they will determine the effect of trade liberalisation on wages and set limits for the conclusions of this chapter. I review below some of these reasons to illustrate how the reaction of wages to events will differ depending on the specific reason for involuntary unemployment. (There is no attempt to provide an exhaustive review of the reasons for involuntary unemployment so, for example, the effects of trade unions on labour markets are not reviewed.)

---

[2] In particular, Venables (1997) shows that the Hecksher-Olin model suggests that trade liberalisation reduces the incentives for factor mobility (and will not create it where there was none before). Thus the European Union policy of opening trade with Eastern European countries could have reduced the differential in factor prices sufficiently to stop the flow of immigrants into countries in the European Union. On the other hand, in the context of the specific-factor model trade liberalisation can increase the return on the mobile factor, e.g. labour, the mass migrations to the US.

[3] Nickell (1990) points out that the main implication of imperfect competition in the product market is to open up the possibility of nominal inertia in price setting. Nominal inertia refers to a sluggish price response to other nominal changes, particularly changes in wage costs. It arises from the fixed costs of adjustment.

One class of reasons for involuntary unemployment refers to efficiency wages. This means that firms set wages and that for some reason they prefer to pay a wage in excess of the wage that equates supply and demand. The rationale is simple: "the firm is not interested in choosing the minimum wage at which demand is satisfied but, rather, in choosing the wage which minimises its cost per efficiency unit of labour", Weiss (1980).

There are various types of efficiency wage explanations. One of them could be the possibility of shirking. This simply recognises that firms are unable to monitor workers costlessly and perfectly. In this case, paying a higher wage increases the cost of losing the job and reduces shirking and hence employment, Shapiro and Stiglitz (1984). They believe that shirking is "a significant factor in the observed level of unemployment, especially in lower-paid, lower skilled, blue-collar occupations." Jaffe et al. (1995) suggests that this may well be the case of pollution-intensive industries. Another efficiency wage explanation could be a situation where firms' have limited information about workers' ability and reservation wages are an increase function of their productivity. Wages have then a sorting effect. Weiss (1980) shows that in this case, paying wages in excess of the market-clearing wage enables the firm to hire from a "better" pool of applicants.[4]

A different class of reasons for involuntary unemployment refers to the effect of information asymmetries on the type of labour contracts offered. For example, it could be that the response to information asymmetries in labour markets is to offer (second-best) contracts that equalise worker's utility across states or across agents, Artis and Sinclair (1996). The different implications of these classes of reasons for unemployment is illustrated in Artis and Sinclair (1996), which compares the case for (revenue-neutral) employment subsidies. They find that if unemployment is explained by efficiency wages as in Shapiro and Stiglitz (1984), then an employment subsidy – a reduction in the cost of employment – will lower unemployment when product demand is inelastic. However, if employment is the result of a second-best contract that equalises utility, then an employment subsidy has no effect.

---

[4] Other efficiency wage explanations include firms' preference to reduce the costs of labour turnover (Salop (1979)) and firms' preference to build loyalty among workers and hence increase workers' effort (Akerlof (1982)).

It is not possible to be certain as to which of these possible explanations for voluntary unemployment will be relevant without focusing on a specific case of reduction in trade barriers and environmental taxation. This discussion nevertheless suggests, in practice, wages may not respond to the abolition of trade barriers and may remain close to the level of the autarky equilibrium.

4.3 Model's assumptions

The crucial assumption made to extend the model is that firms are wage takers. The remaining assumptions are as stated in Ulph (1995) and in Chapter 3. Briefly, these are:

- firms have to commit themselves to enter the industry before the government announces the level of the environmental tax; the structure of the game remains therefore as follows: firms decide to enter (stage 0), governments set environmental taxes (stage 1), firms decide then whether to be operative (stage 2), and finally, firms set their outputs (stage 3);
- production damages the local environment; it is assumed that the marginal damage is linear, i.e. the total damage is a quadratic function of production;
- capital moves freely between countries and firms are owned by residents of third countries;
- each government sets the environmental tax to maximise the country's welfare; this is defined as consumers' surplus plus the revenue from the environmental tax less the pollution damage;
- firms produce under constant returns to scale and set their outputs without taking account of the reaction of their competitors both within the country and (after the abolition of trade barriers) across countries;
- after the abolition of trade barriers, product markets are integrated and governments set the environmental tax independently; in this chapter, I have assumed that they set it non-cooperatively.

The rest of this section presents the additional assumptions made to extend the model used in Chapter 3.

4.3.1 Supply of labour

An individual's utility is a function of the consumption of widgets, x, the number of hours of leisure time, T, and the expenditure in other goods, z. To compare the results obtained hereafter with those obtained in Chapter 3 and to keep the analysis as simple as possible, it is assumed that the utility function is separable. This is a standard assumption in this type of analysis, see for example, Shapiro and Stiglitz (1984). The implication of this assumption is that the individual demand for widgets and the supply of labour are independent in the sense that the price of a widget, p, is not one of the arguments of the labour supply function. The equilibrium level of wages will nevertheless be affected by the equilibrium price in the market for widgets.

We assume that the utility function takes the following form:

$$U(x,T;z) = (ax - \frac{1}{2}x^2) + (bT - \frac{k}{2}T^2) + z \qquad (1)$$

This means that the supply of labour will be linear, k (a positive number) is the slope of the individual's supply function. This is clearly a simplifying assumption since "whatever the answer, [the supply of labour] is unlikely to be linear", Blundell (1992). It is nevertheless a convenient approach to derive qualitative results and intuitions about the equilibrium hence its widespread use in industrial economics.[5] I will also explore in Section 4.7 how the results change when labour supply is fixed.

Individuals are identical and each individual has an endowment of H hours that can be dedicated to work or leisure. She decides to work h hours ($h < H$) so that the hours of leisure available, T, can be expressed as $H - h$. I can therefore express the individual's utility above as a function of the hours worked and the time endowment as follows

$$U(x,h;z;H) = (ax - \frac{1}{2}x^2) + (b(H-h) - \frac{k}{2}(H-h)^2) + z \qquad (2)$$

An individual obtains her income solely from work so that the budget constraint is now

$$z + p\,x = wh \qquad (3)$$

A change in wages generates a substitution effect and an income effect. The substitution effect means that if wages increase, the opportunity cost of not working increases and the individuals spend more time at work. The income effect goes in the opposite direction. It means that an increase in wages will increase incomes and hence the consumption of normal goods. Leisure (or not working) is typically a normal good for most individuals

[5] Awareness about the limitation of the linearity assumption is not new. For example, Bertrand already criticised the extensive use of linear functions (Martin (2002), preface).

so the income effect means that the result of higher wages is less time at work. Overall, the amount of labour supplied will increase with an increase in wages if the substitution effect dominates the income effect. It is usually assumed that when wages are low the substitution effect will dominate; as wages increase, the income effect will eventually dominate hence the standard result of a (non-linear) backward-bending supply function. The implication of the assumptions in this chapter is that there are no income effects. This means that a result that the equilibrium after trade liberalisation will be characterised by a substantive increase in outputs ought to be qualified – but perhaps no more than a result that assumed that unit labour costs are fixed.

Each individual maximises the utility function in equation 2 above by choosing the amount of widgets to consume, *x*, and the number of hours of work, *h*, subject to the above budget constraint being satisfied. The first-order conditions yield the individual's demand for widgets obtained in Chapter 3 and the individual's supply of labour. The latter is

$$w = (b - kH) + kh \tag{4}$$

where k is positive and represents the slope of the individual's supply of labour and $b - kH$ is positive and represents the individual's reservation wage – the minimum wage she is willing to accept to forego one hour of leisure.

Figure 1: linear labour supply functions

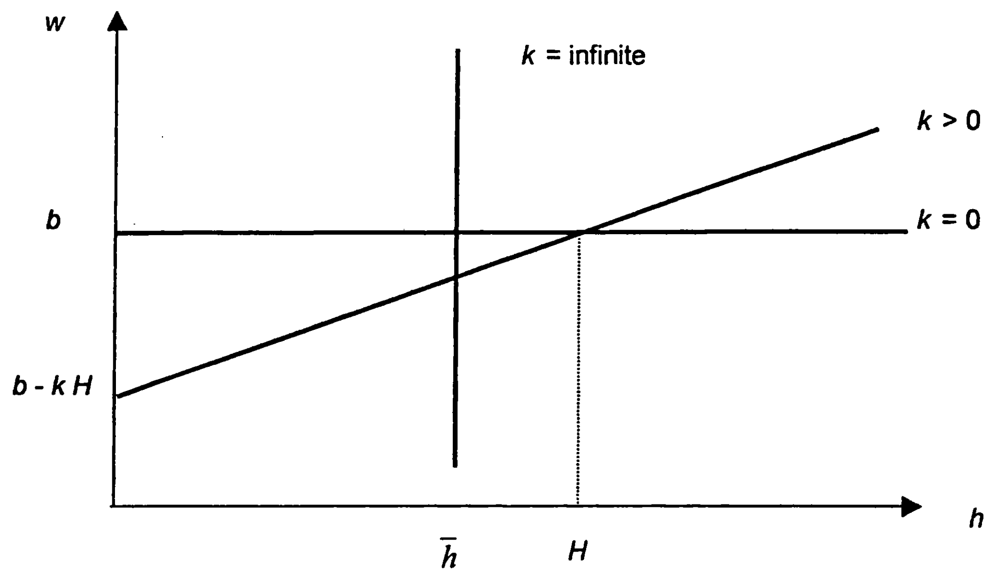


Figure 1 shows equation 4 – the labour supply that assumes $k > 0$. It also shows two other labour supply functions nested within it. A horizontal labour supply represents a case where $k = 0$ and the assumption in Chapter 3 that wages are fixed. A vertical labour supply at $\bar{h}$ means that there a fixed supply of labour. It represents a case where k is infinite and will be introduced in Section 4.7.

For a country with m consumers, the aggregate supply of labour is therefore

$$L_S = m(H'+k'w) \qquad for\ k'w > |H'| \tag{5}$$

where $k' = \frac{1}{k}$ and $H' = \frac{kH-b}{k} < 0$. (Note that increasing k' in equation 5 means reducing k in equation 4.)

4.3.2 Demand for labour and equilibrium wages

The demand for labour is partly determined by firms' production function so I start by specifying it. The model used in Chapter 3 assumed that firms produce under conditions of constant returns to scale in both countries, i.e. constant unit labour costs, c. This means that there is a fixed requirement of labour to produce one widget that is independent of the number of widgets produced. Let us assume that l hours of work are needed to produce one widget and that L hours of work are available to a firm. A firm's production function, the number of widgets produced, y, is then

$$y = L / l \tag{6}$$

Given the cost of each hour of work, w, the unit labour costs are

$$c = w\,l \tag{7}$$

The unit labour costs may therefore differ between countries because wages, w, are different or because the productivity of labour or the number of hours needed to produce one widget, l, is different. The latter is determined by the technology used, which is assumed to be the same in both countries as a result of the free mobility of capital. Thus, differences in unit labour costs reflect differences in wages.

Firms' demand for labour can be derived from the firm's profit maximisation problem. The profit of a firm in country 1 is

$$\pi_1 = P(Y_1)y_1 - y_1(w_1 l + t_1) \tag{8}$$

where Y_1 is the total number of widgets produced in the country and equals $N_1 y_1$, and $P(Y_1)$ is the inverse aggregate demand function ($P(Y_1, Y_2)$, after abolishing trade barriers). The first-order condition is

$$\frac{\partial \pi_1}{\partial y_1} = MR_1 - (w_1 l + t_1) = 0 \tag{9}$$

Re-arranging this, I obtain that a firm's demand for labour in country 1 will be

$$w_1 = \frac{1}{l}(MR_1 - t_1) \tag{10}$$

where $1/l$ is the marginal product of labour.[6] This is the standard expression for the equilibrium wage in a labour market: the equality of the value of the marginal product of labour and marginal revenue. The change here is that the environmental tax is netted from the marginal revenue. The abolition of trade barriers and the integration of product markets affects the relevant demand that firms face. This affects the marginal revenue and the environmental tax so that the demand for labour and the equilibrium wage will also be affected.

4.4 The symmetric equilibrium

This section characterises the symmetric equilibrium to derive the basic intuition when wages are endogenous and the labour supply function is linear.

4.4.1 The autarky equilibrium

The autarky equilibrium represents the equilibrium in each country before trade barriers are abolished and there is no trade between the countries. Given a wage and the assumptions about the labour market, the aggregate output in a standard Cournot oligopoly with N identical firms and m consumers is

$$Y = \frac{Nm}{(N+1)}[a - wl - t] \tag{11}$$

where t is the environmental tax. Using equation 6 above, I obtain firms' aggregate demand for labour

---

[6] From equation 6, $MP_L \equiv \frac{\partial y}{\partial L} = \frac{1}{l}$.

$$L_D = \frac{lNm}{(N+1)}[a - wl - t] \qquad (12)$$

An equilibrium in the labour market is characterised by a level of wages such that the demand for labour, equation 12, and the supply of labour, equation 5, are the same. This is

$$m(H'+k'w) = \frac{lNm}{(N+1)}[a - wl - t] \qquad (13)$$

The government sets taxes as described in Chapter 3 (summarised at the beginning of section 4.1) and firms do not behave strategically in the labour market. Thus the equilibrium level of the environmental tax for a given wage can be taken from Ulph (1995). This is

$$t = \frac{(a - wl)(1 + dmN)}{(N + 2 + dmN)} \qquad (14)$$

The equilibrium wage is obtained from equation 13 above by substituting the environmental tax, equation 14, and re-arranging. This gives

$$w = \frac{Nal - (N + 2 + dmN)H'}{k'(N + 2 + dmN) + Nl^2} \qquad (15)$$

This says that, other things being equal, the equilibrium wage will increase with the reservation wage (H', which is negative). Substituting the equilibrium wage back into equation 14 yields the equilibrium level of environmental tax

$$t = \frac{(ak'+lH')(1 + dmN)}{k'(N + 2 + dmN) + Nl^2} \qquad (16)$$

The environmental tax will be lower the higher the reservation wage in the country.

Operating profits per firm in this country are

$$\pi = m\left(\frac{a - wl - t}{N + 1}\right)^2 \qquad (17)$$

Substituting the equilibrium wage and the environmental tax yields

$$\pi = m\left(\frac{ak'+H'l}{k'(N + 2 + dmN) + Nl^2}\right)^2 \qquad (18)$$

Under the assumption of free movement of capital, new firms enter the market as long as operating profits cover the sunk costs, S. The market structure, as described by the number of firms, will then be

$$N = \frac{(ak'+H'l)\sqrt{\frac{m}{S}} - 2k'}{(k'+k'dm+l^2)} \geq 1 \tag{19}$$

Comparing this expression with the optimal industry size derived in Chapter 3 when unit labour costs were fixed (equation 12 in page 78) reveals two sources of differences: the inclusion of k' and of l. Annex 4.2 shows that the size of the industry is a decreasing function of the number of hours required to produce a widget, l. The annex also shows that the size of the industry is also a decreasing function of k (the inverse of k').

Result 1: relaxing the assumption that unit labour costs are fixed affects the optimal industry size in two different ways with opposing effects so the overall effect is ambiguous.

Equation 19 implies that the condition for the existence of an autarky equilibrium is

$$(ak'+H'l)\sqrt{\frac{m}{S}} > 3k'+dmk'+l^2 \tag{20}$$

4.4.2 The post integration equilibrium

The post integration equilibrium represents the situation after the abolition of trade barriers. As in Chapter 3, it is assumed that it results in the integration of product markets. Firms in the two countries will therefore face the same product demand, which is different from the one that they faced in the autarky equilibrium. Their decisions about production will therefore change. As governments are not committed to the level of the environmental tax, this will also change as a result.

On the assumption that wages are endogenous and that individual firms take wages as given and choose their outputs in a Cournot fashion, the aggregate output of firms based in country 1 (Ulph (1995)) can be written as

$$Y_1 = \frac{N_1(m_1+m_2)}{(N_1+N_2+1)}[a + N_2(w_2 l + t_2) - (N_2+1)(w_1 l + t_1)] \tag{21}$$

Assuming that the two countries are identical (same population, $m_1 = m_2$, same damage to the environment, $d_1 = d_2$, wages, $w_1 = w_2$ and environmental taxes, $t_1 = t_2$), the aggregate output is

$$Y = \frac{2Nm}{2N+1}[a - wl - t] \quad (22)$$

Note that the number of firms in this expression is that from the autarky equilibrium. The demand for labour can be derived from this expression using equation 6 above. This is

$$L_D = \frac{2lNm}{2N+1}[a - wl - t] \quad (23)$$

For a given wage level, the environmental tax is (Ulph (1995))

$$t = \frac{(a - wl)(1 + 2dmN)}{(2N + 2 + 2dmN)} \quad (24)$$

The equilibrium wage is the one that equates demand and supply of labour, equations 23 and 5. This is

$$m(H' + k'w) = \frac{2lNm}{2N+1}[a - wl - t] \quad (25)$$

Substituting the equilibrium tax, equation 24, and re-arranging, I obtain the equilibrium level of wages after the integration of product markets

$$w = \frac{aNl - (N + 1 + dmN)H'}{k'(N + 1 + dmN) + Nl^2} \quad (26)$$

This can be substituted again into equation 24 to obtain an expression for the equilibrium level of tax that is only dependent on exogenous factors.

$$t = \frac{(ak' + H'l)(1 + 2dmN)}{2k'(N + 1 + dmN) + 2Nl^2} \quad (27)$$

Contingent on the number of firms and wages, the operating profits for each firm are (Ulph (1995))

$$\pi = 2m\left(\frac{a - wl}{2 + 2N + 2Ndm}\right)^2 \quad (28)$$

Substituting the equilibrium wage level gives

$$\pi = \frac{m}{2}\left(\frac{ak' + H'l}{k'(N + 1 + Ndm) + Nl^2}\right)^2 \quad (29)$$

Capital moves freely so new entry takes place if the profits exceed sunk costs, S, given the number of firms from the autarky equilibrium. Substituting the number of firms, equation 19, and re-arranging gives the requirement for new entry post integration. This is

$$(ak'+H'l)\sqrt{\frac{m}{S}} < k'\frac{\sqrt{2}}{\sqrt{2}-1} \qquad (30)$$

Bringing together the two entry conditions 20 and 30, I obtain that the condition for new entry in a symmetric equilibrium is now

$$\frac{\sqrt{2}}{\sqrt{2}-1}k' > (ak'+H'l)\sqrt{\frac{m}{S}} > 3k'+dmk'+l^2 \qquad (31)$$

This requires inter-alia that $0.415k' > dmk'+l^2$.

4.3.3 Discussion of the symmetric equilibrium

Three issues about this symmetric equilibrium are explored further: the effect on equilibrium wages, environmental taxes and the conditions for new entry.

The effect on wages of abolishing trade barriers can be seen by comparing the equilibrium level of wages before and after, equations 15 and 26 above. The proposition that wages in the autarky equilibrium exceed wages in the post integration equilibrium results in

$$ak'+H'l < 0 \qquad (32)$$

The condition for the existence of the autarky equilibrium, equation 20, implies, however, that in equilibrium the above expression is positive. Wages will therefore be higher in the post integration equilibrium. Given that the supply of labour is not affected by the abolition of trade barriers, this result implies that firms' demand for labour increases in the post integration equilibrium. This is confirmed by comparing equations 12 and 23. This also suggests that the number of hours worked will increase.

Result 2: the level of equilibrium wages will be higher after the abolition of trade barriers.

Result 3: incomes from work, the product of the number of hours worked and wages, will increase following the abolition of trade barriers.

The effect on environmental taxes can be seen by comparing the level of the environmental tax before and after the removal of trade barriers, equations 16 and 27. One expects the abolition of trade barriers to result in lower environmental taxes. This is because the increase in market size associated with the integration of product markets weakens the distortion caused by Cournot's behaviour. The proposition that abolishing trade barriers results in lower environmental taxes than in the autarky equilibrium implies that

$$k'dm < k'+l^2 \tag{33}$$

The analysis of entry conditions in equation 31 suggested that an equilibrium will always satisfy that $0.415k'-l^2 > dmk'$. This in turn suggests that taxes will be lower after the abolition of trade barriers if it can be shown that

$$k'+l^2 > 0.415k'-l^2 \tag{34}$$

This implies that

$$2l^2 > -0.585k' \tag{35}$$

This condition will always be satisfied given that both l and k' are positive.

Result 4: environmental taxes will be lower after the abolition of trade barriers.

Regarding the entry conditions, equations 20 and 30, it is interesting to explore the effect of the flexibility in unit labour costs introduced, i.e. whether entry in the post integration equilibrium is more likely when wages are endogenous. Equation 29 suggests that in the post integration equilibrium, profits will be larger the smaller the number of firms in the autarky equilibrium. However, the effect of the assumptions about labour markets made here on the optimal industry size in autarky is ambiguous (Result 1). The upshot of these two results is that the effect of assuming that wages are endogenous on the likelihood of entry after the abolition of trade barriers is ambiguous.

