NEW ECONOMIC GEOGRAPHY: MULTIPLE EQUILIBRIA, WELFARE, AND POLITICAL ECONOMY

Frédéric L. Robert-Nicoud

The London School of Economics and Political Science

A thesis submitted for the Ph.D. degree in Economics

London, July 2002

Abstract

This thesis contributes to the body of research known as the new economic geography. According to this paradigm, increasing returns to scale at the firm level, monopolistic competition, and transportation costs interact in shaping the spatial distribution of economic activity.

The introductory chapter lays out the motivation of this thesis and puts it into the perspective of the existing literature.

Chapter 1 introduces a typical model of new economic geography: the nature of the agglomeration and dispersion forces it displays is recurrent in this body of research; the model also displays multiple equilibria. The welfare properties of these equilibria are also analysed.

Chapter 2 completely characterizes the set of equilibria of a wide range of models that are the quintessence of the new economic geography paradigm. The model of chapter 2 is shown to share the qualitative features of these models.

Chapter 3 integrates a simple version of the model chapter 2 within a political economy framework. The welfare analysis of chapter 2 provides the motivation for this theoretical exercise. Chapter 4 seeks to provide an answer to the important but thus far neglected question of what is the mechanism that actually determines the magnitude policies that seek to affect the equilibrium spatial allocation of industries. The geography model is integrated in a fully specified political economy process of policy selection.

Chapter 4 extends the model of chapter 2 to deal with the issue of the 'fragmentation' of the production process when new economic geography forces are at play.

Finally, the analysis of chapter 5 contributes to the growing literature on the labour market imperfections as a driving force for agglomeration. In particular it shows how the hold-up problem can be softened or worsened by the cluster of industries using workers with similar skills.

Table of Contents

Abstract 2

Table of Contents 3

Tables and Figures 5

Acknowledgments 6

MOTIVATION AND LITERATURE REVIEW 8

CHAPTER 1. A SIMPLE MODEL OF AGGLOMERATION WITH VERTICAL LINKAGES AND PERFECT CAPITAL MOBILITY 32

1.1. Introduction 32

1.2. The basic model 33

1.3. Agglomeration and dispersion forces 40

1.4. Symmetric regions and vertical linkages: The FCVL model 43

1.5. Asymmetric regions in the absence of linkages: The FC model 47

1.6. Welfare analysis 50

1.7. Equity and Efficiency in the Footloose Capital Model 50

1.8. Equity and efficiency in the FCVL model 56

1.9. Concluding remarks 61

Appendix 62

CHAPTER 2. THE STRUCTURE OF SIMPLE 'NEW ECONOMIC GEOGRAPHY' MODELS 63

2.1. Introduction 63

2.2. A natural state space for migration-based models 65

2.3. On the number of steady-states and stability analysis 74

2.4. Concluding remarks 78

Appendix 79

CHAPTER 3. THE 'VOTE-MARKET EFFECT', OR, THE POLITICAL ECONOMICS OF INDUSTRY LOCATION 83

3.1. Introduction 83

3.2. The augmented FC model 86

3.3. The 'subsidy effect' on location 91

3.4. The voting model 94

3.5. The vote-market and the net-market effects 100

3.6. Concluding comments 104

CHAPTER 4. GEOGRAPHY AND COMMUNICATION COSTS 107

4.1. Introduction 107

4.2. The simple extended model 110

4.3. Communication costs and 'multinationals' 119

4.4. Preliminary results and break point 123

4.5. Sustain points 125

4.6. 'History of the world, part II'? 130

4.7. Wage inequalities 132

4.8. Conclusion 138

Appendix 140

CHAPTER 5. THE HOLD-UP PROBLEM, INDUSTRY-SPECIFIC SKILL ACQUISITION, AND FUNCTIONAL AGGLOMERATION 144

5.1. Introduction 144

5.2. A simple model 148

5.3. Entrepreneur investment 154

5.4. Core-periphery or dispersion? 155

5.5. Discussion 155

CONCLUSION 157

REFERENCES 161

Tables and Figures

Table 1. Three classes of Economic geography models 17

Figure 1-1: Dispersion and agglomeration forces at the symmetric equilibrium 44

Figure 2-1: Instantaneous Equilibrium 71

Figure 2-2: The Tomahawk Diagram 75

Figure 2-3: Proving the $\phi^{break} > \phi^{sust}$ 77

Figure 3-1. Subsidies and equilibrium location 92

Figure 4-1: Transportation vs. communication costs, 1940-1990 109

Figure 4-2: Dispersion and agglomeration forces at the core-periphery equilibrium ... 114

Figure 4-3: The 'Double Tomahawk' diagram 118

Figure 4-4. Hiring workers abroad 122

Figure 4-5. The 'Loudspeaker' diagram 129

Figure 4-6. Trade liberalisation and welfare 137

Figure 4-7: Break points (parameter values: $\alpha=\mu$, $\sigma=5.31$, $\theta=.3$) 142

Acknowledgments

I have spent four years at the London School of Economics. I could hardly overstate how much I have benefited from this experience, both from an intellectual and a human point of view.

I am especially grateful to Anthony Venables for his teaching and supervision. Stephen Redding was also always willing to give some of his time to talk about research. Both of them provided me with great food for thought.

Richard Baldwin deserves quite a lot of credit as my former supervisor at the Graduate Institute of International Studies. For one, he encouraged me to move to the LSE. Since then, his knowledge and thinking have kept benefiting me, both as an informal advisor and as a co-author. His kind and active support is gratefully acknowledged.

Quite a few people have read parts of this work and, besides the people already mentioned, I would like to thank –without implicating– Fredrik Gallo, Philippe Martin, Henry Overman, Federica Sbergami, and Jacques Thisse for comments and suggestions that helped clarify the exposition of some of these chapters.

I have also benefited from the collaboration of several people while working on related issues, not least Rikard Forslid, Philippe Martin, Niko Matouschek, and Gianmarco Ottaviano. Above all, Federica Sbergami deserves much credit for her coping with my fast-changing mood.

The LSE Economics Department, the Centre for Economic Performance, the graduate Institute of International Studies, and the Janggen-Pöhn foundation all contributed in a materialistic ways to my studies at the LSE. I am extremely grateful to these institutions and to the people who work in their background.

On a more personal touch my time in London provided me with the opportunity to get to know quite a few people. Among my flatmates, I would like to thank Alain Chenaux and Carlo Fiorio for the fantastic time spent together. I discovered London and some of its most interesting aspects with the former; the latter noticeably succeeded in living with me for three years. No girlfriend has ever matched this achievement. The LSE would not have been such a pleasant working environment –and London not such a pleasant place to live in– without the company of Heski Bar-Isaac, Virginie Blanchard, Augustin de-Coulon, Mara Gorli, Pete Longyear, Maryla and Wojtek Maliszewski, Niko Matouschek, Nikolas Müller-Plantenberg, Amedeo Poli, Paolo Ramezzana, Annamaria Reforgiato, and Richard Walker, among many others. Whatever topic we picked, serious or not, it was always a pleasure to exchange views and ideas with all of them.

Last, but in no way least, I also seize this opportunity to thank all my family and friends in Geneva and elsewhere whose company always provided a welcome break on my visits. Their indirect contribution to the kind of achievement this thesis provides is invaluable. Alain and Cathy Chenaux, Barbara Hofer, Pascal Jeannerat, François-Xavier and Valérie Rouxel, Dimitri Ruiz, and Cédric Vaezi even made it to London, and I take this as a strong sign of friendship.

While I was working on this project and to this day, my brother travelled around the world as a volunteer for an NGO; my parents kept giving us moral support and affection; and a good friend of mine was fighting a vicious form of cancer. This thesis is dedicated to them.