Result 5: relative to the assumption that wages are fixed, the effect of the assumptions about labour markets made here on entry in the post integration equilibrium is ambiguous.

4.5 The asymmetric equilibrium

The autarky equilibrium in Section 4.4.1 also describes the equilibrium when the two countries are different. In this section, the focus is the post integration equilibrium and, in particular, the wage differential between countries. As indicated above, relaxing the assumption that unit labour costs are exogenous affects the market equilibrium, stage 3 of the game where outputs are set.

The first issue is whether wages will be different in the two countries. Samuelson (1949) has shown that, under certain conditions, integration of product markets results in the integration of factor markets (capital and labour). These conditions include perfect competition, no transport costs and identical production technologies. If these conditions are met, factors obtain the same remuneration in the two countries even if the factors are not mobile. In this chapter, it is assumed that capital moves freely between countries thus earning the same rate of return. Labour is not mobile so the issue is whether integration of product markets together with the (assumed) integration of capital markets can result in the integration of labour markets so that wages in the two countries are the same.

An equilibrium level of wages must satisfy the profit maximising condition in equation 10 above. This is

$$w_1 = \frac{1}{l}(MR_1 - t_1) \tag{36}$$

Thus, wage equalisation requires that

$$MR_1 - MR_2 = t_1 - t_2 \tag{37}$$

I assumed that the abolition of trade barriers results in product markets integration so that the aggregate demand is

$$p = a - \frac{N_1 y_1 + N_2 y_2}{m_1 + m_2} \tag{38}$$

where y_1 and y_2 are firms' outputs. Assuming that producers behave in a Cournot fashion, the marginal revenue for a firm in country 1 is

$$MR_1 = p - \frac{y_1}{m_1 + m_2} \tag{39}$$

Thus wage equalisation requires

$$\frac{y_2 - y_1}{m_1 + m_2} = t_1 - t_2 \tag{40}$$

This is a very specific requirement that will not be satisfied when each government sets taxes individually (whether cooperatively or non-cooperatively).

Result 6: wages will not be the same in the two countries when governments set environmental taxes individually.

This result suggests, however, that if the two governments agree to set a common environmental tax, $t_1 = t_2$, firms' outputs in the two countries will be identical and so wages. Absent that, nothing in the model has changed to suggest that the differential in taxes will be as required by equation 40. The intuition for this result is as follows. Assuming that wages are endogenous makes the conditions similar to those in models of international trade. Here, firms have to enter the industry before the government sets the level of tax so there is a factor that is specific to the sector in the post integration equilibrium: the capital sunk or the number of firms. This is therefore similar to the standard specific-factor model so one should not expect wages to be the same in the two countries.

Given that wages will not be the same in the two countries, I want to characterise the wage differential in the post integration equilibrium. Using equation 39 above, the profit maximising condition in equation 36 above can be written as[7]

$$lw_1 = p - \frac{y_1}{m_1 + m_2} - t_1 \tag{41}$$

This can be re-arranged as

$$y_1 = (m_1 + m_2)(p - lw_1 - t_1) \tag{42}$$

Recall that y_i is the output of a firm located in country *i*. The demand for labour in country 1 can be derived from this using the production function, equation 6. This gives

$$L_{D1} = lN_1(m_1 + m_2)(p - lw_1 - t_1) \tag{43}$$

The equilibrium level of wages in each country is the one that equates the local supply of labour, equation 5, and the demand above. This is

$$w_1 = \frac{lN_1(m_1 + m_2)(p - t_1) - m_1 H'}{k' m_1 + N_1 l^2 (m_1 + m_2)} \tag{44}$$

---

[7] We are interested in the wage differential. The equilibrium price is included because it is common to both countries and will therefore not affect the direction of the wage differential. The level of tax is also included as if it were an exogenous variable because it is not possible to obtain a simple expression of it when the two countries are not identical, on this see Section 3.5.1 in Chapter 3.

It can be shown that the proposition that wages in country 1 exceed wages in country 2 requires that the following is satisfied

$$(pk'+H'l)(N_1m_2 - m_1N_2) + l^2N_1N_2(m_1 + m_2)(t_2 - t_1) + k'(t_2m_1N_2 - t_1m_2N_1) > 0 \tag{45}$$

Note that although H' is negative, the equilibrium price must be larger than the unit labour cost for the reservation wage, $H'l$. There are then two cases depending on whether the slope of the labour supply, k', is such that $pk'+H'l$ is positive or negative. This expression is positive when $k' \geq 1$ (i.e. $k \leq 1$). A necessary condition for equation 45 to be satisfied is[8]

$$\frac{N_1}{N_2} > \frac{m_1}{m_2} > \frac{t_1}{t_2} \tag{46}$$

This, however, is not completely satisfactory as it depends on taxes which are endogenous (in the post integration equilibrium the number of firms is exogenous). A necessary condition that only depends on the parameters of the model is

$$\frac{N_1}{N_2} > \frac{m_1}{m_2} \tag{47}$$

The intuition behind this necessary condition can be verified by considering the case where the two countries have identical populations and supply of labour and they differ only in the valuation of the environment. The above necessary condition suggests that wages will be higher in the country with the larger number of firms. Equation 19 implies that this will be the country with the lower valuation of the environment. An alternative intuition can be derived from re-arranging equation 47 as a ratio N / m. Given that the number of firms, N, represents the (sunk) capital invested and m the population, this ratio will also represent country's capital intensity (ratio of capital to labour). So a necessary condition for wages to be higher in a country is that capital intensity is higher than in the other country.

---

[8] For the first term in equation 45 to be positive, it is required that $\frac{N_1}{N_2} > \frac{m_1}{m_2}$; for the second term $\frac{t_1}{t_2} < 1$; and for the third $\frac{m_1}{m_2} > \frac{t_1}{t_2}\frac{N_1}{N_2}$. These three conditions together suggest equation 46.

If, however, $pk'+H'l$ is negative (likely to be when $k' < 1$ and $k > 1$) then the necessary condition for satisfying equation 45 is

$$\frac{m_1}{m_2} > \frac{N_1}{N_2} \tag{48}$$

Result 7: whether wages will be higher in one country depends on the slope of the labour supply; if, for example, the slope of the labour supply is larger than 1, then a necessary condition for wages to be higher in a country is that it has a larger number of firms per capita.

4.6 Simulations

This section uses numerical simulations to explore the post integration equilibrium when the two countries have different populations and when they have different valuations of the marginal damage to the environment. The simulation model is very similar to the one used in Chapter 3. The only addition is the specification of a labour market in the market stage – based on equations 13 and 25 above. The model assumes that the environmental tax is set non-cooperatively in the post integration equilibrium. Briefly, the model works as follows. First, the autarky equilibrium is solved; the equilibrium number of firms, environmental tax and wage level are determined. Then trade barriers are abolished and the model solves the post integration equilibrium based on the number of firms in the autarky equilibrium.

Finally, I must specify the value for the slope of the supply function for labour, k', that will be used in the simulation. Here there are various issues that have to be taken into account. First, George Stigler's dictum that all elasticities are 1 in absolute value (as reported, for example, in Heckman (1993)). Second, the empirical evidence about the supply of labour, see, for example, Heckman (1993), Blundell (1992) and Blundell and MaCurdy (1999). Briefly, this suggests that response to changes in wages will vary between males and females; it is generally larger for females. This empirical evidence also suggests that the change in number of hours worked is larger when account is taken of entry and exit decisions into the labour force. The model developed in this chapter is aimed at exploring the robustness of the conclusions in Chapter 3, which assumed fixed unit labour costs. There is no allowance for differences in gender or participation decision. I use an arbitrary value of k in the simulations, 1, and then consider separately (Section 4.7) the results if there is a fixed supply of labour. Thus, I obtain results for three values of k: zero (Chapter 3 – fixed unit labour costs), 1 (the analysis here and the simulations) and infinite (Section 4.7 – fixed supply of labour).

4.6.1 Population asymmetries

The two countries have different populations and, hence, workforces. I have defined four scenarios, A, B, C, and D, where the population increases from A to D. The data for these scenarios and the results of the simulations are summarised in Table A4.1 of Annex 4.3. Section a. contains the data of the base case. Sections b. and c. consider cases where the population in one country is twice and three times the population in the other country respectively. Each of these sections contains two panels. The first panel presents the results of the autarky equilibrium. The second one presents the results of the post integration equilibrium.

The results suggest that in the autarky equilibrium the size of the industry will be larger in the country with the larger population. Wages will be higher in that country which suggests that although the supply of labour is larger, the demand for labour is proportionally larger.

The integration of product markets results in the small country exporting to the large country. Thus, profits exceed sunk costs in the small country and there is new entry. This is consistent with the idea of minimising the overall damage to the environment. As the damage to the environment increases with the level of output in a quadratic form, it is more efficient to allocate additional production to the small country. This seems counter-intuitive given that the countries are setting the environmental tax non-cooperatively. However, the equilibrium price is the same in the two countries and this will be lower when environmental taxes are lower. Thus, both countries will benefit if the country where pollution causes a lower environmental damage carries out some of the additional production. The results also suggest that environmental taxes are lower in both countries in the post integration equilibrium. This is consistent with the results from the symmetric equilibrium (Result 4 above) and with the results obtained when wages were exogenous.

Comparing wages in the autarky and the post integration equilibrium for a given country suggests that in the small country, the abolition of trade barriers increases wages and incomes. It is unclear, however, whether wages will increase in the large country. This seems to depend partly on the proportional difference in size between the two countries.

I compare now wages in the two countries in each of the two states (autarky and post integration). In the autarky equilibrium, wages were higher in the country with the larger population. In the post integration equilibrium, however, wages are higher in the country with a small population. The necessary condition identified in Section 4.5, equation 47, is satisfied.[9] This illustrates that the integration of product markets can have a powerful effect on the wage differential between countries and even invert it as in this case. The economic intuition of this change is as follows. Integration creates additional demand for the product. Given the assumption that damage to the environment is a quadratic function of output and given that the environmental tax is related to output, the small country is in a better position to benefit from the excess demand for the product and export it. This additional demand for the product is sufficient to result in wages in the small country being now higher than in the other country.

4.6.2 Marginal damage asymmetries

Production results in different marginal damages to the environment in the two countries. I have defined four scenarios, A, B, C and D, where the marginal damage increases from A to D. The data for these scenarios and the results of the simulations are summarised in Table A4.2 of Annex 4.3.

A higher valuation of the environment, which is reflected in higher marginal damage to the environment, means that increased production will result in a larger reduction in social welfare, other things being equal. Thus, the tax will be higher in this country and wages smaller. The results of the simulations confirm this intuition. They also suggest that in the autarky equilibrium, consumption and industry size are also smaller in the country where production causes a higher damage to the environment.

---

[9] The ratio of population ($m_1 \div m_2$) is 0.5 or 0.3; and the number of firms in country 1 (the small country) is more than half or a third of the number of firms in the other country.

Following the abolition of trade barriers, it is the country where production causes a lower damage to the environment that exports and where there is new entry. The intuition for this result is the same that was described in Section 4.6.1 for the small country. Environmental taxes are also lower across the board in the post integration equilibrium because the output distortion is reduced as a result of the increase in the size of the product market.

I compare wages before and after the abolition of trade barriers for a given country. This shows that the additional product demand generated by the abolition of trade barriers increases wages and incomes. I compare now wages in the two countries in the post integration equilibrium. The results suggest that wages will be higher in the country where production causes a lower damage to the environment.

The simulations also suggest that the wage differential between the two countries increases with the abolition of trade barriers. The two countries have the same population and they have the same supply of labour. The above change implies that the proportional increase in the local demand for labour is different in the two countries.

4.7 Qualifications

The preceding discussion in this chapter is conditioned by the assumptions of the model and it seems appropriate to discuss some of the key assumptions and how they affect the results.

First, the results are conditioned by the assumptions about the consumers' utility function, which lead to linear product demand and labour supply that are independent. This seems appropriate for a simple check of the robustness of the conclusions in Chapter 3. There are, however, alternative assumptions about labour supply that could be made, see, for example, Stern (1984) for a review of the possibilities. It seems that there may be scope for extending this work to understand the effect of specific cases of reductions of trade barriers using other utility functions. Of the various utility functions in Stern (1984), one that seems intuitively interesting in this context is the Stone-Geary utility function because it allows the specification of a 'minimum' or 'subsistence' level for consumption and labour supply. So, for example, utility could take the following form

$$u(x,T) = (x-\bar{x})^{\alpha_1}(T-h)^{\alpha_2} \qquad (49)$$

where $\bar{x}$ is minimum consumption, T is the maximum labour supply and $\alpha_1 + \alpha_2 = 1$. One of the advantages of this function is that it nests other cases, for example, the Cobb-Douglas system of preferences is obtained setting the 'minimum' levels to zero. Dertouzos and Pencavel (1986) use this functional form to specify a trade union's objective function and show how, by a suitable choice of parameters the model, can nest various other models of union's behaviour discussed in the literature.

Stern (1984) also shows that by taking logarithm, the utility function in equation 49 can be transformed, so that the resulting supply function for labour becomes vertical when the number of hours supplied is close to the maximum labour supply, T. Whether the equilibrium is characterised by the vertical part of the labour supply function depends on the relative size of the supply and demand for labour. The model presented here can be used to explore the effect of abolishing trade barriers if the labour market can be characterised by a fixed supply of labour (and product demand is unaffected). Annex 4.1 shows that in this case, output will be determined by the supply of labour – population – and will not be affected by product demand or government's intervention. These will only affect prices and wages. In this case, the abolition of trade barriers does not change the total supply of labour and the additional demand for widgets created by integration is choked-off by higher wages. The output in each country will be the same as before integration and there is no new entry.

Another assumption in the paper is that population and hence labour is immobile after the abolition of trade barriers and that there is no migration between countries. Assuming that there is perfect mobility of labour after the abolition of trade barriers will change drastically the analysis in two different ways. First, if labour is perfectly mobile, it is as if there is one market for labour and wages will be the same in the two countries. Second, environmental policy will be affected. In its simplest form, the results derived here depend on the population in each country. More generally, assuming that population is mobile can have wider implications for environmental policy. For example, one of the standard results is that if pollution is transboundary, efficient environmental policies will require coordination between countries (by means of treaties or a third party authority). Hoel and Shapiro (2001) show that if there is perfect mobility of population and governments set environmental taxes non-cooperatively the efficient outcome is also Nash equilibrium. However, there may be other Nash equilibria. Thus the need for external coordination changes – from explicit treaties or mandates to a third party to some mechanism for ensuring that the efficient equilibrium is the one achieved.

4.8 Conclusions

I have extended the analysis in Chapter 3 by relaxing the assumption that unit labour costs are fixed. I assume that wages are endogenous and use a very simple model of the labour market – no imperfections, separable utility function, linear supply of labour and no income effects. Of these assumptions, perhaps the crucial one is that there are no market imperfections in the labour market. It means that there are no impediments for the adjustment of wages when trade barriers are abolished. These impediments could arise from various causes such as efficiency wages. In this case, it is as if wages and unit labour costs remained fixed as discussed in Chapter 3.

Using the assumptions made in this chapter, I observe that the environmental taxes will be lower after the abolition of trade barriers because the associated increase in market size reduces the output distortion caused by the oligopolistic behaviour of existing firms. Abolishing trade barriers increases the demand for the product faced by a firm and the demand for labour so that wages tend to be higher.

The analysis also suggests that wages will remain different in the two countries as a result of government intervention – setting the environmental tax. We characterise the wage differential between countries in the post integration equilibrium in the asymmetric equilibrium assuming linear supply functions. This suggests that wages will be higher in the country with the larger number of firms per capita if labour supply is sufficiently elastic. This has been verified by simulating the equilibrium results. The simulations suggest that this will be the country with a lower valuation of the environmental damage or the country with a smaller population.

All these results depend on the functional forms used for the utility function and there are various alternative forms. A feature of an alternative utility function, Stone-Geary, is that it can result in a labour supply that is vertical when the number of hours is close to the maximum number of hours available. I have explored how the results may change if the equilibrium is characterised by such tight conditions in the labour market. For this purpose, I have assumed that the labour supply is fixed. In this case, the abolition of trade barriers does not change the output and there is no new entry.

Annex 4.1: Fixed supply of labour

The text assumes that individuals decide how much time to work as part of their utility maximisation problem so that the supply of labour is flexible. An alternative assumption is that individuals work a fixed amount of time, $\bar{h}$. The aggregate supply of labour is then

$$L^S = m\bar{h} \tag{A1}$$

where m is the number of consumers in the country.

The autarky equilibrium

Firms' demand for labour is not affected by the assumptions about the supply of labour so we can use equation 12. The equilibrium in the labour market is a level of wages such that the demand for labour and the supply of labour, equation A1, are the same. This is

$$m\bar{h} = \frac{lmN}{(N+1)}[a - wl - t] \tag{A2}$$

The equilibrium wage in the autarky equilibrium is then

$$w = \frac{1}{l}\left(a - t - \bar{h}\frac{(N+1)}{lN}\right) \tag{A3}$$

Combining the expression above with the equilibrium level of taxes, equation 14 from the main text, and re-arranging yields

$$w = \frac{1}{l}\left(a - \bar{h}\frac{N+2+dmN}{lN}\right) \tag{A4}$$

$$t = \frac{\bar{h}(1+dmN)}{lN} \tag{A5}$$

Substituting now the equilibrium wage level, equation A4, and the environmental tax, equation A5, into the expression for the country's total output, equation 11 from the main text, gives after some re-arranging

$$Y = \frac{\bar{h}m}{l} = \frac{L^S}{l} \tag{A6}$$

Result A1: prices and environmental taxes do not affect the equilibrium output; the supply of labour and the marginal productivity of labour determine it.

Given a wage level, a firm's operating profits in this Cournot's oligopoly are

$$\pi = m\left(\frac{a - wl - t}{N+1}\right)^2 \quad \text{(A7)}$$

Substituting the equilibrium level of wages, equation A4, and the environmental tax, equation A5, gives

$$\pi = m\left(\frac{\bar{h}}{lN}\right)^2 \quad \text{(A8)}$$

There is free movement of capital so that a new firm will enter the market if profits are sufficient to cover sunk costs. The market structure, as described by the number of firms, is then

$$N = \frac{\bar{h}}{l}\sqrt{\frac{m}{S}} \geq 1 \quad \text{(A9)}$$

Result A2: the market structure in the autarky equilibrium is independent of the size of the product demand and of the marginal damage to the environment; demand affects the market structure only through the size of the population.

Post-integration equilibrium

We follow the same logic as in Section 4.4.2. The equilibrium level of wages is determined from the following equilibrium condition

$$m\bar{h} = \frac{2lNm}{2N+1}[a - wl - t] \quad \text{(A10)}$$

Solving the above for wages and re-arranging as in the previous section gives

$$w = \frac{1}{l}\left(a - \frac{\bar{h}}{l}\frac{N+1+dmN}{N}\right) \quad \text{(A11)}$$

$$t = \frac{\bar{h}}{l}\frac{(1+2dmN)}{2N} \quad \text{(A12)}$$

Substituting these into the equilibrium level of output, equation 22 of the text, yields

$$Y = \frac{\bar{h}m}{l} = \frac{L_S}{l} \quad \text{(A13)}$$

Result A3: the equilibrium level of output is not affected by integration.

Comparing the equilibrium level of wages before and after the abolition of trade barriers, equations A4 and A11, gives the following.

Result A4: the equilibrium level of wages is higher after integration.

Substituting the equilibrium wage level and the environmental tax in the profit function, equation A7, gives

$$\pi(N) = \frac{m}{(N+1)^2}\left(\frac{\bar{h}}{l}\frac{2N+1}{2N}\right)^2 \qquad \text{(A14)}$$

where N is the number of firms in the autarky equilibrium, equation A9. Re-arranging gives

$$\pi(N) = m\left(\frac{\bar{h}}{lN}\right)^2\left(\frac{2N+1}{2(N+1)}\right)^2 \qquad \text{(A15)}$$

New entry post integration requires that operating profits cover sunk costs. The first two terms in the above expression is the profit in the autarky equilibrium, equation A8. Free entry means that, given the number of firms in the autarky equilibrium, profit is sufficient to cover sunk costs. Thus for the profits to exceed the sunk costs in the post integration equilibrium, the third term in equation A15

$$\frac{2N+1}{2(N+1)} > 1 \qquad \text{(A16)}$$

This will not be satisfied since the expression in the denominator is always greater than the expression in the numerator for positive values of N.

Result A5: there will be no new entry post integration.