MOTIVATION AND LITERATURE REVIEW

This thesis contributes to the theoretical literature on spatial economics, also known as economic geography. This strand of the economic literature addresses the specific questions of why economic activities take place in a discrete number of well-defined places and how space does affect competition and various economics activities. Within this area of the economic theory, the so-called 'New Economic Geography' (NEG) paradigm concentrates on the interaction among increasing returns at the firm level, monopolistic competition, and transportation costs or the state of infrastructures. The principle here is that consumers' taste for variety, intermediate inputs differentiation, or both, result in agglomeration economies. These arise as the result of pecuniary externalities. The importance of increasing returns can hardly been overstated: with non-increasing returns the existence of transportation costs results in 'backyard capitalism'.

The resulting trade-off between transportation costs and increasing returns gives rise to a cumulative process typical in models of monopolistic competition. In a spatial setting, this cumulative process results in the cluster of economic activities in a few locations. The scope of these clusters can vary from the urban level, in which case these clusters are called cities, to the broadest, international level. In an international framework, the existence of a limited amount of these clusters draws a line between developed and developing countries, with the former specialised in production processes characterized by large economies of scale and the latter being specialised in more traditional, less sophisticated industries. The issues raised in this thesis are regional in scope; hence these clusters should be understood as industrial clusters.

What do we learn from those models we did not know from neoclassical of international and interregional trade? Quite a lot, as it turns out. The neoclassical trade theory teaches us that the conjunction of constant returns to scale and free trade or free factor mobility (or a mix of both) leads to the convergence of income. This is the famous factor price equalisation theorem. In sharp contrast the NEG aims to explain how the product mix of seemingly similar regions diverges endogenously as the result of the aforementioned cumulative process. Moreover, a dynamic interpretation of these models can account, firstly, for lock-in phenomena, i.e. when initial conditions determine the current outcome beyond and above comparative advantage. Secondly these models can account for the possibility of self-fulfilling expectations and, finally, for the catastrophic changes in the spatial configuration following minor changes in the environment.

This thesis is entitled *New economic geography: Multiple equilibria, welfare, and political economy.* As the title suggests, the main contribution of the present work is threefold.

Firstly one may wonder whether agglomeration is optimal and one may want to identify those who gain and those who lose out from agglomeration. The first point is about the efficiency of the free-market outcome, the latter about the equity of this outcome. Chapter 1 tackles both of these issues in a typical NEG framework.

Secondly this work aims at completing the characterisation of the backbone models of the NEG paradigm. The idea here is to fill a theoretical gap. Despite the simplicity of their framework, NEG models are extraordinary troublesome to manipulate. In particular one must rely on numerical simulations to infer their equilibrium properties. The seminal 'Core-Periphery' model by Krugman (1991a,b) is a case in point. Numerical simulations of this model show three facts:

1. There are no more than five steady-states (and no more than three interior steady-states). In other words, this model displays multiple equilibria;
2. Among them the symmetric steady-state (in which industry is uniformly spread across regions) always exists, but is not always stable;
3. When they exist, both asymmetric interior steady states are always unstable.

At this level of generality it is sufficient to point to the fact that these features form a regular pattern in this paradigm. And yet, more than ten years on no algebraic proof to points 1 and 3 has been put forward. The search of an analytical proof for the remaining points represents a formidable task. In fact most of the third chapter of this thesis is devoted to it. Additionally the analysis in Chapter 2 shows why these facts are regular across widely different models: it turns out that most of these can be written in a more natural state space in which they are all isomorphic. As a corollary, the proof of the three points above is valid for a whole family of models.

The third contribution of this thesis is perhaps the most innovative in this study. It integrates a simple model of economic geography within a political economy framework. The motivation for this exercise starts with the observation that the spatial allocation of industry is an important determinant to the welfare of different groups of people, as Chapter 1 points out. Several contributions show how different policy instruments affect the spatial distribution of industries. But an important question has been put aside in the literature thus far: 'What is the mechanism that actually determines the magnitude of infrastructure spending or of the production subsidy?' Chapter 3 seeks to provide an answer in an economic geography model, which it extends to include a fully specified political economy process of policy selection (a probabilistic voting model to be precise).

The final two chapters of this thesis deal not with core analytical questions, but instead with extensions of geography models. Chapter 4 deals with the issue of the 'fragmentation' of the production process when NEG forces are at play. The idea here is that the presence of decreasing returns in the background sector implies that wages are higher in more industrialised nations, which squares well with empirical evidence. In such a context, firms trade off the benefits of agglomeration economies associated with locating in the more industrialised nation with the low wages of the less developed country. It is assumed that low communication costs allow firms to outsource some of the routinised tasks to the less industrialised nation. This attenuates one of the dispersion forces and, as a result, sustains the viability of the industrialised base. This has an ambiguous effect on each nation's workers' welfare, as we shall see.

The last contribution of this thesis is the topic of Chapter 5. The analysis therein goes beyond the NEG and contributes to the growing literature on the labour market imperfections as a driving force for agglomeration. In particular it shows how the hold-up problem can be softened or worsened by the cluster of industries. The main message of this chapter is thus that agglomeration and labour market pooling (or absence thereof) emerges as the outcome of the interaction between market power (or lack of), the non-verifiability of some investment, and the specificity of this investment to the relationship –which itself depends upon the location decision of various agents. The mechanisms it emphasizes and the empirical predictions that can be derived from it set the theoretical model of Chapter 5 apart from the NEG, which it seeks to complement.

This finishes my discussion of the aims and scope of the thesis. I will now review the relevant literature in more depth. I start by briefly reviewing the literature that is relevant to the whole context of this work. Then I review contributions more specifically related to each chapter, explaining in turn how each of them contributes to the literature. References that are more specific to a given chapter's idiosyncrasies are referred to in the introductory section of the chapter itself.

General literature review

Before briefly reviewing some of the major contributions within the NEG proper, it is useful to step back a little and in following Fujita and Thisse (1996) to distinguish between the different expressions of the agglomeration phenomenon. At the smallest scale, restaurants, cinemas, and small shops cluster in a few neighbourhoods within a city. All of these restaurants, shops, etc. provide almost identical services. As an example consider the catering industry in London: there are an uncountable number of East-Asian restaurants in Soho and there are just as many Bengali and Bangladeshi restaurants in Brick Lane. Why do providers of such homogenous products cluster in the same neighbourhood –or on the same street even in the latter example?

Up one level, we may wonder why industrial districts exist. Presumably, at the regional level, it is hard to imagine that the UK motor industry is spread along the Thames Valley on the grounds of Ricardian or Heckser-Ohlin-Vanek comparative advantage. At the international level the North-South divergent development patterns also reflect a Core-Periphery structure, with the rich North being the industrialised core and the poor South being left behind in the periphery.

One can also distinguish among different agglomeration mechanisms. At the regional level, for instance, industrial clusters can take different forms as e.g. Duranton and Overman (2002) insist. As an illustration, they stress the difference in scale between the cutlery industry in Sheffield and the motor industry along the Thames. The former is localised in one area of Sheffield whereas the latter is spread over more than 100 km. This distinction has long been recognised, for more than a century ago Marshall (1890) identified three sources of agglomeration economies.

The first of these 'Marshallian externalities', as they are often referred to in the literature, is the knowledge spillovers. This is a technological externality that arises via face-to-face interactions. This is probably most relevant at the product development stage (Fujita and Thisse 1996, p. 345, Saxenian 1994). Berliant, Reed and Wang (2000) formalize the interaction between knowledge spillovers and concentration of economic activities.[1]

The second of these externalities regards the localised economies of a thick labour market. In the words of Marshall:

> A localized industry gains a great advantage from the fact that it offers a constant market for skills. (...) (Employers) are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skills as theirs (...). (Marshall, 1920, p. 225)

I shall return to this when discussing the analysis of Chapter 5.

Finally the last type of Marshallian externalities expresses itself as the forward and backward linkages associated with large markets (Fujita, Krugman, and Venables 1999, p. 5). This type of pecuniary externalities in particular stresses the fundamental importance of increasing returns for agglomeration phenomena. With non-increasing returns the distribution of factor endowments and technology alone determine the production patterns. By contrast the production scale has no importance: product specialisation patterns are determined but backyard capitalism prevails. Moreover, when factors are mobile the neoclassical theory cannot account for why cities of the scale observed come to existence.