The economic intuition underlying these results is as follows. The demand for the product and hence demand for labour increases as a result of the integration of product markets. Given that output is determined by the supply of labour, which is fixed and not affected by the abolition of trade barriers, the increased product demand must be choked off by higher wages (Result A4). Output then remains the same (Result A3); profits will be smaller and will not exceed sunk costs.

Annex 4.2: Comparative statics

Proposition 1: the number of firms in the autarky equilibrium, equation 19 in the text, is a decreasing function of l, the number of hours of labour required to produce one widget.

Proof: We need to show that $\frac{\partial N}{\partial l} \leq 0$.

$$N = \frac{(ak'+H'l)\sqrt{\frac{m}{S}} - 2k'}{(k'(1+dm)+l^2)} \geq 1$$

$$\frac{\partial N}{\partial l} = \frac{\left(H'\sqrt{\frac{m}{S}}\right)\left(k'(1+dm)+l^2\right) - 2l\left((ak'+H'l)\sqrt{\frac{m}{S}} - 2k'\right)}{(k'(1+dm)+l^2)^2} \leq 0$$

The denominator of this expression is always positive so it is sufficient to show that the numerator is positive.

Consider now the numerator. H' is negative so the first term in the numerator is negative. The second term in the numerator is positive: l is positive and $N \geq 1$ implies that

$$(ak'+H'l)\sqrt{\frac{m}{S}} - 2k' > 0$$

So the numerator is negative.

Thus the number of firms in the industry will be a decreasing function of the labour requirement.

Proposition 2: under certain conditions, the number of firms in the autarky equilibrium, equation 19 in the text, is an increasing function of the slope of the labour supply function, k.

Proof: We need to show that $\frac{\partial N}{\partial k} > 0$.

$$N = \frac{(ak'+H'l)\sqrt{\frac{m}{S}} - 2k'}{(k'(1+dm)+l^2)} \geq 1$$

Given that $k' = \frac{1}{k}$ and $H' = \frac{kH - b}{k} = H - \frac{b}{k}$, we can express N as

$$N = \frac{\left(\frac{a}{k} + Hl - \frac{bl}{k}\right)\sqrt{\frac{m}{S}} - \frac{2}{k}}{\left(\frac{1}{k}(1+dm) + l^2\right)} \geq 1$$

$$\frac{\partial N}{\partial k} = \frac{\left(\left(\frac{-a}{k^2} + \frac{bl}{k^2}\right)\sqrt{\frac{m}{S}} + \frac{2}{k^2}\right)\left(\frac{1}{k}(1+dm) + l^2\right) - \left(\left(\frac{a}{k} + Hl - \frac{bl}{k}\right)\sqrt{\frac{m}{S}} - \frac{2}{k}\right)\left(\frac{-1}{k^2}(1+dm)\right)}{\left(\frac{1}{k}(1+dm) + l^2\right)^2} > 0$$

The denominator of this expression is always positive so it is sufficient to show that the numerator is positive.

Consider now the numerator. After some re-arranging I obtain

$$\left(\frac{-a}{k^2} + \frac{bl}{k^2}\right) l^2 \sqrt{\frac{m}{S}} + \frac{2}{k^2} l^2 + \left(\frac{Hl}{k^2}\sqrt{\frac{m}{S}}\right)\frac{1}{k}(1+dm)$$

All the parameters are positive numbers so this expression will be positive unless a – the maximum willingness to pay for a unit of x – is very large in relation to all the other parameters.

Annex 4.3: Results of the simulations

Table A4.1: Different populations

a. base case assumptions

	Scenario A	Scenario B	Scenario C	Scenario D
Tastes (demand), a	3	3	3	3
Sunk costs, S	2	2	2	2
Labour requirement, l	0.25	0.25	0.25	0.25
Population, m	3	4	5	6
Marginal damage, d	0.02	0.02	0.02	0.02
Labour supply, b	25	25	25	25

b. population in country 2 is twice larger than in country 1

Autarky equilibrium Country 1; $m_1 = m$	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.876	0.801	0.760	0.739
Wages, w_1	1.249	1.292	1.319	1.337
Consumption, X_1	2.985	4.677	6.387	8.095
Number of firms, N_1	1.219	1.653	2.020	2.337
Country 2; $m_2 = 2$ m				
Tax, t_2	0.739	0.729	0.741	0.765
Wages, w_2	1.337	1.358	1.367	1.371
Consumption, X_2	8.095	11.452	14.698	17.813
Number of firms, N_2	2.337	2.863	3.286	3.636

Post integration equilibrium Country 1; $m_1 = m$	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.569	0.540	0.542	0.558
Wages, w_1	1.439	1.507	1.546	1.569
Marginal damage, d_1Y_1	0.105	0.162	0.218	0.273
Consumption, X_1	4.771	6.698	8.558	10.350
Production, Y_1	5.272	8.106	10.1912	13.665
Profits (exceed sunk costs)	2.079 (Yes)	2.003 (Yes)	1.946 (No)	1.897 (No)
Country 2; $m_2 = 2$ m				
Tax, t_2	0.636	0.633	0.647	0.669
Wages, w_2	1.377	1.375	1.369	1.362
Marginal damage, d_2Y_2	0.181	0.240	0.295	0.348
Consumption, X_2	9.541	13.396	17.116	20.700
Production, Y_2	9.040	11.988	14.762	17.394
Profits (exceed sunk costs)	1.633 (No)	1.461 (No)	1.345 (No)	1.271 (No)

c. population in country 2 is three times larger than in country 1

Autarky equilibrium Country 1; $m_1 = m$	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.876	0.801	0.760	0.739
Wages, w_1	1.249	1.292	1.319	1.337
Consumption, X_1	2.985	4.677	6.387	8.095
Number of firms, N_1	1.219	1.653	2.020	2.337
Country 2; $m_2 = 3$ m				
Tax, t_2	0.733	0.765	0.810	0.861
Wages, w_2	1.364	1.371	1.371	1.366
Consumption, X_2	13.090	17.813	22.235	26.362
Number of firms, N_2	3.085	3.636	4.060	4.394

Post integration equilibrium Country 1; $m_1 = m$	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.541	0.537	0.558	0.591
Wages, w_1	1.540	1.630	1.681	1.710
Marginal damage, d_1Y_1	0.130	0.202	0.272	0.341
Consumption, X_1	4.892	6.698	8.423	10.065
Production, Y_1	6.485	10.079	13.619	17.045
Profits (exceed sunk costs)	2.36 (Yes)	2.323 (Yes)	2.273 (Yes)	2.217 (Yes)
Country 2; $m_2 = 3$ m				
Tax, t_2	0.675	0.701	0.735	0.772
Wages, w_2	1.363	1.348	1.335	1.322
Marginal damage, d_2Y_2	0.262	0.334	0.401	0.464
Consumption, X_2	14.677	20.094	25.268	30.196
Production, Y_2	13.085	25.749	20.071	23.216
Profits (exceed sunk costs)	1.499 (No)	1.32 (No)	1.222 (No)	1.163 (No)

Table A4.2: Different marginal damage to the environment

a. base case assumptions

	Scenario A	Scenario B	Scenario C	Scenario D
Tastes (demand), a	3	3	3	3
Sunk costs, S	2	2	2	2
Labour requirement, l	0.25	0.25	0.25	0.25
Population, m	3	3	3	3
Marginal damage, d	0.01	0.02	0.03	0.04
Labour supply, b	25	25	25	25

b. marginal damage is 1.5 times greater than in the other country

Autarky equilibrium Country 1; $d_1 = 1.2$ d	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.853	0.887	0.920	0.950
Wages, w_1	1.254	1.246	1.239	1.231
Consumption, X_1	3.051	2.954	2.863	2.777
Number of firms, N_1	1.245	1.206	1.169	1.134
Country 2; $d_2 = 0.8$ d				
Tax, t_2	0.841	0.865	0.887	0.909
Wages, w_2	1.257	1.251	1.246	1.241
Consumption, X_2	3.084	3.018	2.954	2.893
Number of firms, N_2	1.259	1.232	1.206	1.181

Post integration equilibrium Country 1; $d_1 = 1.2$ d	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.631	0.678	0.722	0.762
Wages, w_1	1.358	1.344	1.331	1.319
Marginal damage, d_1Y_1	0.052	0.099	0.143	0.184
Consumption, X_1	4.363	4.247	4.138	4.035
Production, Y_1	4.295	4.124	3.970	3.829
Profits (exceed sunk costs)	1.983 (No)	1.950 (No)	1.923 (No)	1.901 (No)
Country 2; $d_2 = 0.8$ d				
Tax, t_2	0.617	0.652	0.686	0.718
Wages, w_2	1.369	1.364	1.359	1.353
Marginal damage, d_2Y_2	0.035	0.070	0.103	0.136
Consumption, X_2	4.363	4.247	4.138	4.035
Production, Y_2	4.431	4.370	4.307	4.241
Profits (exceed sunk costs)	2.064 (Yes)	2.098 (Yes)	2.126 (Yes)	2.149 (Yes)

c. marginal damage is 3 times greater than in the other country

Autarky equilibrium Country 1; $d_1 = 1.5$ d	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.862	0.904	0.942	0.978
Wages, w_1	1.252	1.242	1.233	1.225
Consumption, X_1	3.026	2.908	2.798	2.697
Number of firms, N_1	1.235	1.187	1.142	1.101
Country 2; $d_2 = 0.5$ d				
Tax, t_2	0.832	0.847	0.862	0.876
Wages, w_2	1.259	1.256	1.252	1.249
Consumption, X_2	3.110	3.067	3.026	2.985
Number of firms, N_2	1.270	1.252	1.235	1.219

Post integration equilibrium Country 1; $d_1 = 1.5$ d	Scenario A	Scenario B	Scenario C	Scenario D
Tax, t_1	0.641	0.695	0.744	0.787
Wages, w_1	1.350	1.329	1.311	1.295
Marginal damage, d_1Y_1	0.063	0.118	0.168	0.212
Consumption, X_1	4.365	4.255	4.153	4.058
Production, Y_1	4.196	3.947	3.731	3.541
Profits (exceed sunk costs)	1.923 (No)	1.843 (No)	1.778 (No)	1.724 (No)
Country 2; $d_2 = 0.5$ d				
Tax, t_2	0.605	0.629	0.653	0.676
Wages, w_2	1.378	1.380	1.381	1.381
Marginal damage, d_2Y_2	0.023	0.046	0.069	0.092
Consumption, X_2	4.365	4.255	4.153	4.058
Production, Y_2	4.534	4.563	4.575	4.575
Profits (exceed sunk costs)	2.126 (Yes)	2.213 (Yes)	2.286 (Yes)	2.349 (Yes)

References

Akerlof, G. A. (1982), "Labour contracts as partial gift exchange", *Quarterly Journal of Economics*, vol. 97, pp. 543 – 569

Artis, M. J. and P. J. N. Sinclair (1996), "Labour subsidies: a new look", *Metroeconomica*, vol. 47, pp. 105 – 124

Ashenfelter, O. (1978), "*What is involuntary unemployment*", Proceedings of the American Philosophical Society, vol. 122, April (available also as Working Paper No. 109, Industrial Relations Section, Princeton University, January 1978, from http://www.irs.princeton.edu/)

Blundell, R. W. (1992), "Labour supply and taxation: a survey", *Fiscal Studies*, vol. 13, pp. 15 – 40

Blundell, R. W. and T. MaCurdy (1999), "Labour supply: a review of alternative approaches", in O. Ashenfelter and D. Curdy (eds.), *Handbook of Labour Economics*, Volume 3, Elsevier Science, Amsterdam

Dertouzos, J. N. and Pencavel, J. H. (1981), "Wage and employment determination under trade unionism: the International Typographical Union", *Journal of Political Economy*, vol. 89, pp. 1162 – 1181

Heckman, J. J. (1993), "What has been learnt about labour supply in the past twenty years?", *American Economic Review*, vol. 83 (Papers and Proceedings), pp. 116 – 121

Hoel, M. and P. Shapiro (2001), "*Population mobility and transboundary environmental problems*", CESifo Working Paper No. 488, May (available from http://www.cesifo.de/CES)

Jaffe, A. B., S. R. Peterson, P. R. Portney and R. N. Stavins (1995), "Environmental regulation and the competitiveness of US manufacturing: what does the evidence tell us?", *Journal of Economic Literature*, vol. 33, pp. 132 – 163

Krugman, P. R. and M. Obstfeld (2002), *International economics: theory and policy*, sixth edition, Addison-Wesley, Reading, Massachusetts

Martin, S. (2002), *Advanced industrial economics*, second edition, Blackwell, Oxford

Nickell, S. (1990), "Unemployment: a survey", *Economic Journal*, vol. 100, pp. 391 – 439

Salop, S. C. (1979), "A model of the natural rate of unemployment", *American Economic Review*, vol. 69, pp. 117 – 125

Samuelson, P. A. (1949), "International factor price equalisation once again", *Economic Journal*, vol. 59, pp. 181 – 197

Shapiro, C. and J. E. Stiglitz (1984), "Equilibrium unemployment as a worker discipline device", *American Economic Review*, vol. 74, pp. 433 – 444

Stern, N. H. (1984), "*On the specification of labour supply functions*", Economic and Social Research Council Programme, No. 50, May (also published in R. W. Blundell and I. Walker (eds.), *Unemployment, search and labour supply*, 1986, Cambridge University Press, Cambridge)

Ulph, D. (1995), "*Environmental taxation and entry: a symmetric equilibrium*", Mimeo, University College London

Venables, A. (1997), "*Trade liberalisation and factor movements: an overview*", London School of Economics, Centre for Economic Performance, Discussion Paper 352, July

Weiss, A. (1980), "Job queues and layoffs in labor markets with flexible wages", *Journal of Political Economy*, vol. 88, pp. 526 – 538

Chapter 5

An additional perspective for retail financial products and their regulation

5.1 Introduction

One of the consequences of the progressive withdrawal of the welfare state is that a large proportion of consumers must now take responsibility for their long-term savings decisions, hereafter financial planning. At the same time, governments have felt that consumers cannot rely entirely on market forces for adequate financial planning so there is an increasing interest in regulating the sale of retail financial products.

The standard economic approach to analyse the market for retail financial products is based on models where consumers understand their needs – the outcome of financial planning – and there is asymmetric information about the products between buyers and sellers. So consumers engage in a costly search of the market for the suitable product. The relevant issues here are associated with a process of search, consumers' search costs and the potential gains from search. A good example of this literature is Salop and Stiglitz (1977) where consumers have different search costs and search for the cheapest product.

The purpose of this Chapter is to introduce an additional perspective to analyse the market for retail financial products based on consumers' lack of understanding of their needs. This issue is not unique to financial services. The market for medical services is another context where consumers do not understand their needs. For example, based on a sample of two relatives a consumer believes that the likelihood of developing cancer from smoking is negligible. A doctor has obviously read some of the existing research and has a better understanding of the probability of developing cancer from smoking and will use this knowledge in assessing the patient's symptoms and in recommending a course of action. Similarly, a consumer may infer the probability of getting back into employment based on a sample of two and use this as the basis for establishing her needs of financial products.

This additional perspective is consistent with the understanding of UK regulators that regard advisors as more than a device to reduce consumers' search costs. The economics literature about the regulation of retail financial services clearly identifies the standard issue of information asymmetries and suggests that there are also issues associated with credence goods. A consumer knows that an appliance does not work but does not know the nature of the problem. This means that a crucial aspect of the demand for repairs is that the supplier determines the type of service that is provided with consequences for consumers' welfare, Dulleck and Kerschbamer (2001). There is therefore potential for oversupply (the problem is "minor" and the supplier provides a "major" overhaul of the appliance) and for the overcharging (the problem is "minor" and the supplier charges as if a "major" overhaul was provided). Finally, this additional perspective is also consistent with the approach adopted in Gravelle (1994) to explore the incentive properties of various forms of remunerating financial advisers.

In this context, I develop a very simple model to understand the welfare consequences of consumers' ignorance about their financial needs assuming marginal cost pricing. The model captures the quality of decisions and takes explicit account of the variety of consumers' needs in a population. I then use this framework to explore how regulation might improve on an unregulated equilibrium. I focus on two cases: product regulation and advice regulation. In both cases I make simplifying assumptions such as there is only one product for all consumers and that monitoring of advice is effective. So this is far from being a complete model of the market for retail financial services. The rationale for these assumptions is that a robust understanding of one of the underlying problems in the market for retail financial services – consumers' ignorance about their needs – is required as well as a high-level comparison between the alternatives to address it. Furthermore, these steps should precede the development of a full model of regulated advice, which would make sense if advice regulation dominates product regulation. The overall conclusion of this analysis is that none of these alternatives will make all consumers better off. This is consistent with moves to combine advice regulation and product regulation in the UK. It also means that from an analytical perspective there may be a further stage where the consequences of combining advice regulation with elements of product regulation should be explicitly modelled.

This chapter is structured as follows. Section 5.2 describes the consumer's problem that forms the basis for this perspective on financial services – a consumer's ignorance about her needs. Section 5.3 summarises the literature about retail financial regulation and about information attributes. Section 5.4 presents a model of decision making that reflects this perspective on retail financial regulation. Section 5.5 characterises social welfare in an unregulated equilibrium assuming marginal cost pricing. Section 5.6 explores two potential ways of enhancing welfare: product regulation and advice regulation. Section 5.7 explores some qualifications of the results obtained. Finally, Section 5.8 summarises and concludes.

5.2 The consumer's problem – financial planning

The consumer's problem – financial planning – has been briefly mentioned in the introduction and the purpose of this section is to provide a more detailed discussion.

It is useful to consider a problem similar to financial planning: curing an illness. The consumer may be aware that she is ill but will have few details, if any, about the true condition. So the consumer needs to establish what is the specific illness, e.g. flu, and the specific circumstances that may be relevant to address the problem, e.g. allergies. Then a suitable drug needs to be selected. It is interesting to note that drugs are sold with extensive disclosures (composition, adequate use, side effects, etc.) and, usually, these disclosures are not regarded as sufficient to enable a consumer to make an informed purchase. So an expert's assessment of the problem is required. A recent article in the Harvard Business Review (2002) illustrates how a doctor imparts her knowledge in the assessment of a patient and how the assessment interacts with data about the patient:

> "The other day, for example, a woman – an ex-smoker – came to see me complaining that she'd had a cold for a month. Since colds normally go away in a couple of weeks, this was clearly not a pattern for a cold. So I asked the patient for more data. It turns out her symptoms persisted despite a course of antibiotics. This made a bacterial infection less likely, but the information could still fit a number of other patterns including a virus, an allergy, or even cancer. Eventually, I found that things were worse for my patient during the week than on weekends. As it turned out,

the cleaners were trying a new product in the office where my patient worked."

This suggests that curing an illness depends on the diagnosis of the problem as well as on product information disclosed about the relevant drugs.

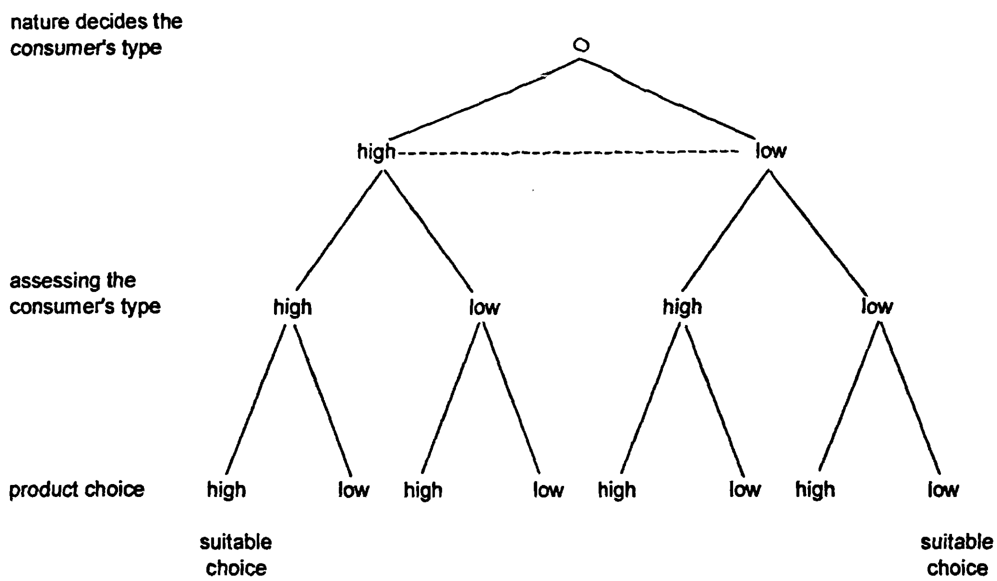
Financial planning represents a similar problem: a consumer needs to establish what she really needs and then choose a suitable product. The first stage – establishing needs – is equivalent to the diagnosis of an illness. It includes gathering information about income and personal circumstances but also considering how they affect the variability of income. A consumer is likely to make an incorrect assessment of her needs of financial products if, for example, she based it on a view about the probability of being unemployed from a sample of two observations or she is not aware of how unemployment relates to income. Once a consumer has taken a view about her needs (correct or incorrect), she will search the marketplace for the relevant product. Disclosures of product details and charges will help her identify the product that suits her needs. However, unlike in the case of curing an illness, the consumer will be able to purchase a financial product even if an expert has not assessed her needs.[1] Later on the consumer will experience the quality of her savings decisions or the error that she may have made. A key aspect of this process is that there is a long lead-time between savings decisions and experiencing the quality of these decisions, which leaves little room for learning-by-doing, Llewellyn (1999).