To be sure von Thünen's (1826) concept of bid rents seeks to explain the distribution of agricultural production around cities without relying on indivisibilities. However it takes the existence of the city as given. As far back as 1940 Lösch (1940) believed that scale economies are important for understanding the spatial configuration of the economy and built a model of monopolistic competition.

The NEG literature to which all but one chapter of thesis belongs falls into the third Marshallian category.[2] The standard in the NEG literature is articulated around the Core-Periphery model introduced by Krugman (1991a,b) and similar models that build

---

[1] See also the discussion of Combes and Duranton (2001) in the section devoted to Chapter 5.
[2] The NEG's relation to the urban and regional literature is discussed in greater detail and certainly with more accuracy by Fujita and Thisse (1996, 2002).

Other studies then ask 'What are the determinants of agglomeration?' If vertical linkages between firms seem to be empirically relevant, the winning prize goes to the labour-pooling argument (Dumais, Ellison and Glaeser 2002, Rosenthal and Strange 2001). In other words, industries that use the same type of skilled workers seem to co-locate most. This suggests that agglomeration may be functional rather than sectoral, to paraphrase Duranton and Puga (2001).

Finally, various papers by Davis and Weinstein (1996, 1999) aim to detangle the causes of country specialisation. In particular, using data on Japanese prefectures (Davis and Weinstein 1999) and data on OECD countries (Davis and Weinstein 1996), they suggest that NEG determinants might be important at the regional level but that traditional Heckser-Ohlin-Vanek comparative advantages drives international specialisation. See also Midlefart-Knarvik, Overman and Venables (2001) for a similar approach using cross EU-country sectoral data.

Chapter 1

The purpose of Chapter 1 is threefold. First of all it introduces the notation as well as the model that will act as the workhorse to subsequent chapters. I also discuss the nature and operation of the agglomeration forces that play a central role in the NEG literature, namely the backward and forward linkages. Lastly I address the issues of equity and efficiency of the free-market equilibrium: the analysis identifies the losers and gainers of agglomeration and the nature of the externalities whose existence makes agglomeration generically sub-optimal; some of this analysis borrows freely from Baldwin et al. (2002).

The model this chapter introduces is a NEG model in which agglomeration stems from the interaction between increasing returns at the firm level, transportation costs, output-input linkages among firms (also called vertical linkages), and perfect capital mobility across all space. The original models based on vertical linkages are due to Faini (1984), Venables (1994, 1996a), and Krugman and Venables (1995). Applications to industrial development include Venables (1996b) and Puga and Venables (1996); applications to trade policy and preferential trade agreements include Puga and Venables (1996, 1997) and Baldwin et al. (2002).

Let me to compare the model I propose in Chapter 1 to the simplest form of the original model –taken from Section 14.2 in FKV– and call it the CPVL model (for

Core-Periphery and Vertical Linkages).[6] The CPVL model allows for two regions, two sectors, and one primary factor of production (labour).[7] Both regions are endowed with the same technology and labour force. The background sector produces a homogenous good under constant returns in a perfectly competitive environment using labour only; its output is freely traded. Restricting parameter values so that this sector is active in each region or country at any equilibrium ensures that labour wage is equalized across space and sectors.[8] By this token, the supply of labour to the sector of interest, which is called 'manufacturing', is perfectly elastic. The manufacturing sector produces different varieties of a horizontally differentiated product using both labour and intermediates under increasing returns. Its output is both consumed by final consumers and used as intermediate inputs by other manufacturing firms. This way firms are said to be vertically linked with each other.

Because there are increasing returns at the firm level the environment in which firms produce is necessarily imperfectly competitive. Indeed each firm is a monopolist in its own variety and hence it prices its output with a mark-up. Free-entry and exit ensures that no pure profit are made at equilibrium. Increasing returns have another implication: the number of firms that can be supported at equilibrium is finite but possibly uncountable. This way, market size matters: a larger market can accommodate a larger number of firms. Since both consumers (who value variety) and producers (who value input diversity) benefit from this, a larger market will be associated, ceteris paribus, with a higher degree of consumer utility and larger profits. This is the nature of the agglomeration force. Essentially it exploits in a spatial setting the kind of circularity causality recurrent in models of monopolistic competition (Matsuyama, 1995).

In the CPVL, workers are perfectly mobile on an intersectoral basis whilst perfectly immobile on an interregional one. In other words, workers can move freely from the background sector to the manufacturing sector within each region, but they cannot migrate from one region to another. To see how this might trigger regional disparities consider the following thought experiment. Imagine that nominal wages are equalized within each region and that initially both regions are perfectly symmetric. Consider the effect of moving one worker from the background to the manufacturing

---

[6] This terminology and the likes to follow are borrowed from Baldwin, Forslid, Martin, Ottaviano and Robert-Nicoud (2002).
[7] Models in which space is continuous are more involved. See e.g. Chapters 6 and 17 of FKV.
[8] The assumption that trade in the homogenous good is free is one of convenience (FKV, Chapter 7). However, see Davis (1998) for a divergent opinion.

sector in a single region only. In the background sector this worker is paid her marginal product. In the manufacturing sector this supplementary worker will generate a product that is priced above the cost of her labour. This is a pecuniary externality for the worker does not take it into account when she decides to join the manufacturing force as she simply compares the nominal wages.

Furthermore, this expansion of output must be accompanied by an increase in demand for intermediates since these factors are complementary. The existence of transportation costs implies that this supplementary demand is biased towards local inputs, which in turn raises the profitability of local firms relative to foreign ones (by mark-up pricing again). This increase in profitability is matched by the entry of new firms (more of them in the local market), which also hire workers from the background sector, so the cycle repeats. This is the essence of the demand (or backward) linkage.

When a new firm starts producing in a region, it also offers a new variety of intermediates that by the prevalence of transportation costs, decreases the production cost of local firms relative to foreign firms. This increases the profitability of the former relative to the latter, so net entry of new firms takes place in the local market, so the cycle repeats. This is the essence of the cost (or forward) linkage.

The discussion of these linkages points to the obvious fact that strictly positive transportation costs are essential for this home-bias to exist and trigger agglomeration forces. When transportation costs are nil, location is irrelevant.

If the forward and backward linkages were left by themselves, all firms would cluster in a single location so as to take full advantage of these agglomeration economies. In the standard model, there is only one centrifugal, or dispersion force to oppose these centripetal forces: the market crowding effect (in the terminology of Baldwin et al. 2002). In a monopolistically competitive environment, there can be no pro-competitive effect. However when entry occurs the market share of all existing firms shrink and since markets are segmented by transportation costs, they do so disproportionably for the firms located in the market in which entry occurs.

To sum up, if one extra worker joins the manufacturing workforce this may or may not trigger a snowball effect that takes the form of a 'catastrophic agglomeration' (in the jargon), depending on whether the linkages are stronger or weaker than the market crowding effect.

Consequently agglomeration will or will not take place as the result of the tension between centrifugal and centripetal forces. The value of the parameters of the model is crucial, and traditionally the NEG has focused its attention on one of them: transportation costs. In the standard models like the CPVL model, agglomeration is the outcome when transportation costs are low whereas dispersion forces prevail when transportation costs are high. It is clear that different assumptions may lead to different conclusions. In Helpman's (1995) model, for instance, the conclusion as far as transportation costs of the manufactures are concerned is the exact opposite.

Table 1. Three classes of Economic geography models

	No Aggl. Forces	Vertical Linkages	Factor Migration	Factor Accumulation
Fixed firm size	Krugman (1980)	CPVL	CP	
Specific factor	FC	FCVL	FE	CC

In the model I propose in Chapter 1, firms also by buy each other's output and use this as intermediates, alongside labour they can poach at a constant wage from the background sector. There are two differences between the model therein and the CPVL model. In the latter there is only one primary factor of production, labour, which is spatially immobile. Hence the agglomeration rents accrue to the workers and it is this, which drives them in or out the background sector out of equilibrium. By contrast, there are two primary factors in the model of Chapter 1: capital and labour. As before, the background sector uses only labour. In the manufacturing sector, labour is used alongside intermediates as a variable input; capital is specific to the manufacturing sector and hence captures the agglomeration rents. Capital is also mobile across regions –but capital owners are not. Hence capital moves in search for the highest nominal return. Capital being mobile and specific to the manufacturing sector, agglomeration rents induce capital to move from one region to another. In short, interregional capital mobility plays the same role here as intersectoral labour mobility in the CPVL model. Despite this important conceptual difference, the dynamic properties of both models are strikingly similar. These two models arguably illustrate different empirical situations, but when it comes to theoretical applications the model of Chapter 1 is more parsimonious and hence is much easier to manipulate. It has the additional advantage of being the natural extension of the 'Footloose Capital' trade model by Flam and Helpman (1987), which it encompasses. For this reason I dub it as the FCVL model (for Footloose Capital and Vertical Linkages).

Table 1 classifies various NEG models according to their agglomeration mechanism and the functional form they assume. In the first column are the original models on which the NEG models of columns 2, 3 and 4 are built. Neither the model of Krugman (1980) nor the FC model of Flam and Helpman (1987) display any self-enforcing agglomeration force by themselves. This is achieved by the addition of vertical linkages among firms (like the CPVL and FCVL models of the second column), factor migration (like the CP and FE models of the third column), of the endogenous accumulation of factors (like the CC model of the last column).

The second column of Table 1 summarizes the earlier discussion. The CPVL model is the original model of agglomeration as driven by vertical, or input-output, linkages. The functional form here assumes that the fixed and variable components of total cost for the typical manufacture use factors in the same intensity, like in Krugman (1980). As a result of free entry, the equilibrium firm size is function of the parameters only; the adjustment variable is the number of firms. The FCVL model, by contrast, is a specific factor model: the fixed cost is made of capital, whereas the variable cost is as in the CPVL model. As a result, the number of firms is fixed by initial endowments; the adjustment variable is the size of firms.

The interaction between vertical linkages, factor mobility, and other factors is not the only possible mechanism that can give rise to agglomeration, as can be inferred from Table 1. I turn to these next.

Chapter 2

Before turning to the purpose of Chapter 2, some additional background is necessary.

This body of research has also identified other sources to forward and backward linkages. I am here surveying two of them: embodied factor migration and endogenous factor accumulation.

Start with the models in which agglomeration is driven by factor migration alongside the usual suspects (monopolistic competition and transportation costs). These are the topic of the second column of Table 1. The Core-Periphery model by Krugman

(1991a,b) –a more complete version of which can be found in Chapter 14 of FKV– is the backbone of the NEG literature.[9] It assumes two ex-ante identical regions, each initially endowed with equal numbers of unskilled and skilled workers. Each factor is specific to the background and manufacturing sectors, respectively. Like in the CPVL model, the background sector is Walrasian and the manufacturing sector is monopolistically competitive. Skilled workers are mobile between regions (unskilled workers are not) and move according to current real wage differences.

As in the CPVL model there is room for circular causation in the form of backward and forward linkages. A simple way to identify these linkages in the CP model is to point at the similarities with the CPVL model.

To begin with consider the forward (or cost) linkage. In the CPVL model, workers 'migrate' from one sector to another so as to equalise nominal wages. When workers join the manufacturing sector, they resemble the firm's owner, for pure profits are eliminated and wages capture all the rents. Hence, when making their occupational choice, workers will compare the costs that are prevailing in the manufacturing sector with those in the background sector, among other things. Since there are vertical linkages in the former, the manufacturing costs are a function of the location equilibrium. Usually then, manufacturing costs differ in the two regions. In the CP model there are no vertical linkages, but interregional migration plays exactly the same role. The cost dependant on the location pattern is no longer a production cost, but a cost-of-living: in the CP model, manufacturing workers have to make a location choice only which they base on the real wages currently prevailing in each region. Therefore, as such a worker immigrates into a region and contributes to the local manufacturing production process; this decreases the cost-of-living in the destination location and increases it in her region of origin. Again, this arises because transportation costs are strictly positive.

In the CPVL model, the backward (or demand) linkage stems from the fact that firms use each other's output as intermediate. In the CP model, the backward (or demand) linkage stems from the fact that skilled workers are also consumers. Consequently when such a worker moves from one region to another, this increases the

[9] See also the excellent and critical survey by Neary (2001). Baldwin et al. (2001, 2002) provide the most comprehensive analysis of the model.

relative profitability of the firms in the latter. Again, this home-market bias holds because transportation costs are strictly positive.

In the CP model, too, if the forward and backward linkages were left by themselves all firms would always cluster in a single region. Moreover there exists a dispersion force alongside these agglomeration forces. Like in the CPVL model some demand emanates from both regions in any configuration. Indeed the workers in the background sector are immobile. The usual market crowding effect applies: the more firms co-locate, the lower is market share on the domestic market to each of them. When this market is large enough and/or transportation costs are low enough, this does not restrain the marginal firm to locate there. If this were not the case, then it would rather go for a large market share in a small market. This trade-off illustrates the tension that exists between agglomeration and dispersion forces.

As I already mentioned, the Core-Periphery model by Krugman (1991a,b) is the backbone of the NEG literature. Regrettably the CP model is astoundingly difficult to work with analytically. None of the interesting endogenous variables can be expressed as explicit functions of the variables that the model tells us are important –trade costs, scale economies, market size, etc. Indeed, the CP model does not even provide a closed-form solution for the principal focus of the whole literature – the spatial distribution of industry. This has forced researchers to illustrate general points with a gallery of numerical examples. While the resulting outcome points to some consistent results, it is less than fully satisfactory from a theorist's perspective; one simply cannot be certain these regularities give a comprehensive description of the results generated by the model.

As Forslid (1999), Ottaviano (2001), and Forslid and Ottaviano (2001) have shown, a large shrunk of the intractability of the Core-Periphery model is due to the specific functional form Krugman is using. To put it simply, they propose to work with a functional form similar to the one I use in the FCVL model instead. In particular both sectors use labour as the variable input with capital specific to the manufacturing sector. In Forslid and Ottaviano's model, capital is embodied and as a result is best interpreted as human capital. Like the manufacturing workers in the CP model, human capital owners, or entrepreneurs, migrate in search for the highest real wage. For this reason, call this model the 'Footloose Entrepreneur' (or FE) model. The cost of living in the FE model replaces the vertical linkages of the FCVL model as the forward linkage. The backward linkage in both the CP and FE models are identical. The market crowding effect needs not be detailed for it is identical in all models surveyed thus far.

As it turns out the FE and CP models are isomorphic. In particular their dynamic behaviour is the same. Since the FE is much easier to work with, it makes it more amenable to use in applications and extensions.

Finally in the class of models of the third column of Table 1, the agglomeration forces stem from the accumulation of factors. The pioneering model in that tradition is Baldwin's (1999) 'Constructed Capital' model.[10] Baldwin (1999) works with the specific capital functional form. Capital is not exogenously given, as in the FE and FCVL models but produced by a Walrasian investment sector using labour only. Investors are forward looking. There are no vertical linkages and no footloose factor. Agglomeration forces stem purely from the fact Capital is assumed to be immobile allowing the returns on capital to diverge across regions. If for some reason capital accumulates faster in a region, income and demand increase faster there, too. Via the usual home-market bias local firms benefit more that foreign ones, which stimulates further entry in the form of investment and capital accumulation, so the cycle repeats. This is a demand linkage. There is no cost linkage in this model. As a consequence of this it is much more amenable –so much so the model is completely solvable analytically. The market crowding effect is the same as in the other models.