Figure 5.1 presents a decision tree that summarises the consumer's problem assuming that the needs of financial products are of two types ("high" and "low"). The dotted line between "high" and "low" is used in the conventional way to indicate that the consumer is not sure of her own type. It illustrates the observation that even if product features are fully disclosed to consumers, choices may still be unsuitable if the assessment of her needs is incorrect.

[1] For example, existing rules for the regulation of advice in the UK allow a consumer to declare that she wants to transact on an "execution only" and purchase products without taking advice from a financial advisor.

It is worth noting that the diagnosis involved in financial planning is qualitatively different from the diagnosis associated with the purchase of risk insurance, e.g. health insurance and motor insurance. In the case of risk insurance, the consumer knows something that the insurer does not know and needs to know to price the risk and avoid a financial loss. In this case, product providers have a private incentive to either find out about the consumer's circumstances or design a set of contracts that will reveal private information, see, for example, Stiglitz (1977). In the case of financial planning discussed here the consumer is simply not sure of her own needs. There is no adverse selection – a consumer simply lacks knowledge about her needs – and no moral hazard – a consumer has no control over the events that can affect her income after the product is purchased. Furthermore, the product provider is not affected by the lack of suitability between the product and the consumer – unless regulation imposes a liability on the firm in the case of unsuitable advice.

Figure 5.1: Decision tree for financial planning



So far I have said little about how consumers determine their savings needs. Consumers could do this by themselves as they do with other purchasing decisions. Consumers could also learn from others' savings decisions before making their own decisions. This learning is, however, unlikely to be useful because it is difficult to observe the full extent of the savings' decisions of other consumers. Furthermore, even when someone observes someone else's savings decisions (the last stage in Figure 5.1), the observation is unlikely to extend to the underlying assessment of the needs.

Alternatively, consumers could ask an expert – a financial adviser – to establish their needs and recommend what product they should buy. Financial advisers know how to establish consumers' needs, have learnt about the range of people's circumstances and needs and have a better understanding than the consumer as to how her circumstances are likely to affect the variability of income. Finally, financial advisers also know how to map consumers' savings aims into broad characteristics of products and how to search the market at a minimum cost. This suggests that there may be two different motivations for seeking financial advice: financial planning (establishing consumers' needs) and minimising the (search) costs of finding a suitable product. The first type of financial advice will answer questions such as what do I need to provide for a comfortable retirement. The second type of financial advice will answer questions such as what is the cheapest unit-linked personal pension in the market.

The suggested approach is consistent with the understanding of UK regulators. Broadly speaking, the basic requirement for advice in the rules of the FSA and of its predecessors is that advice is suitable, FSA (2000). Legal commentators, Page and Ferguson (1992), have explained the rationale for such a requirement by reference to the open-ended nature of the consumer's request for advice and quoted a statement by Securities and Investments Board (paragraph 3.3, SIB (1990)):

> "The [suitability] rule is thus a direct statement of one of the basic aims of investor protection – that if an investor puts his trust in the judgement of an investment firm – explicitly or implicitly asking 'what do I need' – that firm should vindicate his trust by tailoring its advice to his needs."

Tailoring the advice to the consumer's needs requires that those needs are assessed. This is reflected in the training that financial advisers undertake before they are allowed to give financial advice. The training programme is divided into three modules and the last one is aptly entitled "identifying and satisfying client needs".[2] The introduction to the

[2] The other two are entitled "financial services and their regulation" and "protection, investment and savings products".

training material provided by one of the approved training bodies makes it clear that the advisers will be imparting their knowledge to consumers when they provide financial advice (introduction to Chapter 4 in Ingledew (2001)):

> "The collection of information by conducting a fact-find will not, in itself, enable you [a financial advisor] to decide what should be recommended to the client. Collection of information is necessary but without an accurate and detailed analysis, it will not assist you to meet clients' needs.
>
> The second stage of providing financial advice then is to assess and meet the client's needs. [The first stage is the collection of information.] This is of the utmost importance as it is the vital link between understanding a client's circumstances and making recommendations which best suit those circumstances. Furthermore it enables you to demonstrate the 'added value' which you are providing when recommending financial products and services which best meet those needs. After all if clients could link their own circumstances to appropriate financial products and suppliers themselves, then your role as an intermediary would be greatly diminished."

So it is not just about finding the details of the consumer. The advisor has to assess the consumers needs and impart his knowledge by making recommendations that best suit the consumer.

Summing up, the consumers' problem is that they are not sure about their needs and these should be correctly assessed to make a suitable choice. Consumers may use an expert to establish their needs and find the suitable product for them.

5.3 Relevant literature

There are two strands of the economic literature that are relevant to the problem discussed in the previous section. First, there is a growing literature on retail financial regulation. Second, there is a large body of economic theory that is structured around the informational attributes of goods. This section reviews briefly these strands of the literature. The aim is to provide a focused review that sets the context for the model developed later in this chapter.

5.3.1 Literature on retail financial regulation

This is a relatively small body compared with the extensive literature that exists on banking and capital requirements. The reminder of this section reviews aspects of this literature relevant to the consumer's problem – the underlying market imperfections, their consequences and the possible remedies.

5.3.1.1 Market imperfections

Market imperfections provide the economic rationale for financial regulation. Llewellyn (1999) provides a comprehensive analysis. The starting point is a list of the specific features of financial services (page 37). Of them, the following are relevant to retail financial services and to the approach adopted here:

- "value is often critically determined by the personal circumstances of the purchaser";
- retail financial services "are often not purchased frequently and hence the consumer has little experience or ability to learn from experience";
- "the value of a financial contract rises over time whereas the value of other products declines. This lowers the net replacement cost of the latter in the event that, at some time in the future, it needs to be replaced due to a fault";
- "it may be a long time (if at all) before the consumer is aware of the value and faults of a financial contract. This limits the power of reputation as an assurance of good products. Even if, in the long run, reputation is damaged by bad behaviour, consumer wealth is impaired in the meantime."

Llewellyn's paper identifies three main imperfections that provide a rationale for regulating retail financial markets: market failures, lemons and confidence and the gridlock problem.[3] Consider each one in turn.

[3] Llewellyn also identifies market imperfections that are related to prudential regulation such as systemic issues and moral hazard arising from the existence of deposit insurance and lender of last resort. These are not discussed here.

Llewellyn puts forward a list of the relevant market failures in retail financial services (page 21) and emphasises that the existence of a market failure implies that there are also costs from not regulating. The first two market failures are:

- "problems of inadequate information on the part of the consumers"; and
- "problems of asymmetric information (consumers are less well informed than are suppliers of financial services)".

The second market failure is clearly about the standard models of information asymmetry. Thus it is not clear what is the first market failure identified precisely adding to the discussion. We can learn more about Llewellyn's views from Goodhart et al. (1998), which includes Llewellyn as one of the co-authors. Amongst other things, Goodhart et al. (1998) compare the case for regulating retail and wholesale financial services (pages 7 and 8):

> "the case for regulation in the retail financial sector is specially powerful because:
>
> - Problems of asymmetric information are greater at the retail level than in professional wholesale markets. The costs of acquiring information are particularly high for small purchases by retail customers.
> - The individual consumer has only a limited ability and opportunity to acquire the necessary skills to enter into complex financial contracts and assess information."

Here the first issue seems to refer to the standard models of information asymmetry. The second one is less obvious. It seems to be the reason underlying the existence of either experience or credence goods. This suggests that the first market failure in Llewellyn (1999) – "inadequate information" – is related to consumers' limited information about their needs.

Consider now the issue of lemons and confidence. In certain cases of information asymmetries, consumers know that there are good and bad products but cannot distinguish between them at the point of sale. As a result, there is reduced consumers' demand for "good" products. In extreme cases, the market breaks down, Akerlof (1970). Alternatively, a similar situation can result in equilibrium characterised by trade in low quality goods, Chapter 17 in Kreps (1990). The model in this chapter will assume that each consumer has a unit demand and makes one purchase. So it does not take into account how information problems may affect consumers' demand.

Finally, consider Llewellyn's suggestion of gridlock. This arises when all firms know how they should behave towards their customers but adopt a different strategy because they secure a short-term advantage and they have no confidence that all their competitors will behave as they should. So, for example, all firms know that they should provide advice of a certain minimum quality but they do not do so because they have no assurance that all competitors will behave in the same way. Thus the role of regulation is to set a common minimum standard that applies equally to all firms. This is the main argument that I will use in Section 5.6.2 to explain why financial advice needs to be regulated.

Spencer (2000) analyses the nature of the information asymmetries that exist in the financial markets. He suggests that financial securities should be classified as credence goods because they involve an element of professional management. For example, the performance of shares in a company depends, amongst other things, on the quality of the management. This may become apparent, if any, long after the purchase of the shares. Spencer also suggests that this problem also arises in the context of fund management and financial advice. It is also noted that, in some cases, a bundle of credence and search goods will be traded jointly. The example suggested is an execution-only transaction where the underlying security is a credence good and the execution-only has some elements of a search good. This is also consistent with the approach suggested here for financial advice – see Figure 5.1 – where there are elements of credence good (assessing the consumer's needs) and search (given an assessment of consumer's needs).

Finally, Page and Ferguson (1992) explore consumer protection in financial services from a legal perspective and discuss the economics of the information problems that provide a rationale for financial services regulation. They also suggest a similar distinction between search (available information) and assessing the information obtained.

It is worth noting that the view that market imperfections can justify government intervention in retail financial services is not universally accepted. Benston (1998) provides a comprehensive critique of financial regulation – not only of retail financial regulation. Benston agrees with the theoretical rationale that market imperfections provide. His main contention seems to be that retail financial services are not different from other goods and services. This means that Llewellyn's list of specific features of retail financial services (see above) also characterises other goods and services. So, for example, regarding the long time that may take a consumer to become aware of the value and faults of a financial contract, Benston asks (page 60):

> "Might not a paint store recommend a poorer product on which it has a higher margin? It might be years before the consumer will find that the recommended product was incorrectly described."

These concerns lead Benston to suggest that in practice, the costs of regulatory intervention will exceed the benefits.[4] To an extent, this is a valid concern because the existence of defined market imperfections does not necessarily mean that the benefits of any measure that aims to correct it will exceed its costs. This could be because the relevant market imperfections are not material enough, the costs of intervention are too high or both, as suggested by the Chairman of the UK's Financial Services Authority (Davies (1998)):

> "regulatory intervention is only likely to be justified if we have the right answers to the following questions:
>
> - what is the nature of the market imperfection (if any) causing a problem?
> - are there ways in which that imperfection can be tackled with outcomes that deliver a net improvement in welfare, or does any possible cure causes worse problems elsewhere?"

[4] Benston (1998) also expressed concerns of a "public choice" nature. These are valid issues about the design of a regulatory body that should be addressed through the choice of a set of accountability requirements. Briault (2002) summarises the main accountability requirements imposed on the UK's Financial Services Authority.

The difference between the two views is that Benston tends to regard as self-evident that the costs will exceed the benefits whereas Davies and others believe that such judgements need to be made on a case by case basis.

5.3.1.2 Consequences of market imperfections

Consider now how the various consequences of market imperfections are relevant to retail financial services. First, consumers may end up paying prices above marginal costs. This could be the result of product differentiation, market power or both but, in the case of retail financial products, it is exacerbated by the information problems as suggested, for example, in Nelson (1970). James (2002) shows that on average consumers in managed unit trusts or life office products pay about 50% more than they should pay.[5]

Second, transactions costs may be higher than they should be in the absence of market failures and are likely to deter some consumers from making a purchase, Llewellyn (1999). Llewellyn also distinguishes between various types of transaction costs: the cost of searching alternative products (search cost); the cost of agreeing precise contract terms (bargaining cost); the cost involved in monitoring post-contract behaviour (monitoring cost); the cost of ensuring that contracting parties deliver on the contract (enforcement cost); the cost of ascertaining the characteristics of the product (verification cost); and the cost of securing redress in the event of contract failure (redress cost).

Third, as suggested in Llewellyn (1999), there may be a direct utility loss that arises from the unsuitability of the product as suggested, for example, by the losses arising from pensions mis-selling in the UK. The latest estimate is that the cost of compensating those affected will amount to £11.8 billion, in addition to administrative costs of £2

---

[5] James (2000, 2002) suggests measuring the price of investing through broad equity unit trusts and life offices in the UK as the price of a managed portfolio that yields the market return on £1. James (2002) shows that, on average, one must invest £1.80 in an actively managed fund to obtain the market return on £1. The paper also estimates that the cost of investing in a benchmark fund ("minimum cost efficient markets fund") is in the region of £1.20. The latter is above the cost of a best-practice index tracker because it is derived by adding to the cost of such index tracker, the cost of management, research and trading sufficient to create and sustain an efficient market. James' estimate of the cost of investing in an actively managed fund is equivalent to about 3% per annum, which is consistent with other estimates (Sandler (2002), page 131).

billion, paragraph 5.39 in Sandler (2002).[6] Another example is the relatively small persistency of long-term retail financial contracts such as pensions. The data suggests that within four years of purchase 43% of consumers contributing to a personal pension had stopped contributing if the pension was sold through a company representative and 38% if the pension was sold through an independent financial adviser, PIA (2001). Some early termination is always inevitable as a result of unexpected changes in circumstances and the concern is with the extent of cancellations, Johnson (2000).

Fourth, consumers may not be making adequate provision to maintain their standards of living after retirement. This may not be entirely irrational for consumers with low incomes given the existence of state pension and means tested benefits that may replace most of their income, see Sinclair (1998) for a discussion. There is nevertheless evidence that consumers may experience a sudden drop of income as they enter into retirement. For example, Banks et al. (1998) show that the extent of the fall of expenditure after retirement cannot be reconciled with a view that savings were expected to be adequate to maintain pre-retirement living standards. More recently, Sandler (2002) has focused on this and suggested (paragraph 3.36) that part of this shortfall can be attributed to features of consumer behaviour that are associated with the diagnosis of the problem, such as balancing consumption over time and working out the necessary level of savings.

5.3.1.3 Remedies to the market imperfections identified

Finally, consider what the literature tells us about the possible remedies to the market imperfections identified. There are different views about what regulation could achieve. So it is useful to start by setting out what remedies should achieve where there seems to be a wider consensus. As Llewellyn (1999) puts it (page 46):

> "regulation should not impede competition but should enhance it and, by addressing information asymmetries, make it more effective in the market place".

---

[6] The pensions mis-selling is a case of deficient advice. Those affected were mainly employees that were already in an occupational pension scheme or had the possibility of joining it. (An occupational pension scheme tends to be a preferable alternative because it may be non-contributory or because the employer tends to match the employee's contribution.) The important aspect of this episode is that personal pensions may be a suitable product for many consumers but not everyone.

There are then different views as to how regulation should work to enhance competition. Benston (1998) offers some thoughts on regulation but they are mainly concerned with prudential regulation and with the structure of financial regulation.[7] Benston's suggestion for regulating retail financial services is rather limited (page 120):

> "to establish an independent agency that would serve as an Ombudsman for consumers who believe they have been mistreated by a financial services firm or salesperson".

So the role of this agency would be limited to help consumers pursue their claims. Benston also suggests that if the claim is validated the Ombudsman "can try to persuade industry participants to change, attempt to educate consumers, refer the matter to legal authorities or suggest legislative action" (page 120).

Llewellyn's views about financial regulation are influenced by various observations. First, consumers may demand regulation and be willing to pay for it.[8] Second, reputation cannot be relied upon to fully address market imperfections in retail financial services because reputation effects are too diffuse in long-term contracts. Third, experience of disclosure suggests that firms will not disclose relevant information in a useful format unless mandated to do so.

Taking these observations into account, Llewellyn suggests conduct of business regulation that establishes the appropriate behaviour and business practices in dealing with consumers, including guidelines for the objectivity of advice. This gives rise to a number of issues that are discussed in Section 5.6.2. Llewellyn's contribution is to suggest a delivery mechanism for conduct of business regulation. This is based on the observation that regulators have various instruments that can be combined in a variety of ways and with various degrees of intensity. He suggests the following approach:

---

[7] The issue regarding the structure of financial regulation is whether the same agency should deal with prudential regulation and consumer protection in retail financial services. Benston (1998) suggests that requiring the same agency to do both tasks will be detrimental to the interests of consumers.

[8] Lewellyn recognises that if consumers believe that regulation is a free good, the demand is distorted. His suggested solution is to make consumers aware that regulation is supplied at a cost even if the price cannot be precisely calculated.

regulators set clear objectives and principles; firms agree with the regulator how their business procedures meet these objectives and principles, which is the basis for a contract between the firm and the regulator; the regulator requires the firm to deliver on its agreed standards and applies sanctions in the case of non-performance.

One of the developments in recent years in the UK has been the introduction of voluntary elements of product regulation setting minimum standards for charges, access and terms for savings accounts, unit trusts, insurance, mortgages (CAT standards – for charges, access and terms) and for personal pensions (stakeholder pensions). This has challenged the extent to which regulation of retail financial services is based on advice regulation.

Johnson (2000) suggests that there are limits to what conduct of business regulation can achieve. First, there is evidence that suggests that the current system of conduct of business regulation results in contributions that do not seem to be adequate to maintain consumers' standards of living after retirement. (This was briefly discussed at the end of the preceding sub-section.) Johnson notes that for pensions, at least, the issue appears to be little contributions rather than little penetration of pensions. Second, consumers have trouble understanding financial products. Johnson notes that this is part of a general problem with numeracy and that it translates into a lack of confidence or unjustified confidence.[9] Third, consumers' lack of understanding is likely to result in an inappropriate product being bought. So Johnson concludes that product regulation and, in particular, the government decision to introduce product standards in financial services is a useful complement to consumer education (now one of the statutory objectives of the FSA) and conduct of business regulation.

Product regulation received recently a boost as a result of the recommendations of the Sandler's review of savings in the UK. It recognised that product regulation will mainly protect consumers "from major detriment", paragraph 10.10 in Sandler (2002). It recommends "the introduction of a suite of simple and comprehensible products" that

---

[9] Johnson provides the following example of unjustified confidence: 40% of consumers in social classes C2DE believe that "a bank is the best place to go to get independent financial advice" despite the fact that almost all the major banks only offer advice about their own products.

could be sold safely without regulated advice that shares many features with CAT standards and existing stakeholder pensions, paragraph 10.12 in Sandler (2002).[10]

James (2002) analyses the evidence about the cost of investing – see footnote 5 – and makes a similar assessment as to the limits of conduct of business regulation. However, James makes his own suggestions for product regulation. This is, broadly speaking, a unit trust which guarantees a minimum level of performance and is backed up by the fund manager own capital.

Summing up, for the most part commentators agree that there is more than information asymmetries between buyers and sellers in retail financial services and suggest that there is also an issue of credence goods.[11] The consequences of these market imperfections, are excessive prices, transaction costs and direct utility losses. There are different views about the extent of these market imperfections and their consequences. Traditionally, conduct of business regulation has been used in the UK to address these market imperfections and their consequences. This has been challenged in recent years with the introduction of elements of product regulation that exist alongside conduct of business regulation.

5.3.2 Literature on information attributes

The previous discussion has suggested that, in addition to the standard problem of asymmetric information, consumers of retail financial services also have an information problem that arises from a limited understanding of their needs. This additional information problem is hinted in the literature dealing with financial services regulation but it has not been considered in detail. I now turn to the literature on information attributes to explore the extent to which we can draw from it.

---

[10] The main differences with existing product regulation arises from specific requirements for "with-profits" stakeholders (Sandler (2002), paragraph 10.27) and limits on investment risk assumed by current CAT-marked mutual funds (Sandler (2002), paragraphs 10.36 and 10.41).

[11] Tirole (1988) suggests that the provision of warranties about the quality of the product can turn credence goods into search goods. To the best of my knowledge, there have been no attempts to do so in the context of financial planning. However, there have been attempts to do so in the context of retail financial products with disappointing results (Sandler (2002), chapter 6). This is the case of with-profits products, which guarantee a minimum sum and undertake to "smooth" returns.