Before proceeding further, I note as an aside that the grids of Table 1 are not hermetic: it is perfectly possible to imagine a model combining two or more of the agglomeration forces. In fact, with a European context in mind, this is exactly what Puga (1999) does with the CP and CPVL models to study the impact of labour mobility (or absence of) on regional divergence. Faini (1984), whose work it is fair to say considering the excitement that followed the publication of Krugman (1991a,b) was overshadowed at the time it came out, proposes a model of capital accumulation in which the production of a Walrasian final good uses capital, labour, and non-traded intermediate inputs. In this sense it merges the last two columns of Table 1 for yet another functional form (not shown). The purpose of Chapter 2 is twofold. First, it

[10] Applications and extensions involve endogenous growth models. See Baldwin and Forslid (2000), Baldwin, Martin and Ottaviano (2001), and Martin and Ottaviano (1999, 2001). The engine of endogenous long run growth stems from spillovers in R&D, possibly due to the non-rivalry in knowledge (Romer 1990). If spillovers are localised, then innovation will take place in any region that takes the lead, as in Englman and Walz (1995).

claims that all the models listed in Table 1 are isomorphic (with some qualification for the CPVL model) in a sense that I shall make precise below.

Moreover, it completes the analytical analysis of the CP model –and so confirms that the gallery issued by numerical simulations is complete. To sustain the first claim I show in this chapter that a judicious choice for the state variable and judicious collections of the structural parameters render these models identical (or almost so in the case of the CPVL model).

Then I use this fact to prove that all the models share the same dynamic features. In this way I complete the characterization of the CP model by simply conveying this exercise for the FE model, which is much simpler to work with. In doing so this chapter contributes to an eleven-year old body of the literature that assigned itself the task of completing the analytical study of the CP model.

In his canonical paper, Krugman (1991b) took the simple Dixit-Stiglitz monopolistic competition trade model with trade costs and added labour migration driven by real wage differences. Using simplifying assumptions, leaving the dynamics of the model in the background, and relying on numerical simulations, Krugman showed how this model behaved in a radically different way as compared to the trade model.

Further contributions to our understanding of the model include Puga (1999), who linearized the model around its symmetric steady state, and this way provided an analytical solution to an important parameter of the model, the 'break point'. Baldwin (2001) used formal tests to assess the global stability properties of the steady states of the model and introduced forward-looking expectations (together with migration costs) on the migrants' side. This exercise showed that most properties of the original model remain true in a more formal and orthodox setting. Interestingly, when migration costs are low, 'history versus expectations' issues of the type first described in Matsuyama (1991) arise.[11] Ottaviano (2001) conveys the same type of exercise in the FE setting.[12] Neither Puga (1999) nor Baldwin (2001) provides an analytical proof for the fact that

---

[11] However, Karp (2000) shows that the equilibrium indeterminacy vanishes under certain conditions when agents have 'almost common knowledge' (in the sense of Rubinstein 1989) about economic fundamentals rather than common knowledge.

[12] See also Krugman (1991c) and Matsuyama and Takahasi (1998) on this issue in a related model. The latter authors conduct welfare analysis and show how the coordination failures between migration decisions of individuals and entry decisions of firms typically result in inefficiencies at the long run equilibrium location.

there are generically three interior and two corner steady states. This chapter closes this gap.

Other related contributions include Herrendorf, Valentinyi, and Waldman (2000), Tabuchi and Thisse (2001), and Ottaviano, Tabuchi, Thisse (2002). The latter paper proposes yet another functional form for migration-driven models of agglomeration (and hence fills a virtual additional cell to column of Table 1 labelled 'Factor Migration'). Their work adopts a quasi-linear quadratic utility function and in doing so avoids some of the unappealing implications of the Dixit-Stiglitz monopolistic competition and iceberg trade costs. In particular firm prices change as location changes allowing for real competition effects. Moreover this model too is analytically solvable. The remaining papers reconsider the 'bang-bang' properties of the model. Tabuchi and Thisse (2001) show that this property is due to the assumption that potential migrants are homogenous. Taste heterogeneity adds a dispersion force into the model that smoothes the location equilibrium: the correspondence that maps transportation costs into the location equilibrium becomes continuous. Herrendorf, Valentinyi, and Waldman (2000) convey the same exercise in Matsuyama's (1991) setting in which scale economies are external to the individual agents.

Chapter 3

The topic of Chapter 3 is probably the most innovative of this thesis.

The de-location process associated with trade integration has been a major concern for European policy makers for decades. It is reflected, for instance, in the quadrupling of cohesion spending as a share of the EU budget since 1986, and in the important level of spending by member states on their disadvantaged regions such Germany's Eastern Länder and Italy's Mezzogiorno. Much of this spending is explicitly aimed at preventing, delaying or even reversing the agglomeration of economic activity in favoured regions.

Using the FC model introduced in Chapter 1, the aim of the present chapter is to address the issue of agglomeration during a process of regional integration in a framework where regional policy is determined by political economy forces. More precisely, taking a laissez-faire equilibrium as a benchmark, the analysis show how politics and economic integration interact in both directions to speed up or slow down the agglomeration process that results from integration.

With few exceptions, the economic geography literature has not considered policy issues since it is concerned mainly with the positive analysis of exogenously rising levels of openness. When it has, instruments of regional policy were either taken as exogenous (Baldwin and Robert-Nicoud 2000, Martin 1999, Martin and Rogers 1995, Ottaviano and Thisse 2002), or addressed the specific issue of inter-regional tax competition (Andersson and Forslid 1999, Baldwin and Krugman 2001, Kind, Midelfart-Knarvick and Schjelderup 2000, Ludema and Wooton 2000). Likewise, Persson and Tabellini (1992) consider a model where two policy makers, each from a different region or country, compete for the mobile factor (capital) by setting taxes. They consider how equilibrium redistributive policies are affected by economic integration in a more classical environment (i.e. they focus on public good provision in populations with heterogeneous factor ownerships rather than on economic geography issues). By contrast, this chapter assumes that both regions belong to a single, centralised constituency and focus exclusively on the interaction between spatial redistribution politics and geography in a framework of electoral competition.

In a related context, Rauch (1993) assesses the role of the developers of industrial parks in coordinating location decisions by individual firms. Firms are initially agglomerated in a location that no longer has the comparative advantage (the nature of agglomeration forces is left unspecified). Under some conditions, there is a first-mover disadvantage and as a consequence firms fail to coordinate and are stuck in the 'wrong' location (that is, history matters). In this context, Rauch shows that 'land developers' (see e.g. Henderson 1985) can circumvent this coordination problem by subsidising the first firms to move, while charging a positive price to the land slots allocated to firms moving at a later stage. In this way, the land developer makes non-negative profits and firms that move first are compensated for not taking advantage of the location economies that accrue to the initial location. Rauch also cities some evidence that supports his theory. In this chapter, the central government plays the role of the land developer, but its role is not to coordinate firm relocation (the model features a unique locational equilibrium, so there can be no coordination failure). Rather, political candidates seek a location equilibrium that maximizes their political support. Generically, the outcome is not the utilitarian optimum.

The paper by Cadot, Röller and Stephan (2001) provides empirical support for the idea that candidates will craft their policy platforms to please regions that have a lot of 'swing voters'. Using French panel data, these authors show that electoral concerns (as well as lobbying activities) are significant determinants to the spatial allocation of regional transportation infrastructure investments. In particular, they instrument for the proportion of swing voters in different regions and show this explanatory variable to be statistically highly significant. Since they take the location of industries as given, however, their study is not a direct test of the model of this chapter but it provides strong support to the political mechanism it is assuming.

The following chapter extends the FCVL model of Chapter 1.