There is a large body of literature where one of the parties to a transaction is not sure about his needs or about his characteristics. One could then split the literature according to whether the party who is not sure about his features behaves strategically (or not). There is a strand of the literature where the agent, who is not sure of his own features, behaves strategically. Examples of that literature are Prendergast and Stole (1996) and Ottaviani and Sorensen (2001). The contexts of these papers are different but in both cases there is an agent – a manager and a professional adviser, respectively – who does not know his own ability and receives a private signal that he can exploit. Typically, this is far from being relevant to consumers of retail financial services.

The other strand of the literature deals with credence goods. This starts from the observation that in services such as the provision of repair services "contrary to the basic assumption of conventional demand theory, the consumer is unaware of the ability of the repair service to satisfy a given want" (Darby and Karni (1973), page 67). So unlike the previous class of models, consumers do not behave strategically.

The literature on credence goods assumes that the consumer knows that she has a problem but does not know how serious it is. Typically, it is assumed that the problem could be either a major or minor problem. Sellers are experts who determine consumers' needs. Dulleck and Kerschbamer (2001) provide a unifying treatment of the existing literature. This literature assumes that the expert makes a perfect diagnosis of the problem to focus on the incentives to provide an inefficient treatment or to overcharge. An inefficient treatment means that an expert can try to misrepresent a minor problem – requiring minor treatment – as a major problem. So it could be a case of either undertreatment (providing a cheap treatment when an expensive treatment is required) or overtreatment (providing an expensive treatment when a cheap treatment is required). Overcharging means that the expert charges the consumer for a treatment that is more expensive than the treatment provided.

An example of this literature is Wolinsky (1993) which focuses on consumers' search for multiple opinions as a device to mitigate expert's incentive to misrepresent the consumer's needs. (There is no possibility of learning about one's own problem.) Consumers incur a fixed 'search' cost to obtain a diagnosis of the problem – this includes both the cost of searching an expert and the cost of undertaking the diagnosis. Assuming that experts announce in advance the price of the two treatments, Wolinsky shows that the market equilibrium is highly dependent on search costs. If they are sufficiently low, then some experts specialise on the minor problem while others specialise in the expensive treatment. Customers sample first one of the experts specialising in the minor problem. If the first expert recommends the expensive treatment then the consumer will go to an expert specialising in the major treatment. So some consumers may incur search costs twice but there is no fraud in this equilibrium.[12]

As these papers are not written with financial services in mind, the effect of an incorrect choice is just paying higher prices rather than the direct utility loss associated with lack of suitability. For example, Dulleck and Kerschbamer (2001) assume that in the case of unnecessary undertreatment there is no direct loss of consumer utility – utility is zero. This may well be the appropriate way to think about appliance repairs, which is the usual reference point in this literature and where the result of an inadequate repair is just no extra utility from the appliance. It also suggests that inferring from standard models of credence goods might underestimate the welfare loss associated with an unregulated equilibrium in retail financial services.

This discussion suggests that it is worth noting the difference between informational attributes of goods (search or credence goods) and a description of consumer behaviour (the relevance of a search process). For example, Salop and Stiglitz (1977) deal with search goods but the consumer may not necessarily search the market. At the same time, a search process (albeit a different one) is also relevant to credence goods where consumers have to search for a diagnosis of the problem. And, as suggested above, this search process and its cost are important to an understanding of credence goods. This suggests approaching financial planning using a model that:

- uses a simple process of search, where relevant;
- specifies financial planning as in the credence goods literature; and
- captures the extent of utility losses from unsuitable purchases.

[12] The equilibrium satisfies that in the case of consumers requiring the major treatment, the expected cost of this sequential search are lower than the cost of approaching directly the expert specialising in the major treatment.

An example of a paper that shares some of the features suggested here is Gravelle (1994), which explores the welfare properties of alternative approaches to remunerate financial advisers – commission and fees. The paper does not refer to the literature about credence goods but the set up of the problem shares some of its features. Consumers are described by two parameters: the "gross benefit" and the "mismatch" parameters. The net benefit to the consumer of a financial product is the difference between these two parameters. Initially consumers know the value of the gross benefit but not the value of the mismatch parameter, only its distribution within the population. Advisers contact consumers and provide information that enables them to discover their true mismatch parameter. Consumers buy the product if the net benefit exceeds the price of the product.[13] The main difference with the model developed here is that the consumer's true type cannot be perfectly assessed before the purchase with the consequent need to model the extent of utility loss from an unsuitable purchase.

5.4 The model

5.4.1 Overview

We approach financial planning as a credence good. Consumers know that they need to make long-term savings decisions but they do not know for sure their needs. Recall that there is no adverse selection – a consumer simply lacks knowledge about her needs – and no moral hazard – a consumer has no control over the events that can affect her income after the product is purchased.

Consumers' decision making process can be described by reference to the timeline in Figure 5.2.[14] Consider briefly each aspect of the process.

A consumer has some knowledge about her true type (as defined by her needs) and believes that her true type is one of various possible types – the initial set of beliefs.

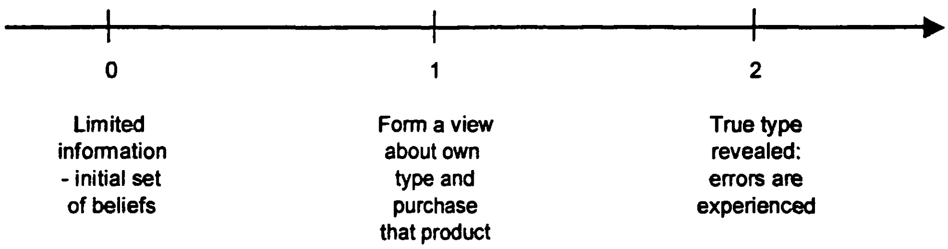
---

[13] The paper shows that the fee system is not necessarily superior to the commission system. The intuition is that consumers gain from advice even if they do not buy the product when informed. However, consumers will not become informed under a fee for advice when they pay *first* because brokers will set the fee for advice reflecting the weak market position of consumers. So too few consumers become informed.

[14] This description is based on a framework for analysing consumer detriment under conditions of imperfect information developed by the Office of Fair Trading, OFT (1997).

These beliefs are correct in the sense that the set does indeed contain the consumer's true type. The size of the initial set of beliefs represents the extent of a consumer's ignorance about her true type – the larger the more ignorant – and the extent of the potential mistakes.

Figure 5.2: Timeline for financial planning



Next, the consumer forms a view about her true type and chooses the product that matches her perceived type. I consider three alternative ways of arriving at a view about the consumer's type and hence purchasing the product. In an unregulated equilibrium consumers make up their own mind and pick one of the products in the initial set of beliefs. (These results will be used as a benchmark.) I then discuss two regulatory approaches aimed at improving welfare – product regulation and advice regulation.

After the purchase, a consumer's true type is revealed so, for example, the consumer experiences the variability of her income – the true set of beliefs. Note that this could be years after the purchase decision.

The model developed here has certain similarities with models of horizontal product differentiation.[15] The main assumptions of this class of models (see, for example, Martin (2002) or Tirole (1988)) include consumers with different preferences evenly spread along the market, buying one unit of the good and incurring a transport (search) cost, products that are imperfect substitutes and localised competition. The literature starts

---

[15] Horizontal product differentiation refers to cases where consumers do not have a common preference ordering. The opposite is vertical product differentiation where consumers have a common preference ordering.

with Hotelling (1929).[16] The main conclusion from this literature (Thisse and Norman (1994)) is: "to recognise the central role of product differentiation in the workings of the market process and the determination of prices. [...] However, product differentiation endows each firm with a degree of local monopoly power allowing it to price above marginal cost." However, the model developed here differs in a number of ways from the standard models of horizontal product differentiation. For example, consumers do not know their true type when they buy the product – unlike in the standard models of product differentiation. In addition, I assume marginal cost pricing to assess ex-post levels of utility whereas standard models of product differentiation are used to explain differences between marginal costs and prices.

5.4.2 Consumers

Each consumer must plan its finances (assess her needs) and choose a product without knowing her true needs.[17] As suggested in Section 5.2 consumers cannot really learn from the decisions of other consumers.

Suppose that assessing a consumer's needs of financial products requires the identification of two parameters. The first parameter, y, represents known factors such as the expected level of income. The second parameter, z, is not known with certainty at purchase. For example, it represents factors such as the variability of income, which partly depends on events such as moving jobs and getting divorced and on other financial products that the consumer may already have.

---

[16] Hotelling starts from the observation that, in reality, equilibria that result from competition among the few is relatively stable, i.e. small reductions in prices do not tend to take away a firm's profits as implied by Cournot's model. Hotelling goes on to explain the observed stability as a result of product differentiation where a firm's demand is proportional to the length of the market served. The equilibrium is then characterised by minimum product differentiation. This analysis remained unchallenged for about 50 years. d'Aspremont et al. (1977) show that under Hotelling's formulation when firms are too close a marginal change in price can attract the entire demand from the rival. The equilibrium is only stable if one assumes that firms' strategies exclude undercutting the rival's price (inclusive of transport cost). Alternatively, i.e. if one makes different assumptions about the demand – quadratic transportation costs, the equilibrium is characterised by maximum product differentiation. This class of models has also been used to explore, inter-alia, the role of brands in entry. Schmalensee (1978) assesses the equilibrium in an industry characterised by horizontal product differentiation, fixed costs and relative immobile brands (ready-to-eat breakfast cereal industry). This suggests that there would be equilibria where the established brands earn excess profits but no potential entrant (or established firm) finds it attractive to launch a new brand.

[17] I am therefore excluding the possibility that consumers wrongly decide not to make provision.

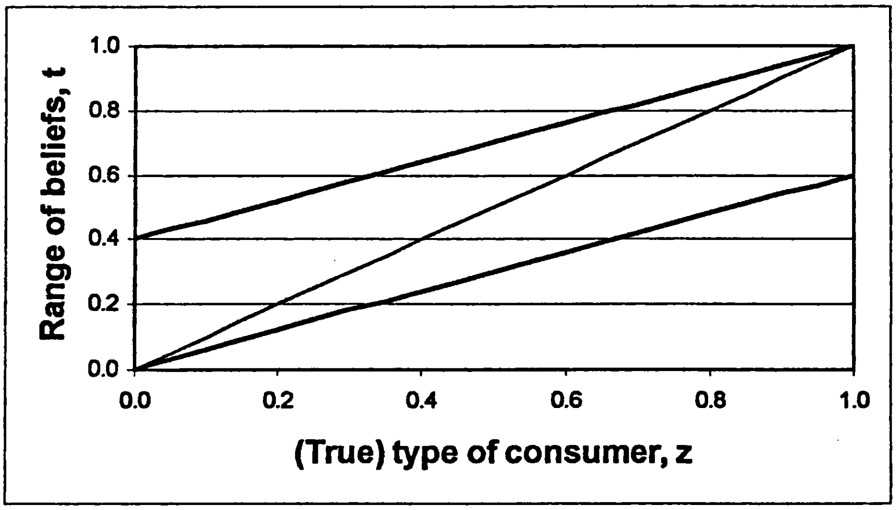
Suppose for simplicity that all consumers have the same income (known features) and only differ in the unknown parameter, z. Suppose further that consumers in the population can be sorted by z and that the higher it is the larger the financial provision required. For example, if z represents factors such as the variability of income, then the higher it is the more provision is needed. I also assume that the consumer's true type (as defined by her needs), z, is a continuous variable lying in the close interval [0, 1]. Consumers are uniformly distributed on [0, 1].

A consumer is not completely ignorant about these unknown factors that represent her true type and has an initial set of beliefs about it. In particular, we assume that a consumer of type z believes with certainty that she is of type t where t is uniformly distributed in the interval

$$[z - \varepsilon z, z + (1 - z)\varepsilon] \quad (1)$$

and where ε is a measure of consumer's ignorance about her true type. Figure 5.3 below shows the initial set of beliefs of different type of consumers for certain values of ε. The lower (higher) line represents the lower (upper) bound in the range of beliefs for a value of ε of 0.4. The diagonal represents the case of full information (ε is 0) so the beliefs match the consumer's true type.

Figure 5.3: Range of consumers' beliefs (for $\varepsilon = 0.4$)



The figure shows that different consumers have different beliefs and that they all share the same level of ignorance about their own needs. A consumer of type $z = 0$, who needs a small amount of financial provision, believes that her true type ranges between [0, ε]. This consumer's mistake is to make an excessive financial provision. However, a consumer of type $z = 1$, who needs a substantial amount of financial provision, believes that her true types ranges between [1 − ε, 1]. This consumer's mistake is to make insufficient financial provision. Finally, a consumer of type ½ believes that her true type ranges between [½ (1 − ε), ½ (1 + ε)]. In this case, the direction of the mistake is ambiguous: the consumer may make too much or too little financial provision. Summing up, consumers are different. They make different types of errors though the magnitude of the potential error (the degree of consumers' ignorance) is the same.

Finally, a consumer incurs a search cost, s, to plan her finances. This represents the direct cost to the consumer (e.g. value of own time) or the fees paid to a financial adviser or both. As in Wolinsky (1993), I start by not differentiating between these categories. The search cost may nevertheless be a fixed value or a function of some variable and this is discussed in the context of the unregulated equilibrium and of the regulatory approaches discussed in Section 5.6.

I characterise a product, z^*, according to the type of consumer that that it suits – the factor unknown to the consumer at the point of sale that it represents. Thus a consumer of type z, incurs a search cost s, purchases a product of type z^* and obtains the following utility

$$u(z,s,z^*) = y - s - pz^* - \gamma\,(z - z^*) \tag{2}$$

where y is income, p is the price per unit and $\gamma\ (z - z^*)$ is a function that represents the direct utility loss from purchasing an unsuitable product.

This formulation assumes that consumers observe the price of the product and the search cost (the cost of advice, where relevant). This is a strong assumption when one has in mind the market for retail financial products in the UK.[18] It is nevertheless a useful assumption to focus on the effect of a consumer's ignorance about her true type. This

[18] A new disclosure regime was introduced in 1995. As a result, firms sell financial products making projections about future values based on their charges and use common assumptions for the growth rate, set by the regulator. There are, however, certain costs that are not disclosed because they are charged directly to the fund such as the costs incurred in buying and selling the underlying portfolio, James (2000).

formulation also suggests that searching for the lowest cost product would not necessarily maximise a consumer's welfare, even if all costs were disclosed, because the product could impose a significant utility loss. An analogy would be searching for a portfolio that maximises a consumer's welfare. In the world of the capital assets pricing model, a consumer maximises welfare by combining the market portfolio[19] and borrowing or lending at the riskless rate. This requires information about a consumer's preferences between risk and return, which may or may not be available. Here I posit a situation where a consumer has limited information about her true type.

Consider now the function representing the direct utility loss from buying an unsuitable product, $\gamma(z-z^*)$. If a consumer buys product z^* where $z^* > z$, she is effectively making provision in excess of her needs and will be covered for any eventuality affecting her income. The only extra cost incurred is financial: the cost of making an excessive provision. On the other hand, if a consumer buys a product z^* where $z^* < z$, the consumer is not making an adequate financial provision and will not be covered for all possible variations of her income. A particular case that satisfies this description is where

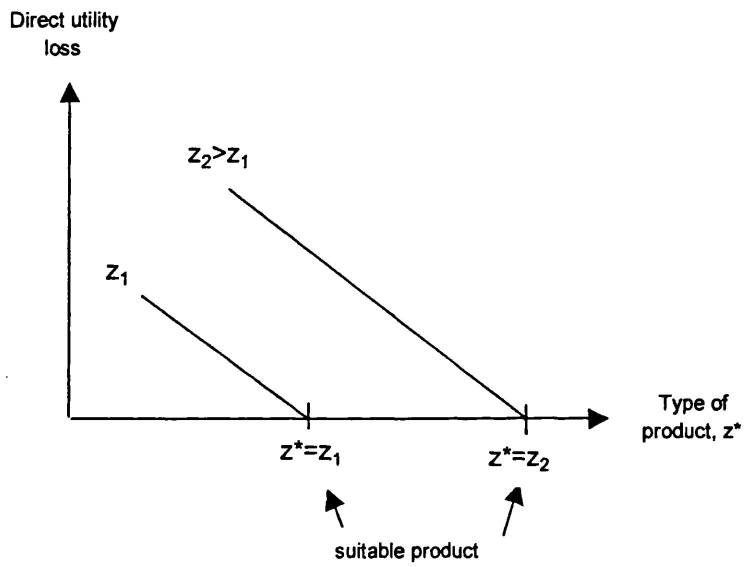
$$\gamma(z-z^*) = \alpha\, Max(z-z^*, 0) \tag{3}$$

This is similar to the assumption made in standard models of horizontal product differentiation about the transportation cost. The main difference is that the transport cost tends to be symmetric unlike here. (Other differences in the supply side will be highlighted later.) Figure 5.4 shows the direct utility losses for two consumers of (true) type z_1 and z_2 that purchase different products, z^*.

The figure illustrates the point made earlier about consumers not being identical. So, for example, for purchases to the left of $z^* = z_1$, these two consumers experience different direct utility losses. For purchases between $z^* = z_1$ and $z^* = z_2$ only the consumer of type z_2 suffers a utility loss. (The consumer of type z_1 suffers a financial loss because she is providing in excess of what she requires.) And, for purchases to the right of $z^* = z_2$, none of the consumers experience a direct utility loss.

[19] This is a portfolio where the proportion invested in any risky asset equals to the market value of that asset divided by the market value of all risky assets.

Figure 5.4: Direct utility losses for two types of consumers (z_1 and z_2)



5.4.3 Firms and retail financial products

I assume that each firm offers one product and characterise a firm by the product it offers. This is far from reality in financial services where multi-product firms are the norm. This assumption is, however, a convenient starting point to model this market. In any event, the assumption has a limited effect on the analysis. First, in the case of product regulation, there is only one product so there is no scope for a multi-product firm. Second, in the case of advice regulation, the assumptions about advisers' remuneration – see below – insulate them from any specific considerations about this.

A retail financial product mainly specifies the way in which money would be accumulated. This should include – amongst others – a schedule of contributions, the accumulation of interests (the exposure to risk), the term of the contract, the tax credits available and the conditions for stopping contributions and withdrawal of funds.[20] At one extreme, there are products that can suit any consumer and accommodate any changes in consumers' circumstances and allow stopping payments and withdrawals. At the other extreme, there are (pure) financial contracts that do not accommodate changes

---

[20] In reality, savings products such as pensions also specify a set of contingencies that result in a payment from the firm to the consumer, e.g. a fixed payment in the event of death. This, however, is hardly the main element of the package or the main reason for its purchase. There are pure risk insurance contracts available that provide just that.

in a consumer's circumstances through the term of the contract. Thus, I characterise a retail financial product, z^*, in terms of the factor unknown to the consumer at the point of sale that it represents. So, for example, if z represents the variability of income then products can be described by the changes in a consumer's circumstances that the product can accommodate. I assume that there are no economies of scale and that there are as many products as types of consumers so z^* takes a value between [0, 1]. This is the other main difference with the standard model of product differentiation.

The cost of a retail financial product depends on the extent to which changes in consumers' circumstances can be accommodated, as this requires, for example, more expenditure on systems and staff. If funds accumulated are only returned at the end of the contract's term, the costs will be lower than where the contract accommodates any variation in consumer's circumstances during the contract's term. Suppose that the costs of a product are a linear function of the provision, $c_1 z^*$. Thus more provision is more expensive but there are no economies or diseconomies of bundling products. It is also assumed that $\alpha \geq c_1 > 0$ to guarantee that if a consumer of type z knows for certain her true type, i.e. $\varepsilon = 0$, she makes the adequate provision, i.e. $z^* = z$. Firms announce the unit price, p, that they charge for the product.

5.5 Unregulated equilibrium

This section characterises a simple unregulated equilibrium to illustrate the welfare effects of consumers' information problems.

In an unregulated equilibrium consumers take responsibility for their financial planning. I assume that consumers incur a fixed search cost, s, to obtain the retail financial product that they believe to suit them. As in Wolinsky (1993), this search cost accounts for the time and effort required to ascertain what product they believe is required and to find that product in the market.

Consider first an ideal situation where there are no information problems once the search cost is incurred, consumers pay the marginal cost and there are no fixed costs. The latter means that there is a retail financial product that suits the characteristics of each consumer. So the consumer incurs the search cost, s, and learns her true type, z and chooses $z^* = z$. The consumer's ex-post utility is

$$\upsilon(z,s) = y - s - c_1 z \qquad (4)$$

Several factors, however, make this full information equilibrium unlikely. First, the consumer is unable to identify with certainty her needs and hence the suitable product even after incurring the search cost. (As mentioned earlier, this is one of the main differences with the literature on credence goods.) Second, there are economies of scale and/or sunk costs at the product level. This means that not all the possible product varieties would be profitable and that a smaller number would be offered. As a result some consumers would not find the product that matches their needs. Third, firms may have a degree of market power. This means that firms can charge prices in excess of marginal costs.

All these factors are present in different degrees in financial services and are relevant to understanding financial planning. Here I want to focus on consumers' ignorance about their own needs so I make some simplifying assumptions. The main simplifying assumption that I make is that there is marginal cost pricing. I also assume that sellers do not behave strategically so unsuitable purchases occur because needs cannot be perfectly assessed.

In an unregulated equilibrium each consumer makes up her mind in period 1 about her true type. So in period 1, a consumer of type z incurs a search cost s and forms a belief about her true type being t. She maximises ex-ante utility from purchasing z^*, $u(z, s, z^*)$, so she buys the product that matches her beliefs about her true type, $z^* = t$.

There is no learning from other consumers or experience, so different consumers of (true) type z will form different beliefs, t. Thus, they end up buying different products and experiencing different levels of ex-post utility. There is a dispersion of behaviour within each consumer type and there are no systematic errors. There is no reason to assume that the distribution of these beliefs is different for each type of consumer so consumers will end up uniformly distributed across all products.[21] Thus the average ex-

[21] This is a simplified view. In reality, firms use advertising and other means to attract consumers.

post utility of a consumer of type z with an initial set of beliefs represented by ε that buys one of the products offered will be the average of the utilities – equation 2 – over the range of purchases that might occur – equation 1.[22] This is

$$\bar{\upsilon}^{UR}(z,s,\varepsilon)=\frac{1}{\varepsilon}\int_{z-\varepsilon z}^{z+(1-z)\varepsilon}u(z,s,t)dt \qquad (5)$$

Given equation 2 above, this can be written as

$$\bar{\upsilon}^{UR}(z,s,\varepsilon)=y-s-\frac{c_1}{\varepsilon}\int_{z-\varepsilon z}^{z+(1-z)\varepsilon}t\,dt-\frac{1}{\varepsilon}\int_{z-\varepsilon z}^{z}\alpha(z-t)\,dt \qquad (6)$$

After some re-arranging, I get

$$\bar{\upsilon}^{UR}(z,s,\varepsilon)=y-s-\frac{c_1}{2}[\varepsilon+2z(1-\varepsilon)]-\frac{\alpha z^2}{2}\varepsilon \qquad (7)$$

As one would expect, if there is no consumers' ignorance about their own needs, $\varepsilon = 0$, the average ex-post utility from equation 7 equals the ex-post utility from equation 4. More interesting is that the average ex-post utility in equation 7 is a decreasing function of consumer's ignorance, $\frac{\partial \bar{v}^{UR}}{\partial \varepsilon}<0$. The intuition of this result is simple, as ε increases so does the range of products from which consumers buy the product. This increases the possibility of errors and therefore the average ex-post utility decreases. The average ex-post utility is also a decreasing function of the unit cost of the product, c_1, of the search cost, s, and of the unit cost of utility losses from an unsuitable product, α.

Result 1: consumers' ignorance about their needs reduces consumers' average ex-post utility even when products are priced at marginal cost.

5.6 Regulatory approaches

Let us consider how this information problem can be addressed. If all consumers were both identical and ignorant about their characteristics, a regulator could address this information problem by identifying the product that is suitable to all consumers and then

[22] The integral in equation 5 is scaled down by $1 / \varepsilon$ because the length of the interval over which consumers' beliefs are distributed is ε – see equation 1.

mandating that only that product is sold.[23] Letting firms compete in the supply of that product would maximise social welfare provided that the sunk costs and economies of scale of providing the product were not as significant as to result in a small number of firms entering the market.

However, consumers are different so this approach will not necessarily maximise social welfare. A welfare maximiser regulator could consider two broad alternatives. First, the regulator can still specify the details of the product that would be offered to all consumers, despite their heterogeneity. I refer to this as "product regulation". Second, the regulator could specify how financial products are sold, including the extent to which sellers should assess consumers' needs. Firms would be free to choose the features of the retail financial products offered and (subject to adequate disclosures) to compete in prices. I refer to this as "advice regulation". The UK is an example of a regulatory system based on advice regulation.[24] This is the motivating example but it is worth stressing that I do not aim to model that so, for example, the analysis does not take account of the various distribution channels that exist.

The next two sub-sections will characterise each of these approaches in the context of the model developed in Section 5.4 and compare the resulting ex-post utility with the ex-post utility in an unregulated equilibrium. The third sub-section provides some simulations to compare the three approaches.

5.6.1 Product regulation

Consumers are not sure about their own needs. Product regulation does not address this problem but its consequences: the unsuitable choices that consumers can make. The intuition of this approach is that even if consumers' needs are different, it may be

---

[23] This assumes that the State provides the regulator with sufficient powers to mandate that only products meeting these requirements are sold.

[24] In the UK regulatory context, a distinction is made between "suitable advice", which tends to refer to a product being suitable to a consumer given a firm's range of products, and "best advice", which refers to the best product within a class product given the consumers' characteristics. Company representatives (and agents of product providers – appointed representatives) provide suitable advice and individual financial advisors provide best advice. In this analysis, I abstract from that distinction and note that the idea of a product that matches the consumer's characteristics is broadly consistent with best advice. I use the term "suitable product" in this sense.

possible to limit the utility loss that can arise from consumers' ignorance about their needs by specifying product details.

As indicated earlier, the UK is an example of a system based on advice regulation and this has been challenged in recent years. As discussed earlier (Section 5.3.1.3), the rationale for product regulation is based on the limits of advice regulation. However, it is worth stressing that it is widely accepted that, even with product regulation, there are utility losses because all consumers will buy the same product despite their different circumstances. Thus HM Treasury presented its product standards for unit trusts and other financial products as follows (HMT (1998)):

> "Benchmarked ISAs [Individual Savings Accounts] should always offer savers a reasonable deal. The deal may not be the very best on the market, but savers using products which meet – or better – the standards should not get ripped off".

Three further reasons are offered in favour of product regulation. First, the assessment of consumers' needs and search will be simpler – in particular when there is only one product. Second, where consumers have information problems and product variety is large, price competition might be weak. If so, product regulation that is structured as minimum standards may encourage firms to compete in the supply of a more homogenous product with consequent efficiency effects. Finally, product regulation can be a useful alternative to the costs of advice regulation (more on this in Section 5.6.2). The concern is that the cost of regulated advice may make the provision of advice to consumers with small contributions uneconomical.

The model discussed here also captures the first of these additional reasons for product regulation. The last one is also captured because the results for product regulation are compared with the results for advice regulation in Section 5.6.3.

The regulator observes the distribution of the true type of consumers but does not observe the type of a specific consumer and cannot match consumers to specific products. A welfare maximiser regulator could mandate that all products sold meet specific minimum standards, like CAT standards and stakeholder pensions. This could be modelled by assuming that product regulation requires any product z^* to satisfy that

$z^* > z_{min}$ where z_{min} is set by the regulator. The main advantage of this approach to model product regulation is that it nests the unregulated equilibrium, $z_{min} = 0$. I have not adopted this approach here because modelling a minimum product in this context is not easily tractable from an analytical perspective. So for the purposes of this work, I have assumed that the regulator specifies one product for all consumers.

So although a consumer could form a view (in period 1) about her type, t, and buy $z^* = t$, she will now buy the regulated product, $z^* = \hat{z}^*$. In this case, search costs will be minimal and there is little need (if any) to assess consumers' needs so I assume for simplicity that there are no search costs, $s = 0$. This is consistent with Sandler's suggestion that products meeting minimum standards be sold without advice, i.e. no assessment of needs, and that the provider (or adviser) only provides information (Sandler (2002), paragraph 10.46). Using equation 2, I denote the consumer's utility as $u(z, 0, \hat{z}^*)$.

The regulator chooses the regulated product, $\hat{z}^*$, to maximise social welfare. This is the sum of the utility to all consumers given the specific product selected

$$w^{PR}(\hat{z}^*) = \int_0^1 u(z, 0, \hat{z}^*)dz \tag{8}$$

Given equation 2 above, this becomes

$$w^{PR}(\hat{z}^*) = y - c_1\hat{z}^* - \int_{\hat{z}^*}^1 \alpha(z - \hat{z}^*)dz = y - c_1\hat{z}^* - \frac{\alpha}{2}(1 - \hat{z}^*)^2 \tag{9}$$

The optimal choice of product, $\hat{z}^*$, is then

$$\hat{z}^* = 1 - \frac{c_1}{\alpha} \tag{10}$$

Given the assumption that $\alpha > c_1$, the regulated product will be a value in the interval (0,1). So the optimal form of product regulation requires less provision (lower z) when the cost, c_1, increases and more provision when the utility loss, α, increases. The latter is not surprising because a direct utility loss only arises if there is under provision – see Section 5.4.2 above.

Substituting 10 into the consumers' utility function in equation 2, I get the ex-post utility of a consumer of type z that buys the regulated product. This is

$$
\upsilon^{PR}(z) = \begin{cases} y - c_1\left(1 - \frac{c_1}{\alpha}\right) & 0 \le z \le \left(1 - \frac{c_1}{\alpha}\right) \\ y - c_1\left(1 - \frac{c_1}{\alpha}\right) - \alpha\left[z - \left(1 - \frac{c_1}{\alpha}\right)\right] & \left(1 - \frac{c_1}{\alpha}\right) < z \le 1 \end{cases} \tag{11}
$$

Note that because the regulator identifies the optimal form of product regulation based on the range of consumers' true types, a consumer's ex-post utility will be independent of her degree of ignorance and beliefs about her true type.

The first line of equation 11 represents the ex-post utility of those that are providing in excess of their needs, given the optimal form of product regulation. This is independent of a consumer's true type because there are no direct utility losses and they all pay the same amount. The second line in equation 11 represents the ex-post utility of those that are providing less than required. This is a decreasing function of z – the higher it is the larger the utility loss that arises from buying the (optimal) regulated product. It also shows that the ex-post utility of those that are providing in excess is larger than the ex-post utility of those that are providing less than required.

The comparative statics are not usual in that changes in the parameters affect the regulated product. This in turn determines whether a consumer is providing in excess or less than required. Thus the comparative statics focus on changes in ex-post utility assuming that a consumer does not switch between these categories. Equation 11 suggests that the ex-post utility is a decreasing function of the unit cost of utility losses, α, for consumers that are providing in excess. For those consumers that are providing less than required, the ex-post utility is an increasing function of α.

Let us compare now the ex-post utility that results from product regulation and from the unregulated equilibrium. These are summarised in Table 5.1 for some illustrative values of the parameters.

Table 5.1: Comparison of unregulated equilibrium and product regulation

True type of consumer	Unregulated equilibrium: Full information $\varepsilon = 0$ (equation 4)	Unregulated equilibrium: Completely ignorant $\varepsilon = 1$ (equation 7)	Product regulation (equation 11)
$z = 0$	$y - s$	$y - s - \frac{c_1}{2}$	$y - c_1\left(1 - \frac{c_1}{\alpha}\right)$
$z = 1 - \frac{c_1}{\alpha}$	$y - s - c_1\left(1 - \frac{c_1}{\alpha}\right)$	$y - s - \frac{c_1}{2} - \frac{\alpha}{2}\left(1 - \frac{c_1}{\alpha}\right)^2$	$y - c_1\left(1 - \frac{c_1}{\alpha}\right)$
$z = 1$	$y - s - c_1$	$y - s - \frac{c_1}{2} - \frac{\alpha}{2}$	$y - c_1\left(1 - \frac{c_1}{\alpha}\right) - c_1$

Consider the case where consumers' ignorance is low, e.g. $\varepsilon = 0$. The table suggests that product regulation will make a consumer of type $z = 0$ worse-off than in the unregulated equilibrium if

$$s < c_1\left(1 - \frac{c_1}{\alpha}\right) \tag{12}$$

A similar result is obtained for a consumer of type $z = 1$. The right hand side of equation 12 is positive – see earlier discussion about the regulated product, equation 10. This suggests that product regulation will make consumers worse-off if the search costs in the unregulated equilibrium are sufficiently low. The intuition of this result is very simple. If consumers' ignorance is small, so is the possibility of buying a product that does not really suit the consumer after the search costs are incurred. However, the regulated product is designed on the basis of the range of consumers' types and ignores consumers' ability to choose the appropriate product. Thus introducing product regulation – forcing everyone to buy a specific product – results in lower levels of ex-post utility if the search costs in the unregulated equilibrium are sufficiently low. So consumers are better off making up their own mind.

The exception is the consumer who happens to be of the type of the regulated product. Product regulation will make this consumer worse-off if $s < 0$. Search costs are positive, so this consumer will be better off under product regulation. The intuition is that in an unregulated equilibrium the consumer devotes resources (search cost) to find the suitable product but not in the case of product regulation where the regulated product happens to be the suitable one.

Result 2: when consumers' ignorance is small and search costs in the unregulated equilibrium are sufficiently low, product regulation makes almost everyone worse off.

Consider now the case where consumers' ignorance is large, e.g. $\varepsilon = 1$. Product regulation will make a consumer of type $z = 0$ worse-off than in the unregulated equilibrium if

$$y - c_1\left(1 - \frac{c_1}{\alpha}\right) < y - s - \frac{c_1}{2} \tag{13}$$

After some re-arranging, it can be shown that this requires that search costs satisfy the following

$$s < c_1\left(\frac{1}{2} - \frac{c_1}{\alpha}\right) \tag{14}$$

Note that the right hand side of this condition will be negative if $\alpha < 2c_1$, which may or may not be satisfied given the assumption that $\alpha > c_1$. The important thing to note is that if the right hand side of equation 14 is negative (α is sufficiently low or c_1 sufficiently large), the condition will not be satisfied because search costs are positive. In that case, this type of consumer will be better off under product regulation regardless of the quantum of the search cost.

Consider now the case of a consumer of type $z = 1$. Product regulation will make this consumer worse-off than in the unregulated equilibrium if

$$y - c_1\left(1 - \frac{c_1}{\alpha}\right) - c_1 < y - s - \frac{c_1}{2} - \frac{\alpha}{2} \tag{15}$$

After some re-arranging, this requires that search costs satisfy the following

$$s < c_1\left(\frac{1}{2} - \frac{c_1}{\alpha}\right) + c_1 - \frac{\alpha}{2} \tag{16}$$

Note that the right hand side of this condition will be negative if $\alpha > 2c_1$, which may or may not be satisfied given the assumption that $\alpha > c_1$. Again, the important thing is that if the right hand side of equation 16 is negative (α is sufficiently large or c_1 sufficiently small), the condition will not be satisfied because search costs are positive. It suggests that this type of consumer will be better off under product regulation regardless of the search cost.

Let us consider now what may happen in a population with different types of consumers. For example, suppose that $\alpha > 2c_1$. In this case equation 16 is not satisfied regardless of the search costs – the right hand side is negative. Consumers of type $z = 1$ are better off under product regulation. Consider the position of consumers of type $z = 0$. The previous discussion suggests that if $\alpha > 2c_1$ the right hand side of equation 14 will be positive. So if the search cost are not sufficiently high then this type of consumers will be worse off with product regulation.

Result 3: when consumers' ignorance is large, product regulation could make some consumers better off and some worse off. If the unit cost of utility loss is large, α, those made better off by product regulation are those needing a substantial provision.

5.6.2 Advice regulation

Consumers are not sure about their needs of financial products and as a result they may make unsuitable choices. Advice regulation addresses the cause of the problem by introducing a specific process for increasing consumers' understanding about what they need and hence the likelihood of an informed choice. The intuition for this approach is suggested by the comparative statics of the ex-post utility in the unregulated equilibrium in equation 7: other things being equal, utility increases when consumers' ignorance about their needs decreases. So if the intervention of a financial adviser can reduce consumer's ignorance, the consumer's ex-post utility may be improved. Providing advice uses resources. For example, the adviser must learn about the products in the market and must spend time assessing a consumer's needs. The issue is whether the utility gain from a more informed choice more than offsets the cost of advice.

An adviser does not observe a consumer's true needs but his skills enable him to be more precise about the consumer's true needs than the same consumer. This is different from some of the literature on credence goods where the expert makes a perfect assessment of the consumer's problem after the diagnosis, Dulleck and Kerschbamer (2001).

Furthermore, in retail financial services an adviser's effort is unobservable, as, for example, in Pesendorfer and Wolinsky (2000). So consumers cannot assess the quality of advice at the point of sale. In this situation, an adviser could always increase profits by reducing costs and therefore the quality of advice without this action being noticed.[25] At the same time, if all advisers behave in this way, the equilibrium quality of advice will be such that consumers may well be indifferent about taking advice.

Principal-agent theory suggests that controlling financial advisers' remuneration will be a crucial element to provide advisers with the appropriate incentives to maximise a consumer's welfare.[26] Identifying the optimal remuneration structure is a principal-agent problem beyond the scope of this analysis. However, one of the elements of the principal-agent problem is the agent's (adviser's) participation constraint. This constraint requires that the optimal form of remuneration covers the agent's costs. Here, I assume that advisers' only remuneration comes directly from consumers and that this covers the resource costs regardless of the recommendation provided.[27] This form of remuneration – and in particular the latter aspect – eliminates a bias to recommend a product because the commission rate is higher than for other products so there is no incentive for a systematic selling of unsuitable products. However, it would not eliminate (or reduce) the advisers' incentive to increase profits by reducing the quality of advice and costs. So the provision of advice needs to be regulated. I assume that a regulator sets a level for the quality of advice[28] and that its monitoring and the possibility of sanctions avoids the deterioration of the quality of advice.

---

[25] Dowd and Hinchliffe (2001) present anecdotal evidence about this type of behaviour in the UK during the late 1980s.

[26] This suggests that bringing forward the introduction of a new disclosure regime for retail financial products (see footnote 18) will not have necessarily prevented the pensions mis-selling (see footnote 6).

[27] This approach to remunerate advisers is different from the typical approach for remunerating advisers in the UK, which is based on a commission paid by the product provider and varies between products and with the level of contributions. The form of advice remuneration assumed here is also different because it excludes the possibility of product providers remunerating independent advisers by indirect means such as free computer equipment, seminars, etc. Some of these are allowed within certain boundaries in the UK, FSA(2000).

[28] This means regulating the conduct of sales. In the UK, this currently includes setting training requirements for advisors and requiring that advice is suitable.

These are strong assumptions but they are nevertheless useful because I am interested in comparing product regulation and advice regulation. If it can be shown that advice regulation dominates product regulation based on this simple approach, it would be then worth exploring how to devise an optimal remuneration structure that addresses the specific principal-agent issues of advice as well as its monitoring and enforcement.

All consumers are required to take advice before purchasing a product. However, some consumers might prefer not to take advice if they can obtain a product that matches their characteristics at a lower cost. This could be because the consumer can assess her own needs and search the market at a lower cost than the adviser. In this case, the social welfare identified here will represent a lower bound for the level of welfare.

Paying for advice can have additional welfare effects if some consumers are excluded from the market because they cannot afford the cost of advice. In the modelling exercise, I am setting this possibility aside because I am assuming that all consumers have the same level of income. Thus the analysis here answers the following question: assuming that advice is affordable, is social welfare improved by the provision of advice? If advice is not affordable, then the policy choices are an unregulated market and product regulation. More generally, a system of financial regulation does not have to be based on either advice or product regulation and I will offer some thoughts about this in Section 5.7.

The process of advice works as follows. An adviser does not observe a consumer's true type, the consumer's own z. I assume that an adviser has a set of skills and knowledge that allows him to assess the likely true type of a consumer, i.e. the likely value of z, more accurately than the consumer. So in period 1 the adviser meets the consumer and assesses her needs. The adviser then concludes that the consumer's true type is t, which is uniformly distributed in the interval

$$[z-\eta z, z+(1-z)\eta] \tag{17}$$

where $0 \le \eta \le 1$. A consumer maximises her ex-ante utility and accepts the adviser's recommendation, who then provides, $z^* = t$. So advisers are similar to consumers in that they identify a range containing the consumer's true type, see equation 1 above. If $\eta < \varepsilon$ then a consumer may be better off under advice regulation. The parameter η depends on the quality of advice, q. The assumption that the skills and knowledge of advisers result in the adviser having a better understanding of the consumer's true type than the same consumer can be modelled as

$$\eta = \varepsilon(1-q) \tag{18}$$

where ε represents consumers' ignorance about their circumstances and q denotes the quality of advice and ranges between [0, 1]. So the range of ex-post mistakes that may occur is always reduced as a result of advice. Note that as the quality of advice, q, approaches 1, the possibility of mistakes is eliminated.

In this case, the search cost represents the cost of advice. This is the time and effort that both the consumer and the adviser use to undertake the requisite analysis of consumers' needs and search for a suitable product. Suppose that the cost of advice depends on the quality of advice, q, and is independent of the consumer type (or product recommended). For example, let us assume that the cost of advice is c_2q^2, where c_2 is the cost per unit of quality of advice.