Chapter 4

The purpose of Chapter 4 is to study the phenomenon of fragmentation of vertically integrated production processes from a NEG perspective. The term 'fragmentation' refers to the breaking up of such processes into parts and components, which might then be internationally traded, in which case this qualifies as intra-product trade (Jones and Kierzkowski 1990). To take a specific example,

> 'Textiles and electronic products may be designed and marketed in Hong Kong, but they are largely produced in the Pearl River Delta.' (Arndt and Kierzkowski 2001, p. 1)

The neoclassical paradigm then predicts that each locations or countries produce components according to their comparative advantage. Before the breaking up of the process, the production of the vertically integrated process took place according to the average factor intensity of the end product. What permits this physical breaking up across national borderers is the reduction in coordination and communication costs of all sorts, from the convergence of legal systems to technological innovation in the telecommunication sector.

The model in Venables (1996a) is best suited to study the interaction between the location of the production of components and that of final goods in a NEG framework. This model is the initial version of the CPVL model in which the manufacturing sector is split into two distinct sub-sectors: the upstream sector, which produces intermediates, and the downstream sector, which uses them together with primary factors to manufacture different varieties of a final good. In this chapter, I want to go beyond this re-labelling exercise and study instead the fragmentation of the *services* to the firm.

Indeed, a similar phenomenon can be observed in the case of services provided to the head of the production unit, also referred to as the 'front office'. Specifically this chapter looks at how reductions in communication costs – together with variations in transportation costs – affect the spatial distribution of economic activity. 'Communication costs' are understood here as encompassing the cost of coordination and of conveying information between the head of the production unit and other workers. Unlike transportation costs proper, the object of the communication costs is immaterial.

Low communication costs allow a firm to physically separate different activities that used to be performed in a single location. The actual production of the final output is performed by the 'front office'. 'Back office' tasks involve include data entry, data processing, and database management; financial and accounting services; processing assurance claims; and computer software development. Nowadays, the unbundling of back and front office production is widespread. I review some anecdotic evidence in the introductory section of Chapter 4.

In NEG models all forms of costs are collected into one parameter: the iceberg transportation costs. The aim of this chapter is to disentangle communication costs from trade and transportation costs and furthermore to assess how this framework conveys new insights to help understand the growing importance of these phenomena that have taken some importance recently – primarily the growing importance of unbundled back office work.

The analysis of this chapter is related to earlier work as follows. In the paper by Gao (1999) headquarters are intermediate-inputs intensive whereas final production is labour intensive. In this chapter I make the opposite assumption. The intermediates to HQ's are specialised business services –for example lawyers, insurers, or banks. This is what Gao's (1999) model captures. On the other hand, this chapter analyses the fragmentation of routinised business services (e.g. call centres) in industries in which localisation economies are more important at the manufacturing stage. In this sense the two are complementary. Duranton and Puga (2001) allow for a multi-industry, multi-city setting in which firms trade-off the benefits of becoming multi-location –benefits associated with the localisation economies in both business services and specific intermediates– with the managerial cost-saving associated with a spatially integrated firm.

Harris (1995, 2001) argues that better transportation and communication technology together with the quasi non-rival nature of the latter is a major force driving the fragmentation of the production process. In his model operating global communication networks involve large scale economies; consequently, trade results from the specialization in component production to take advantage of these scale economies. In an essentially static, partial equilibrium framework Gersbach and Schmutzler (2000) study how local spillovers between plants and knowledge spillovers within multi-plant firms (a shorthand for communication costs) interact to generate agglomeration.[13]

Jones and Kierzkowski (1990, 2001) propose a general framework to consider fragmentation. Campa and Goldberg (1997) and Feenstra (1998) provide some indirect evidence for this phenomenon at the industry level for the UK and the US respectively. Helleiner (1981) reports that a growing part of international trade takes the form of intra-corporation trade, which also provides indirect support to the story.

In this chapter the model extends the FCVL model in two ways. First, following Krugman and Venables (1995) I assume that the labour supply to the manufacturing sector is no longer perfectly elastic. This adds a dispersion force that is not sensitive to variations in transportation costs. As a result agglomeration of industry in a unique location arises at intermediate levels of trade barriers only. This suggests the rather optimistic view according to which falling transportation costs might have created a North-South duality in the past, but that fostering 'globalisation' beyond the current standards will naturally reverse that trend. I reconsider this issue.

Second, I amend this scenario to allow for the possibility for firms to hire workers from abroad to carry on some back office task. In the model this possibility arises when communication costs are low, having the effect of integrating the labour market. As a result Northern firms may be increasingly able to outsource some back office jobs to Southern countries whilst retaining their Northern location for the processing of final goods so as to take advantage of the agglomeration economies associated with the cluster. In doing so two effects emerge. Firstly some good news for the poor country: the integration of the labour markets should bring a convergence in

---

[13] Their model implicitly assumes that by definition of the term 'spillovers' multi-plant firms do not internalise these flows. Since they play a central role in the analysis, this is somewhat unsatisfactory. Moreover, the model makes the rather strong assumption that knowledge 'spills over' to firms sharing the same location better than it would do from two plants of the same firm located in two different locations.

nominal incomes. However, the bad news is that the prospects for the South industrialising are postponed, and hence the poorest countries forego the agglomeration rents for an extended time (namely, trade costs would have to decrease further to re-industrialize the South). The net outcome of these two effects on people's welfare is unclear.

Chapter 5

Chapter 5 goes beyond the NEG and contributes to the growing literature on the externalities of the labour market as a cause for agglomeration. In particular, it shows how the hold-up problem can be softened or worsened by the cluster of industries employing workers with similar skills. The central tenet of this chapter is thus: agglomeration and labour market pooling (or absence of) emerges as the outcome of the interaction between market power (or lack of), the non-verifiability of some investment, and the specificity of this investment to the relationship –which itself depends upon the location decision of various agents. The assumption that investment is not verifiable by a third party is both reasonable in the context of human capital accumulation and important. If it were observable and verifiable, then it would be contractible. See e.g. Hart (1995) for a discussion of this issue.

The pooling of the labour market brings us back to the earlier quotation of Marshall (1890). The literature has put forth several motives for labour market pooling. In Krugman (1991a) firms facing idiosyncratic risks pool their labour force so as to reduce aggregate volatility of regional employment. In Combes and Duranton (2001) firms that co-locate benefit in drawing from a common pool of trained workers. It follows that firms can poach each other's workers, having two effects. Firstly when a worker moves from one firm to another brings with her the knowledge of the product of her former employer; this enhances the degree of competition in the goods market. In turn this reduces the firm's revenue (note that this effect is minimal in the case of functional agglomeration). Secondly the ability of poaching a competitor's workers enhances the degree of competition in the labour market, increasing wages and costs. Presuming that some knowledge is embodied in workers, the authors note that labour mobility across firms diffuses knowledge acquired in various workplaces. If workers tend to primarily consider jobs in the same geographical area as their current one, this provides a theoretical foundation for the diffusion of knowledge spillovers.[14]

In Hesley and Strange (1990), workers with heterogeneous skills and firms with heterogeneous skill requirements expect the quality of matches to improve with city size as in Kim (1990). This is because there are increasing returns to scale in production and because information regarding skills is private. Since all agents are mobile these agglomeration economies will ensure that a discrete number of cities exist at equilibrium. The dispersion force at work comes from the fact that land prices increase with city size. As is common in the urban literature cities are mono-centric by assumption. Hence as the size of the city increases, commuting costs increase for the additional workers joining the city, which in turn translates into high land rents near the 'central business district'. The equilibrium number and size of cities are the result of the tension between the congestions costs and the agglomeration economies.

The theoretical setting of Chapter 5 is perhaps most closely related to Rotemberg and Saloner's (2000). They assume a simple environment in which two very different parties are co-dependent to produce a given good. One party has all market power; the other is to make a non-verifiable investment. Under these strong assumptions, they get a strong result: agglomeration takes place as it solves the hold-up problem.