The average ex-post utility of a consumer of type z who gets advice of quality q and unit search cost, c_2, is $\overline{\upsilon}^{AR}(z,c_2,q)$. This is similar to the average ex-post utility developed for the unregulated equilibrium, equation 7, with a different cost of search (now, the cost of advice) and a different parameter for the level of ignorance (see equation 18). This is

$$\overline{\upsilon}^{AR}(z,c_2,q) = \overline{\upsilon}^{UR}[z, c_2q^2, \varepsilon(1-q)] =$$

$$= y - c_2q^2 - \frac{c_1}{2}\left[\varepsilon(1-q) + 2z\left(1-\varepsilon(1-q)\right)\right] - \frac{\alpha z^2}{2}\varepsilon(1-q) \tag{19}$$

Thus the quality of advice, q, affects the ex-post utility in two opposite ways. First, it increases utility – recall that $\overline{\upsilon}^{UR}(.)$ was a decreasing function of consumer's ignorance, ε. Second, it reduces utility by imposing additional costs – the costs of advice.

A welfare maximiser regulator will choose the quality of advice, q, that maximises social welfare. Social welfare is

$$w^{AR}(c_2,q) = \int_0^1 \overline{\upsilon}^{AR}(z,c_2,q)\,dz \tag{20}$$

Substituting from equation 19 and re-arranging gives

$$w^{AR}(c_2,q) = y - c_2 q^2 - \frac{c_1}{2} - \frac{\alpha\varepsilon(1-q)}{6} \tag{21}$$

I obtain the socially efficient level of advice by differentiating w^{AR} with respect to q and solving for it. This is

$$\hat{q} = \frac{\alpha\varepsilon}{12c_2} \tag{22}$$

This suggests that when the consumer's level of ignorance, ε, increases the socially optimal quality of advice, $\hat{q}$, increases. The effect of an increase in the unit cost of utility losses from unsuitable purchases, α, is the same. As one would also expect, if consumers have perfect knowledge about their (true) types, $\varepsilon = 0$, then the optimal quality of advice is zero. Finally, the result shows that the optimal quality of advice, $\hat{q}$, goes down when the cost of advice increases.

Substituting the optimal quality of advice, $\hat{q}$, into equation 19, I obtain an expression for the average ex-post utility of a consumer of type z, given the optimal quality of advice. Using equation 7 and after some re-arranging, I can express this

$$\bar{\upsilon}^{AR}(z,c_2,\hat{q}) = \bar{\upsilon}^{UR}(z,0,\varepsilon) + \frac{\hat{q}\,\varepsilon}{2}\left[c_1(1-2z) + \alpha\left(z^2 - \frac{1}{6}\right)\right] \tag{23}$$

where $\bar{\upsilon}^{UR}(z,0,\varepsilon)$ represents the ex-post utility in the unregulated equilibrium (equation 7) assuming that $s = 0$. The second term in equation 23 makes the analysis of the comparative statics less straightforward. These will be illustrated in the next section using some simulations. It nevertheless suggests that for ε close to zero the comparative statics properties will be very similar to those from the unregulated equilibrium – see $\hat{q}$ in equation 22 so the second term is effectively multiplied by ε to the square.

Let us consider now the difference between the average ex-post utility in the unregulated equilibrium (equation 7) and under advice regulation. This is

$$\bar{\upsilon}^{AR}(z,c_2,\hat{q}) - \bar{\upsilon}^{UR}(z,s,\varepsilon) = \frac{\hat{q}\,\varepsilon}{2}\left[\alpha z^2 - 2zc_1 + \left(c_1 - \frac{\alpha}{6}\right)\right] + s \tag{24}$$

If advice regulation increases ex-post utility, then the above expression must be positive. A sufficient condition for this to happen is that the expression in square brackets in 24 above must be positive. This requires [29]

$$4c_1^2 - 4\alpha\left(c_1 - \frac{\alpha}{6}\right) < 0 \tag{25}$$

After some re-arranging, this gives

$$\frac{c_1}{\alpha}\left(1 - \frac{c_1}{\alpha}\right) > \frac{1}{6} \tag{26}$$

Given that $\alpha > c_1$, c_1/α is smaller than one. The left-hand side of equation 26 reaches a maximum when this ratio is 0.5. So the condition is satisfied for values around 0.5 – in fact, between 0.22 and 0.78. This means a sufficient condition for advice regulation to result in ex-post utility being larger than in the unregulated equilibrium for all types of consumers is that the unit cost of products, c_1, and the unit cost of utility losses, α, are not too dissimilar.

Result 4: a sufficient condition for all consumers to be better off under advice regulation than in an unregulated equilibrium is that the unit cost of utility losses is about twice the unit cost of provision.

In addition, there may be cases where ex-post utility of some consumers (not all) under advice regulation is higher than in an unregulated equilibrium. This can be illustrated as follows. Suppose that $\alpha = 9c_1$. Equation 24 suggests that a consumer of type $z = 0$ will not necessarily be better off than in an unregulated equilibrium.[30] However, the consumer of type $z = 1$ will always be better off under advice regulation.

---

[29] This is an expression of the form a z^2 + b z + c. If advice regulation improves consumers' welfare regardless of a consumer's true type then the expression in square brackets must be positive for all z. This means that the equation will not have a solution and that the intercept – c – must be positive. We focus on the first condition because, in this case, if the first condition is satisfied the second one is also satisfied. The equation will not have a solution if the term inside the square root of the standard solution of a quadratic equation is negative.

[30] Assuming that $\alpha = 9c_1$ and $z = 0$ in equation 24 means that advice regulation will make such consumer better off than in an unregulated equilibrium if $s - \frac{1}{4}\varepsilon \hat{q} c_1 > 0$ which may or may not be satisfied.

Result 5: there are circumstances where advice regulation could make some consumer better off but not all of them.

5.6.3 A comparison of the various equilibria

So far I have compared product regulation and advice regulation against the unregulated equilibrium. The aim of this section is to compare the three equilibria using some simple simulations to understand how consumers' ignorance affects the preferences of different types of consumers over the three alternatives. (Recall that in the case of advice regulation and in the unregulated equilibrium I am, in fact, referring to the average ex-post utility of a consumer.)

Consider equilibria characterised by the parameters in Table 5.2.

Table 5.2: Illustrative values for the simulations

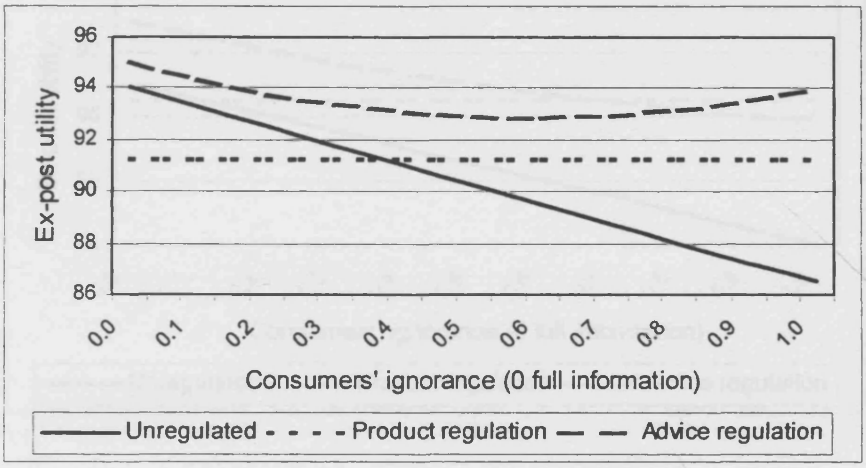
Parameter	Value
Income – y	100
Unit cost of utility losses from unsuitable purchases – α	20
Unit cost of provision – c_1	5
Unit cost of advice – c_2	1.5
Search and diagnosis costs (unregulated equilibrium)	1.0

The parameters have been chosen to illustrate the results of the model so the cost of advice is relatively low given the income so that advice is affordable. Also the parameters satisfy the requirement that $\alpha > c_1$. Finally, the ratio of c_1 to α is 0.25 so the results in the previous section suggest that advice regulation will make all consumers better off than in the unregulated equilibrium.

Figure 5.5 shows the ex-post utility for the type of consumer that needs the largest amount of provision ($z = 1$) for various levels of ignorance about her needs.

Figure 5.5: Ex-post utility of consumers of type $z = 1$

(consumers with the largest need of financial provision)



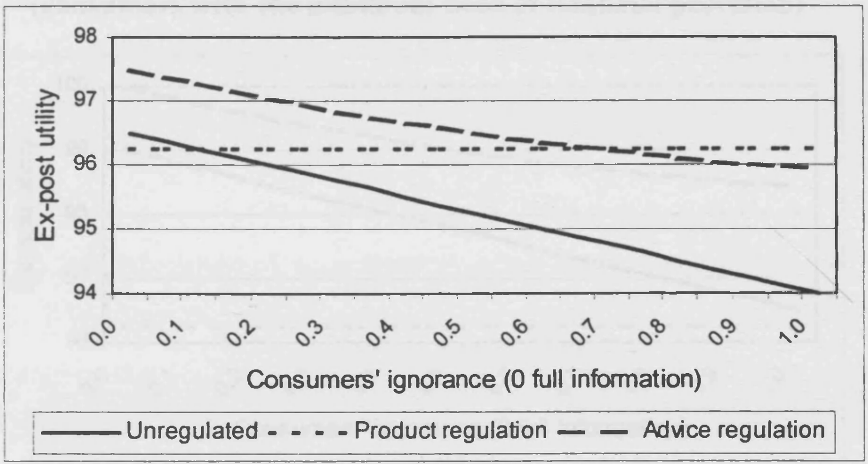
This figure shows that the ex-post utility of this type of consumer under a regime of advice regulation is larger than under any of the alternatives – product regulation or an unregulated equilibrium. In this case under advice regulation, the ex-post utility goes down when consumers' ignorance is low and then increases. The intuition is as follows. Advice regulation reduces the range of errors that consumers can make so on average there is a gain compared with the unregulated equilibrium. This gain has to be set off against the cost of advice. Thus the ex-post utility goes down when consumers' ignorance is low (low values of ε) because the gains from advice (i.e. losses avoided) are not enough to compensate for the cost of advice. This situation is then inverted when consumers' ignorance is large (large values of ε).[31]

This figure also suggests that given the unregulated equilibrium the gain from introducing advice regulation increases with the extent of consumers' ignorance. Finally, this figure also confirms results 2 and 3: the relationship between ex-post utility under product regulation and in the unregulated equilibrium depends on the level of consumer ignorance.

---

[31] For other values of z, the first effect described here is also observed unlike the second one so the ex-post utility of advice regulation may not follow the same pattern.

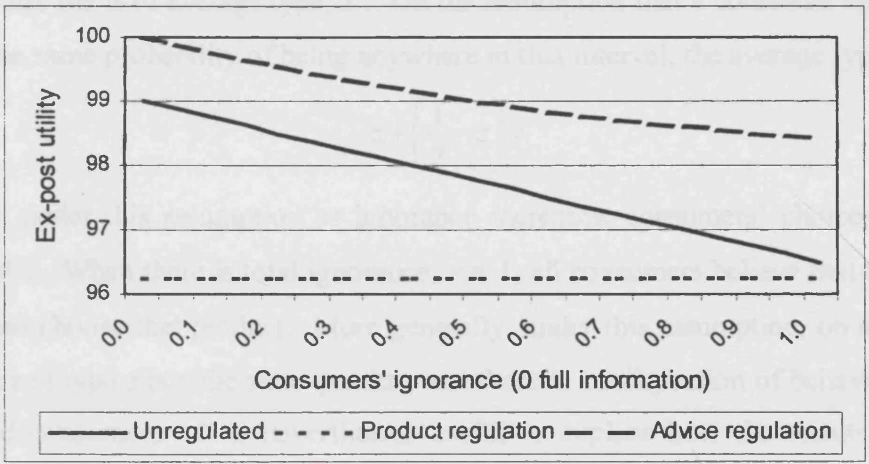
Figure 5.6: Ex-post utility of consumers of type $z = 0.5$
(consumers with the "average" need of financial provision)



Consider now the ex-post utility of a consumer that needs the average amount of financial provision (z = 0.5). This is Figure 5.6. The qualitative results are similar to those in Figure 5.5 but there is an interesting difference. For large levels of ignorance, this consumer prefers product regulation to advice regulation. The intuition is very simple. The regulator optimises product regulation. Given the parameters used here this is $\hat{z}^* = 0.75$. So under product regulation, this consumer gets a product that is reasonably close to her needs (z = 0.5). However, under advice regulation there is a chance that the consumer gets a product that is far from her needs. This effect together with the cost of advice results in this type of consumer being better off under product regulation when consumers' ignorance is large.

Finally, consider Figure 5.7. This represents the case of a consumer that needs the lowest amount of financial provision (z = 0). In this case, advice regulation is the preferable option regardless of the level of consumer's ignorance. Also, this type of consumer will prefer an unregulated equilibrium to product regulation even if consumers' ignorance is large. The intuition is simple. The regulated product results in a substantial provision being made ($\hat{z}^* = 0.75$) and this type of consumer needs a small financial provision. So the (average) utility loss in the unregulated equilibrium is smaller than the utility loss that results from purchasing the regulated product.

Figure 5.7: Ex-post utility of consumers of type $z = 0$
(consumers with the minimum need of financial provision)



These simulations suggest some conclusions. First, if consumers' ignorance is low, consumers may prefer advice regulation regardless of the consumer's (true) type. Second, if consumers' ignorance is large the preferences of consumers that need a lot of provision (large z) and the preferences of those consumers that need little (low z) are similar; both types of consumer prefer advice regulation to product regulation. Finally, when consumer ignorance is large product regulation may be the preferred option for those consumers that need 'average' amount of provision (e.g. z of 0.5). All in all when consumers' ignorance is large, it is unlikely that there will be one approach that will make all consumers better off. This is in line with the analytical results obtained in the previous sections.

5.7 Qualifications

The model presented in this paper is limited by the specific assumptions made. In this section, I explore how one could relax some of these assumptions.

First, the paper is based on the assumption that the consumer convinces herself that she is of type t and is unaware of her lack of information. This is different from more standard models of information asymmetry where the consumer is aware of her lack of information. For example, Wolinsky (1993) assumes that although the consumer does not know the type of her problem, the probability that a problem will be either large or small is common knowledge. In the case of the model of financial planning presented here it means that in the unregulated equilibrium the consumer incurs the diagnosis costs and becomes aware that her true type lies in the interval in equation 1. So a consumer believes that she is of average type, $\bar{t}$. On the assumption that a consumer believes that she has the same probability of being anywhere in that interval, the average type is

$$z+\left(\frac{1}{2}-z\right)\varepsilon \qquad (27)$$

Note that under this assumption as ignorance increases, consumers' choices converge towards 0.5. When there is total ignorance, $\varepsilon = 1$, all consumers believe that they are of type ½ and choose that product. More generally, under this assumption, on average, all consumers of type *z* buy the same product and there is no dispersion of behaviour within a class of consumers. It is nevertheless useful to explore how the results would be affected. The annex shows that also in this case different types of consumers will be better off under different regimes.

Another assumption of the model is that consumers' ignorance about their needs, ε, is a parameter. The model presented here could also be extended to explore learning, i.e. how consumers optimise the resources that they spend on financial planning in an unregulated equilibrium. In its simplest form, this could work as follows: consumers start with an exogenous level of ignorance, ε, and the degree of ignorance after the diagnosis is $\varepsilon(1-l)$ where l is the amount of learning, normalised between [0, 1]. Let us assume that in this case search costs can be represented by the following function $l^2 s$. Consumers' average ex-post utility with learning, $\bar{\upsilon}_L^{UR}$, is as in equation 7 with different parameters

$$\begin{aligned}\bar{\upsilon}_L^{UR} &= \bar{\upsilon}^{UR}(z, sl^2, \varepsilon(1-l)) = \\ &= y - sl^2 - \frac{c_1}{2}[\varepsilon(1-l) + 2z(1-\varepsilon(1-l))] - \frac{\alpha z^2}{2}\varepsilon(1-l)\end{aligned} \qquad (28)$$

Consumers then choose the optimal amount of learning about their needs. Taking the first derivative of the above and solving for l results in the optimal amount of learning, l^*. This is

$$l^* = \frac{\varepsilon}{4s}[\alpha z^2 - c_1(2z-1)] \qquad (29)$$

As one would expect, other things being equal, the optimal level of learning increases with the initial level of ignorance, and decreases with the cost of learning. The effect of a consumer's type on the optimal level of learning changes; it increases up to $z = \frac{c_1}{\alpha}$ and then decreases for larger values of z. This is just an illustration and the analysis could be extended further, for example, to explore how product choice will be affected by learning.

I have modelled separately advice regulation and product regulation. The conclusion that none of these alternatives will make all types of consumers better off is consistent with moves in recent years to combine elements of product regulation and advice regulation. From an analytical perspective, it suggests that it may be worth exploring the consequences of combining product regulation and advice regulation. This is not straightforward and it will require modelling the relationship between these forms of regulation. This could be undertaken within the context of a model of advice that takes into account the preferences of certain consumers for a regulated product and of other consumers for advised products. The former will require mainly search and the latter will require an assessment of their needs and search.

Finally, the model in this chapter could be extended by relaxing the assumption of marginal cost pricing to model the interaction between consumers' ignorance about their features and price setting behaviour. This would enable us to learn about the determinants of price dispersion, which is a feature of the market for retail financial services.[32] One possibility is that consumers incur different costs to search for a diagnosis as in Salop and Stiglitz (1977) model of search. Another issue that could be explored is the effect of firms' location. In the current model, consumers are located on a horizontal spectrum so location matters for firms. (A firm located next to the left (right) side of the spectrum is unlikely to have competitors to the left (right).) An alternative is to assume that consumers are located on a circle so that no location is a priori better than another as in Salop (1979) model of a circular city.

[32] For example, Alfon (2002) presents evidence of price dispersion in the market for personal pensions in the late nineties.

5.8 Concluding remarks

This chapter has introduced an additional perspective for retail financial products and their regulation based on consumers' ignorance about their needs and on the concept of financial planning. Typically, financial planning requires establishing the consumer's needs and searching the product that meets those needs. This is consistent with the regulator's view of the process of financial advice. In this context, financial advice can be regarded as a response to consumers' demand for financial planning.

The economic literature on retail financial products identifies the standard problem of asymmetric information between buyers and sellers and suggests that there may be an issue of credence goods. This literature focuses on situations where the consumer knows that she has a problem but is not sure about the nature of the problem. So in these cases there may be search for a diagnosis rather than just searching for a product.

I have developed a very simple model to start thinking about these issues in the context of retail financial services and their regulation that combines elements of search and of credence goods. So consumers are not sure about their needs and these cannot be perfectly established. The model focuses on the differences between consumers within a population and on the direct utility losses that arise from unsuitable purchases. If a consumer's needs are underestimated, she incurs two costs: the price of the product – proportional to the level of provision – and a direct utility loss. Alternatively, if her needs are overestimated, the cost is only financial.

In the unregulated equilibrium consumers are responsible for their financial planning. Consumers have certain beliefs about their needs where the size of the range of beliefs represent the ignorance about their needs. They incur a fixed search cost and convince themselves that they have a certain need and buy the relevant product. The result shows that consumers' ignorance about their needs reduces their ex-post utility even when products are priced at marginal cost.

I then use this framework to compare two extreme approaches to improve consumers' welfare: product regulation and advice regulation. Product regulation overcomes consumers' ignorance about their needs by limiting choices to one product whose features are selected by a welfare maximiser regulator. Thus almost all consumers incur direct utility losses, which may be smaller than in an unregulated equilibrium. However, there is no need for a diagnosis and the search costs associated with it are almost avoided.

The other alternative is advice regulation. An adviser is an expert who can learn the consumer's initial range of beliefs about her needs and can use his knowledge and experience to assess more accurately the needs and recommend a product. So advice is not perfect in the sense that it will not lead to a perfect identification of consumers' needs. However, advice opens a set of principal-agent issues. These are dealt with in the model by assuming that advisers are remunerated only by consumers (an amount that covers costs and allows a normal profit to be earned) and that there is no systematic mis-representation (so consumers may still be recommended an unsuitable product but not systematically). A welfare maximiser regulator sets the quality of advice that maximises the sum of all consumers' utility. The cost of search now includes the relevant cost of the advice.

The main result of these alternative approaches to improve consumers' welfare is that different types of consumers will be better off under different regimes. In its simplest form, this means that some consumers obtain a larger ex-post utility from product regulation than from the unregulated equilibrium (Results 2 and 3). A similar result is obtained when product regulation, advice regulation and the unregulated equilibrium are compared using some simple simulations.

This work is far from being a complete model of financial planning and the model can be extended in various ways, for example, by relaxing the assumption of marginal cost pricing. The main result of this work is broadly consistent with moves to combine advice regulation and product regulation in the UK. From an analytical perspective, it suggests that it may also be worth exploring the consequences of combining product regulation and advice regulation.

Annex 5.1: An alternative assumption about consumers' ignorance

In the main text, I have assumed that the consumer is not certain about her needs, which I denoted by *z*, and that she forms the belief that she is of type *t* which is uniformly distributed in the interval

$$[z-\varepsilon z, z+(1-z)\varepsilon] \qquad \text{(A1)}$$

This is different from more standard models of information assymetry where the consumer is aware of her lack of information. For example, Wolinsky (1993) assumes that although the consumer does not know the type of problem that she has, the probability that a problem will be either large or small is common knowledge. In the case of the model presented here this would mean that the consumer is aware that her true type lies in the interval in A1. If so, on average, a consumer believes that she is of average type, $\bar{t}$. On the assumption that a consumer believes that she has the same probability of being anywhere in that interval, the average type, $\bar{t}$, is

$$z+\left(\frac{1}{2}-z\right)\varepsilon \qquad \text{(A2)}$$

In this Annex, I explore how this alternative assumption about the search process affects the main result.

Consider first the unregulated equilibrium. This is exactly the same as in Section 5.5: there is a variety of products, consumers take responsibility for their financial planning and there is marginal cost pricing. In period 1, a consumer of type *z* incurs a search cost, *s*, and forms a belief about her type, $\bar{t}$ as in A2. The consumer maximises the ex-ante utility from purchasing product *z**, *u* (*z*, *s*, *z**), and so chooses z* = $\bar{t}$. In period 2, the consumer's true type *z* is revealed and the consumer experiences her ex-post utility given the choice of $\bar{t}$ in period 1, $\upsilon_A^{UR}(z,s,\varepsilon)$. (I use throughout the Annex the subscript "A" to distinguish from the expressions in the main text.) As discussed in Section 5.4.2, the ex-post utility is

$$\upsilon_A^{UR}(z,s,\varepsilon)=y-s-c_1\bar{t}-\alpha Max(z-\bar{t},0) \qquad \text{(A3)}$$

Substituting equation A2 into A3 and re-arranging we obtain,

$$
\upsilon_A^{UR}(z,s,\varepsilon)=\begin{cases} y-s-\dfrac{c_1}{2}\left[\varepsilon+2z(1-\varepsilon)\right] & z\le\dfrac{1}{2} \\ y-s-\dfrac{c_1}{2}\left[\varepsilon+2z(1-\varepsilon)\right]-\alpha\varepsilon\left(z-\dfrac{1}{2}\right) & z>\dfrac{1}{2} \end{cases} \tag{A4}
$$

The comparative statics are the same as those of the unregulated equilibrium in the main text – equation 7 – and Result 1 still applies.[33] It can also be shown that the level of ex-post utility under this assumption is higher than the level of ex-post utility in equation 7. The intuition is that the consumer uses all the information available – the distribution of possible types in A1 – to form her beliefs about her true type at the point of sale.

Product regulation is based on the information available to the government about the distribution of consumers' true types. This is not affected by assumptions about what consumers know so the regulated product remains the same as in the main text.

As in the main text, I want to compare the ex-post utility in the unregulated equilibrium with the ex-post utility under product regulation. The aim is to verify that the result that different consumers will be better off under different regimes is robust to this change in the assumptions about consumers' information. I illustrate this focusing on one value of consumers' ignorance, $\varepsilon = ½$. The table below shows the level of consumer ignorance for different values of z.

Table A5.1: Comparison of unregulated equilibrium and product regulation

z	**Unregulated equilibrium (equation A4)** $\varepsilon = ½$	**Product regulation (equation 11 in the main text)**
0	$y-s-\dfrac{c_1}{4}$	$y-c_1\left(1-\dfrac{c_1}{\alpha}\right)$
1	$y-s-\dfrac{3}{4}c_1-\dfrac{\alpha}{4}$	$y-c_1\left(1-\dfrac{c_1}{\alpha}\right)-c_1$

[33] Note that for $z > ½$, $\dfrac{\partial\upsilon^{UR}}{\partial\varepsilon} < 0$ because we also assume here that $\alpha > c_1$.

Suppose that $\alpha = 5c_1$ (recall that we have assumed that $\alpha > c_1$) and let us consider the situation of a consumer of type $z = 0$. If so, ex-post utility in the unregulated equilibrium exceeds the ex-post utility under product regulation if search costs in the unregulated equilibrium are sufficiently low.[34] The situation of a consumer of type $z = 1$ is different. The results in the table suggest that if $\alpha = 5c_1$, she will be better off under product regulation regardless of the search costs in the unregulated equilibrium.[35]

Consider now the case where financial planning takes the form of regulated advice. The process of advice is the same as in Section 5.6.2. In brief, an adviser does not observe a consumer's true type. The adviser has a set of skills that enables him to assess that the consumer's true type is $[z - \eta z, z + (1 - z)\eta]$ where $\eta = \varepsilon(1 - q)$ and q denotes the quality of advice. (The latter is normalised as a value between [0, 1].) The adviser also believes that the consumer's true type is uniformly distributed on that interval. So on average, he advises a consumer of unknown type z that her true type is

$$z + \left(\frac{1}{2} - z\right)\varepsilon(1 - q) \tag{A5}$$

Note that as q is closer to 1 the adviser's recommendation is closer to the consumer's true type.

The average ex-post utility of a consumer of type z who gets advice of quality q is $\upsilon_A^{UR}(z, c_2, q)$. As in the main text, this is similar to the utility in the unregulated equilibrium in equation A4 with a different diagnosis cost and a different level of ignorance

$$\upsilon_A^{AR}(z, c_2, q) = \upsilon_A^{UR}[z, c_2 q^2, \varepsilon(1 - q)] = \tag{A6}$$

$$= \begin{cases} y - c_2 q^2 - \frac{c_1}{2}\left[\varepsilon(1-q) + 2z\left(1 - \varepsilon(1-q)\right)\right] & z \le \frac{1}{2} \\ y - c_2 q^2 - \frac{c_1}{2}\left[\varepsilon(1-q) + 2z\left(1 - \varepsilon(1-q)\right)\right] - \alpha\varepsilon(1-q)\left(z - \frac{1}{2}\right) & z > \frac{1}{2} \end{cases}$$

---

[34] In fact, in this particular case, the condition about the search cost is: $s < \frac{11}{20}c_1$.

[35] I find that the ex-post utility in the unregulated equilibrium will exceed the ex-post utility under product regulation if the search costs are negative. The latter is unfeasible.

A welfare maximiser regulator chooses the quality of advice, q, that maximises social welfare. Social welfare is

$$w_A^{AR}(c_2,q)=\int_0^1 \upsilon_A^{AR}(z,c_2,q)\,dz=$$
$$=\int_0^1\left[y-c_2q^2-\frac{c_1}{2}\left[\varepsilon(1-q)+2z\left(1-\varepsilon(1-q)\right)\right]\right]dz-\int_{½}^1 \alpha\varepsilon(1-q)\left(z-\frac{1}{2}\right)dz \qquad \text{(A7)}$$

Re-arranging gives the following expression for social welfare

$$w_A^{AR}(c_2,q)=y-c_2q^2-\frac{c_1}{2}-\frac{1}{8}\alpha\varepsilon(1-q) \qquad \text{(A8)}$$

We obtain the socially efficient level of advice by differentiating $w_A^{AR}(c_2,q)$ with respect to q and solving for it. This is

$$\hat{q}=\frac{\alpha\varepsilon}{16c_2} \qquad \text{(A9)}$$

This is very similar to the optimal quality that we derived earlier. The difference is that beforehand the number in the denominator was 12 so other things being equal now the optimal quality of advice is smaller. The rationale for this is as follows. Regulators maximise welfare taking into account the interaction between consumers' ignorance, the cost of advice, etc. If, on average, advisers recommend the average product that may suit the consumers, the optimal quality that a regulator needs to impose to maximise welfare will be smaller.

Substituting the optimal quality of advice from equation A9 into equation A6, we obtain an expression for the average ex-post utility of a consumer of type z. Using equation A4 and after some re-arranging the ex-post utility under advice regulation given the optimal quality of advice can be expressed as

$$\upsilon_A^{AR}(z,c_2,\hat{q})=\begin{cases}\upsilon_A^{UR}(z,0,\varepsilon)+\varepsilon\hat{q}\left(c_1\left(\frac{1}{2}-z\right)-\frac{\alpha}{16}\right) & z\le\frac{1}{2}\\ \upsilon_A^{UR}(z,0,\varepsilon)+\varepsilon\hat{q}\left((\alpha-c_1)\left(z-\frac{1}{2}\right)-\frac{\alpha}{16}\right) & z>\frac{1}{2}\end{cases} \qquad \text{(A10)}$$

where $\upsilon_A^{UR}(z,0,\varepsilon)$ is equation A4 assuming that $s=0$.

I want to compare this ex-post utility with the level that results from the unregulated equilibrium. The difference in ex-post utilities is

$$\upsilon_A^{AR}(z,c_2,\hat{q})-\upsilon_A^{UR}(z,s,\varepsilon)=\begin{cases}\varepsilon\hat{q}\left(c_1\left(\frac{1}{2}-z\right)-\frac{\alpha}{16}\right)+s & z\le\frac{1}{2}\\ \varepsilon\hat{q}\left((\alpha-c_1)\left(z-\frac{1}{2}\right)-\frac{\alpha}{16}\right)+s & z>\frac{1}{2}\end{cases}\qquad \text{(A10)}$$

So if the relevant expression in equation A10 is positive the consumer will be better off under advice regulation. Let us suppose that $\alpha=9c_1$ to explore this (recall that we have assumed that $\alpha>c_1$). Consider the situation of a consumer of type z = 0. If so, ex-post utility in the unregulated equilibrium exceeds the ex-post utility under advice regulation if search costs in the unregulated equilibrium are sufficiently low.[36] The situation of a consumer of type z =1 is different. If $\alpha=9c_1$, advice regulation will make this consumer better off regardless of the search cost in the unregulated equilibrium.

Finally, I have re–run the simulations in Section 5.6.3 using the same assumptions to compare the three regimes under the alternative assumption made here. Figure A5.1, A5.2 and A5.3 show the ex-post utility for all possible levels of consumers' ignorance for a type of consumer. The charts confirm the analytical intuition and, like in the main text, they show that different consumers would be better off under different regimes.

Figure A5.1: Ex-post utility of consumers of type z = 1
(consumers with the maximum need of financial provision)

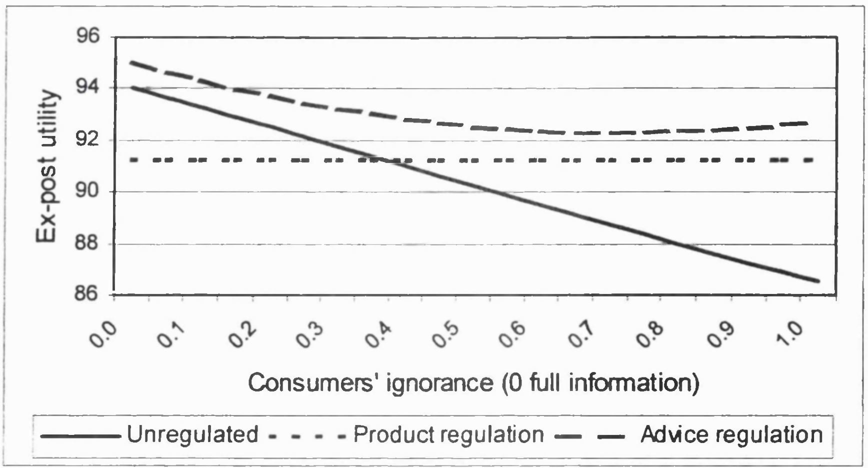


Figure A5.2: Ex-post utility of consumers of type $z = 0.55$ (consumers with the "average" need of financial provision)

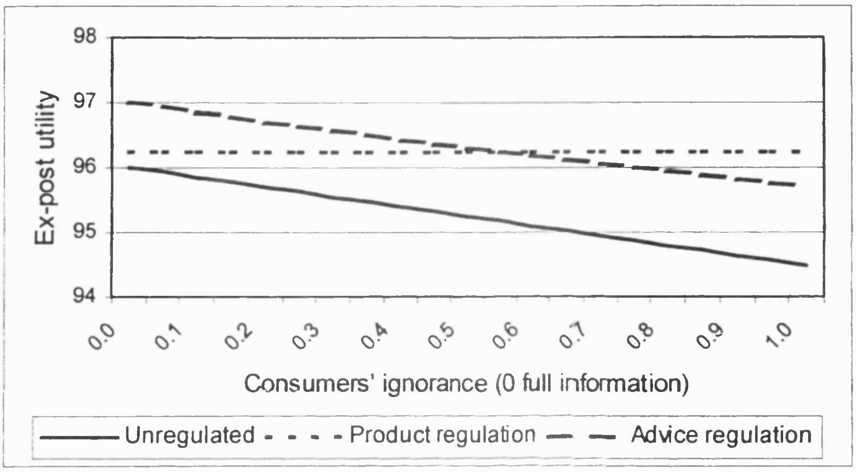
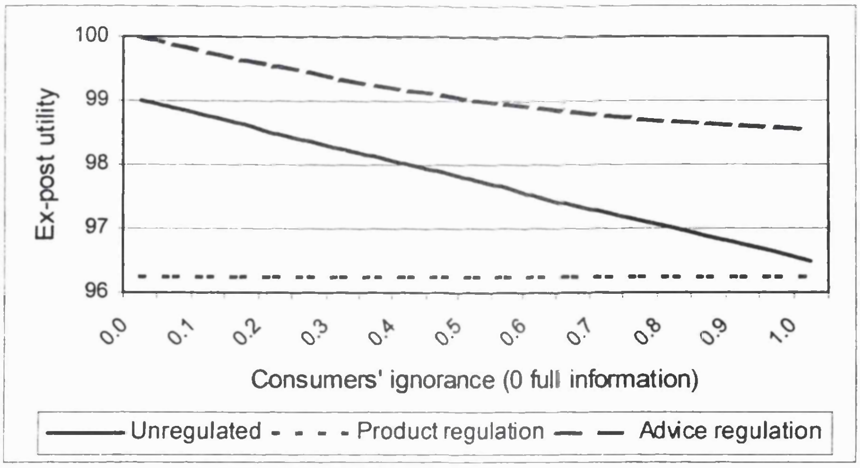


Figure A5.3: Ex-post utility of consumers of type $z = 0$ (consumers with the minimum need of financial provision)



---

[36] In fact, in this particular case, the condition about the search cost is: $s < \frac{c_1 \hat{q} \varepsilon}{16}$ where $\hat{q}$ is the optimal quality of advice from equation A9.

References

Alfon, I. (2002), "*To switch or not to switch, that's the question. An analysis of the potential gains from switching pension provider*", FSA Occasional Paper 18, September

Akerlof, G. (1970), "The market for 'lemons': quality uncertainty and the market mechanism", *Quarterly Journal of Economics*, vol. 84, pp. 488 – 500

Banks, J., R. Blundell and S. Tanner (1998), "Is there a retirement savings puzzle?", *American Economic Review*, vol. 88, pp. 769 – 788

Benston, G. (1998), "*Regulating financial market: a critique and some proposals*", Hobart Paper no. 135, Institute of Economic Affairs, London

Briault, C. (2002), "*Revisiting the rationale for a single national financial services regulator*", LSE Financial Markets Group, Special Paper No. 135, February

d'Aspremont, C., J. J. Gabszewicz and J.-F. Thisse (1979), "On Hotelling's 'Stability in competition'", *Econometrica*, vol. 47, pp. 1145 – 1150

Darby, M. and E. Karni (1973), "Free competition and the optimal amount of fraud", *Journal of Law and Economics*, vol. 16, pp. 67 – 88

Davies, H. (1998), "*Why regulate?*", 1998 Henry Thornton Lecture, City University Business School (available from http://www.fsa.gov.uk/pubs/speeches/index-1998.html)

Dowd, K. and J. Hinchliffe (2001), "Paternalism fails again. The sorry story of the Financial Services Act", in K. Dowd, *Money and the market*, Routledge, London

Dulleck, U. and R. Kerschbamer (2001), "*On doctors, mechanics and computer specialists or where are the problems with credence goods*", Working paper 101, Department of Economics, University of Vienna (available from http://www.univie.ac.at/Wirtschaftswissenbschaften/)

Financial Services Authority (2000), "*The conduct of business sourcebook*", Consultative Paper 45, February

Goodhart, C. A. E., P. Hartmmann, D. Llewellyn, L. Rojas-Suarez, L. and S. Weisbrod (1998), *Financial regulation. Why, how and where now?*, Routledge, London

Gravelle, H. (1994), "Remunerating information providers: commission versus fees in life assurance", *Journal of Risk and Insurance*, vol. 61, pp. 425 – 457

HM Treasury (1998), "*Making savings easy*", May (available from http://www.hm-treasury.gov.uk/Documents/Financial_Services/Savings/fin_sav_index.cfm?)

Harvard Business Review (2002), "Spotting patterns on the fly. A conversation with birders David Sibley and Julia Yoshida", *Harvard Business Review*, November, pp. 45 - 49

Hotelling, H. (1929), "Stability in competition", *Economic Journal*, vol. 39, pp. 41 – 57

Ingledew, S. (2001), "*Identifying and satisfying client needs. Study course FP3*", Chartered Insurance Institute, fourth edition, London

James, K. R. (2000), "*The price of investing*", FSA Occasional Paper 6, February

James, K. R. (2002), "*Waiting for Ariadne: a suggestion for reforming financial services regulation*", Centre for the Study of Financial Innovation, London, July

Johnson, P. (2000), "*New directions in financial regulation: CATs, Stakeholders and consumer education*", FSA Occasional Paper 11, September

Kreps, D. M. (1990), *A course in microeconomic theory*, Harvester Wheatsheaf, Hemel Hempstead

Llewellyn, D. (1999), "*The economic rationale for financial regulation*", FSA Occasional Paper 1, April

Martin, S. (2002), *Advanced industrial economics*, second edition, Blackwell, Oxford

Nelson, P. (1970), "Information and consumer behaviour", *Journal of Political Economy*, vol. 78, pp. 311 – 329

Office of Fair Trading (1997), "*Consumer detriment under conditions of imperfect information*", Research Paper, August

Ottaviani, M. and P. N. Sorensen (2001), "*Professional advice: the theory of reputational cheap talk*", London Business School, December (available from http://www.london.edu/mottaviani/)

Page, A. C. and R. B. Ferguson (1992), *Investor protection*, Butterworth

Personal Investment Authority (2001), "*Survey of the persistency of life and pension policies*", October

Pesendorfer, W. and A. Wolinsky (2000), "Second opinions and price competition: inefficiency in the market for expert advice" (available from http://www.princeton.edu/~pesendor/work/asher.pdf)

Prendergast, C. and L. Stole (1996), "Impetuous youngsters and jaded old-timers: acquiring a reputation for learning", *Journal of Political Economy*, vol. 104, pp. 1105 – 1134

Sandler, R. (2002), "*Medium and long-term retail savings in the UK*", The Stationery Office, London (available from http://www.hm-treasury.gov.uk/)

Salop, S. (1979), "Monopolistic competition with outside goods", *Bell Journal of Economics*, vol. 10, pp. 141 – 156

Salop, S. and J. E. Stiglitz (1977), "Bargains and ripoffs: a model of monopolistically competitive price dispersion", *Review of Economic Studies*, vol. 44, pp. 493 – 510

Schmalensee, R. (1978), "Entry deterrence in the ready-to-eat breakfast cereal industry", *Bell Journal of Economics*, vol. 9, pp. 305 – 327

Securities and Investment Board (1990), "*Regulation of the marketing of investment product and services*", Consultative paper

Sinclair, P. (1998), "*Why do the poor save so little?*", Birmingham University Economics Discussion Paper, No. 98–20

Spencer, P. D. (2000), *The structure and regulation of financial markets*, Oxford University Press, Oxford

Stiglitz, J. E. (1977), "Monopoly, non-linear pricing and imperfect information: the insurance market", *Review of Economic Studies*, vol. 44, pp. 407 – 430

Thisse, J. F. and G. Norman (1994), "Introduction", in J. F. Thisse and G. Norman (eds.), *The economics of product differentiation*, Vol. I, Edward Elgar Publishers, Cheltenham.

Tirole, J. (1988), *The theory of industrial organisation*, MIT Press, Cambridge, Massachusetts

Wolinsky, A. (1993), "Competition in a market for informed experts' services", *Rand Journal of Economics*, vol. 24, pp. 380 – 398