The analysis of Chapter 5 will stress that a very specific conjunction of assumptions is needed for this result: agglomeration is the efficient spatial organization of production if the party that has to make important relation-specific investments faces a potential hold-up problem. Now, assume the contrary case in which the firm that has to make an unverifiable investment. In this situation the firm it does better by locating somewhere in the 'periphery' and forming a one-firm town so as to grab a larger share of the surplus generated at the production process. This, in turn gives it higher incentives to invest in the first place. Hence, in this specific example, agglomeration worsens the hold-up problem, turning Rotemberg and Saloner's (2000) result upon its head. To sum up the respective identity of the party that exerts market power and of the one that makes an industry-specific investment is crucial. This suggests a potentially rich theory in which the relative importance of the investments of the parties involved,

---

[14] The empirical paper by Jaffe, Trajtenberg and Henderson (1993) suggests that knowledge spillovers are local in scope.

the specificity of this investment to their relationship, and their outside options all mater.

The question 'does competition solve the hold-up problem?' addressed in Felli and Roberts (2001) is clearly related to ours: agglomeration affects the hold-up problem in our setting via its effect on the degree of competition, as measured by the degree of market power. Felli and Roberts, however, assume away efficiency arising from market power (or absence thereof) and concentrate on inefficiencies arising from matching frictions. Accordingly their measure of degree of competition is different to ours (and is very specific to their setting). In their setting, heterogeneous firms compete a-la Bertrand for matches with heterogeneous workers once relation-specific investments have been made. The numbers of both firms and workers are finite with the number of workers being smaller than the number of firms; each party is matched with at most one agent and there is exogenous heterogeneity in quantitative abilities on both sides.

In the matching process they are assuming the party that is on the 'long' side of the market (the workers) is the residual claimant of the surplus generated by the match. Consequently when workers alone invest, they face the correct investment incentives and investment is efficient. If instead firms engage in investments before the Bertrand competition game starts, then they face a hold-up problem and under-invest as a result. However in this case aggregate inefficiency is bounded above by the inefficiency that would arise if the best firm matched with the best worker in isolation. This is a remarkable result, as this means that inefficiencies due to the hold-up problem do not cumulate in the presence of workers' competition for the matches (Felli and Roberts 2001). In other words, competition solves (a part of) the hold-up problem.

Other important and related papers are Acemoglu and Shimer (1999), Bolton and Whinston (1993), and Grossman and Helpman (2002). Acemoglu and Shimer (1999) consider which labour market institution could solve the hold-up problem that results from search frictions. In their model, firms alone make investments prior to matching with workers. With ex-post bargaining over the surplus firms' investments are held up and hence firms under-invest in the first place. If instead firms were able to post wages and workers to direct their search towards different firms, then these authors show that the decentralized economy would achieve an efficient outcome under relatively mild conditions: even if the workers were able to observe any two wage offers at a time only, each firm in effect Bertrand-compete for workers, as long as the pair of firms is random. Firms that have acquired more capital are able to propose higher wages. They have the incentive to doing so because posting higher wages fills vacancies faster. So, investment (and entry) is constraint efficient. As the authors note, it is not sure that this remarkable result extends to more complex environments, e.g. one in which both sides engage in non-verifiable investments.

The other papers relate the boundaries of the firm with the hold-up problem, following the seminal papers by Grossman and Hart (1986) and Hart and Moore (1990). See also Hart (1995) and Holmstrom and Tirole (1989) for non-technical treatments.

A concluding chapter wraps up the results of the analysis of Chapters 1 to 5 points to some questions related to the issues tackled in this thesis but left aside here. They constitute material for further research.

Chapter 1. A SIMPLE MODEL OF AGGLOMERATION WITH VERTICAL LINKAGES AND PERFECT CAPITAL MOBILITY

1.1. Introduction

This chapter introduces the workhorse model that will be used in later chapters. It is a 'New Economic Geography' model that nests the 'Footloose Capital' (or FC) model of Flam and Helpman (1987), to use Baldwin et al.'s (2002) terminology.[15]

The original model in the 'New Economic Geography' (or NEG) is Krugman's (1991) 'Core-Periphery' (or CP) model in which agglomeration relies on (skilled) labour migration. In an alternative class of models starting with Venables's (1996) paper agglomeration arises as the interaction between labour mobility *between* sectors *within* the same region and input-output (or vertical) linkages among firms (see the introductory chapter). Empirically vertical linkages are a stronger explanation of international agglomeration patterns than labour migration. Also, capital mobility is much more prevalent than labour migration. It is therefore useful to have a model in which capital mobility and vertical linkages together sustain agglomeration.

This can be achieved by extending the FC model to include intermediate inputs. The resulting model, call it the FCVL model, retains the same qualitative properties of Fujita, Krugman, and Venables' (1999, chapter 14.2) CPVL model, which greatly simplifies the original model by Venables (1996a).[16]

This chapter presents this model in great lengths. However, the characterization of its stability properties are left to Chapter 2 because this latter chapter is dedicated to the stability properties of NEG models in general –and not just the FCVL model.

The remainder of the chapter is organised as follows. The section below sets up the structure of the model. Section 1.3 discusses the agglomeration and dispersion

---

[15] The FC model is strictly speaking a "New Trade' model. However, it can also be interpreted as geography model since it has a spatial dimension. See Martin and Rogers (1995).

[16] 'FCVL' stands for Footloose capital-vertical linkages and 'CPVL' stands for Core periphery-vertical-linkages.

forces displayed in this model. Section 1.4 solves the equilibria for ex-ante symmetric regions. Section 1.5 introduces the FC model as a special case of the nesting model. Sections 1.6 to 1.8 conveys the welfare analysis. Section 1.9 concludes.

1.2. The basic model

The model developed in this section shares functional forms with Flam and Helpman's (1987) trade model and builds on Dixit and Stiglitz's (1977) framework of monopolistic competition. The novelty here is to add agglomeration forces a-la Venables (1996a). As I proceed, I make choices of units and of the numéraire that are standard in the NEG literature.

Tastes and production

Consider a country consisting of two regions or countries, j=1,2 (in the applications of subsequent chapters, one interpretation or the other will be more appropriate; in this chapter, however, I use the two terms interchangeably). The typical individual is assumed to supply one unit of labour L (the reward of which is w) and k units of capital K (the reward of which is π) inelastically. There is a measure L of workers and a measure K of capital in this economy, so the typical worker owns k=K/L units of capital and, as a consequence, her income is w+kπ. Tastes for a typical individual in j take a Cobb-Douglas form in which j spends a share μ of her income y_j on a composite good M (for 'manufacturing') and a share 1-μ on the homogenous good A (for 'agriculture', say). The composite good M comes in N different varieties. Tastes over the different varieties are captured by a CES, 'love-for-variety', functional form, with an elasticity of substitution σ between any pair of varieties. The dual of this, the indirect utility function of region j's representative consumers, can therefore be written as:

$$\textbf{(1-1)} \qquad V_j = \frac{y_j}{p_A^{1-\mu} G_j^{\mu}}; \; G_j \equiv \left(\int_{i=0}^{N} p(i)^{1-\sigma} di \right)^{\frac{1}{1-\sigma}} ; \; 0 < \mu < 1$$

where p(i) is the consumer price of variety i, G_j is the true CES price index over the N varieties of the manufacturing good, and p_A is the price of good A (the reason p_A and π are not indexed by j will become clear shortly). Each firm i$\in$[0,N] produces a different variety, the buyer price of which is p(i); this brings us to production.

Each producer enjoys monopoly power over his own variety. No producer has any incentive to produce a variety already being produced by another producer, for she would then directly compete for the market of that variety with the incumbent producer. As a result her profits be lower. Hence, N is also the number (mass) of firms operating in sector M.

The manufacturing sector M is the usual monopolistic competition sector a-la Dixit and Stiglitz (1977). It produces a CES aggregate under increasing returns. Specifically, each firm needs a fixed amount a_{KX} of capital K to start producing and a constant amount β of a Cobb-Douglas composite input made out of labour (with share $1-\alpha$) and intermediates produced by sector M itself (with share α) for each unit of output it produces. Mathematically, the cost function for the typical firm located in j is given by:

(1-2) $$C_j(x_j) = a_{KX}\pi_j + \beta x_j w_j^{1-\alpha} G_j^{\alpha}; \quad 0 < \alpha < 1$$

where G_j is the same CES price index as in (1-1), π_j is the cost of one unit of K prevailing in j, and x_j is a typical firm output. Observe that the same G_j enters (1-2) and (1-1); this means that the elasticity of substitution among varieties of manufacture is the same for consumers and for firms. Qualitatively, this is an innocuous assumption, but this is required to keep the analysis manageable. See also Fujita et al. (1999, chapter 14).

The background sector A produces a homogenous good under constant returns using labour only: the per-unit output labour requirement is a_{LA}. A is assumed to be freely traded (hence p_A is the same in both regions) and parameter values are chosen so that no region ever specializes in M (we make the 'no-specialisation' condition more precise below); further, we choose A as the numéraire. These imply $a_{LA}w_j=p_A=1$, $j\in\{1,2\}$. As a consequence, we can rewrite (1-1) as $V_j=y_jG_j^{-\mu}$.

By contrast to A, interregional trade in M is subject to Samuelson-type iceberg transportation costs $\tau\geq 1$. That is, in order to sell one unit abroad a firm has to ship τ units. The difference $\tau-1$ melts in transit (hence the name). Monopolistic pricing yields the usual relation $p_j(1-1/\sigma)= \beta w_j^{1-\alpha}G_j^{\alpha}$ for the producer price of a typical firm in j. The term in the right-hand side is the marginal cost, and σ is the perceived elasticity of demand; this requires to impose $\sigma>1$ as a regularity condition. We choose units so that $\beta=1-1/\sigma$, hence $p_j=G_j^{\alpha}$. In Dixit-Stiglitz monopolistic competition transportation costs are fully passed onto consumers so the producer price p_j holds irrespective of the market served.

Would be entrepreneurs bid for units of capital. Free entry and exit in M ensures that these entrepreneurs make no profits, so the operating profits of a typical firm active in j just cover the capital reward π_j:

(1-3) $$\pi_j \equiv \frac{p_j x_j}{\sigma}, \quad p_j = G_j^{\alpha}$$

Normalize a_{KX} to 1. By symmetry among all varieties and full-employment of capital this implies N=K. Furthermore, we normalise K to 1 and we define n as the share of N operating in j=1. Therefore, we can rewrite the price indices in (1-1) as

(1-4) $$\Delta_1 = n\Delta_1^{\alpha} + \phi(1-n)\Delta_2^{\alpha}; \quad 0 \le \Delta_j \equiv G_j^{1-\sigma} \le 1, \quad 0 \le \phi \equiv \tau^{1-\sigma} \le 1$$

and Δ_2 is defined analogously. Note that the definitions of Δ_1 and Δ_2 are implicit and simultaneous. The variable G_j and the primary parameter τ usually come raised at the power $1-\sigma$, so it is more convenient to use Δ_j and ϕ instead. The parameter ϕ measures the degree of free-ness of inter-regional trade in manufactures M (it is zero when trade is prohibited and equals unity when trade is perfectly free). Like ϕ, Δ_j lies in the unit interval because $n \in [0,1]$ and $\alpha<1$. (This claim is easily made by contradiction.)

Endowments and factor mobility

Potentially, the two regions differ in size: region 1 is endowed with a share s of world labour and world capital stock alike; assume $s \geq ½$ without loss of generality.[17] Labour is embodied and immobile; capital is disembodied and perfectly mobile in the long run (consequently, $n \neq s$ is possible).[18] Both workers and capital owners are themselves immobile. We further assume that the capitalists own a perfectly diversified portfolio, namely, each of them own the same share of each firm.[19] Hence, their portfolio return is $\pi \equiv n\pi_1 + (1-n)\pi_2$. In the long run, capital is perfectly mobile, so

[17] I relax the assumption of identical relative endowments in section 1.5.
[18] What 'long run' means in the context of this model will be made clear in Section 1.4.
[19] More on this in footnote 21 below.

$\pi_1=\pi_2=\pi$ must hold whenever there are some active firms in both countries.[20] Also, remember that $w_1=w_2=1$ holds by free trade in A and by the choice of numéraire.

All these imply that aggregate income Y in region 1, say, is equal to $Y_1=s(L+\pi)$. Location 1 expenditure on M is given by $E_1=\mu Y_1+\alpha n p_1 \beta x_1$. The term μY_1 in the previous expression is the share of final demand (and, since there are no savings, income) spent on M; it follows from (1-1). The second term in the expression for E_1, $\alpha n p_1 \beta x_1$, is the share of intermediate demand spent on M that emanates from other manufacturing firms. It can be inferred from (1-2) using Shepard's lemma. By analogy, location 2 expenditure on M is given by $E_2=\mu Y_2+\alpha(1-n)p_2\beta x_2$.

To close the model, note that the value of total output in sector M at producer prices must equal the value of global M-sector private expenditures, viz. $np_1x_1+(1-n)p_2x_2=\alpha\beta[np_1x_1+(1-n)p_2x_2]+\mu[L+\pi]$. Making use of the pricing rules and the free-entry condition (1-3), we get $\pi=\mu L/[(1-\alpha)\sigma+\alpha-\mu]$.[21] Observe that the equilibrium π is function of parameters and exogenous endowments only; in particular, this expression holds for any n. Importantly, it does not depend upon ϕ or τ.

As an aside, we now have everything at hand to make the no-specialisation condition more precise: if all firms cluster in a single location, we require the labour supply of this region to be larger that the labour these firms demand so that sector A is active in both regions and the law of one price prevails on market A. Mathematically, this requires $\min\{L_1,L_2\}>(1-\alpha)\beta\sigma\pi$. Using the equilibrium expression for π, the closed-form condition is $(1-s)>(1-\alpha)(\sigma-1)/(\sigma(1-\alpha)+\alpha-\mu)$. We assume it holds throughout.

Short run and long run equilibria

In the short run capital is immobile while it becomes mobile only in the long run. Thus, in a short run equilibrium consumers maximize utility, firms maximize profits, and all markets clear.

Define q_j as the ratio of the actual operating profit in region j to the equilibrium value of π, that is, $q\equiv\pi_j/\pi$, and e_j as the share of expenditure that emanates from region j,

---

[20] More precisely, the necessary condition for an equilibrium is $n_j(\pi_j-\pi)=0$, with the possibility that $\pi_j<\pi$ if $n_j=0$.

[21] Economic consistency imposes $\pi>0$, which requires $(1-\alpha)\sigma+\alpha-\mu>0$. This always holds because $\sigma>1$ and $1>\mu$. If K had not been normalised to unity, the expression for π in the text should be divided by K.

viz. $e_j \equiv E_j/\sigma\pi$.[22] Together with the expressions for p and E_j above, we obtain the following closed form solutions for e_j:

(1-5) $e_1 = s + \alpha\beta(n-s) + \alpha\beta n(q_1 - 1);\ e_2 = (1-s) - \alpha\beta(n-s) + \alpha\beta(1-n)(q_2 - 1)$

We then use Sheppard's lemma and Roy's identity to get the demand for a typical variety, $p_j^{-\sigma}\mu E_j/\Delta_j$. Using this alongside (1-2), (1-3) and (1-5), we obtain the following operating q-ratios:

(1-6) $$q_1 = \left(\frac{e_1}{\Delta_1} + \phi\frac{e_2}{\Delta_2}\right)\Delta_1^{\alpha};\quad q_2 = \left(\frac{e_2}{\Delta_2} + \phi\frac{e_1}{\Delta_1}\right)\Delta_2^{\alpha}$$

It is obvious from the definition of π and the q's that $n \in (0,1)$ implies, first, $q_1<1$ if, an only if, $q_2>1$ and (and conversely) and, second, $q_1=1$ if, an only if, $q_2=1$. In words, the firms located in a given region make above normal operating profits whenever the firms located in the other one make below normal operating profits.

The system (1-3)-(1-6) completely characterizes the so-called *instantaneous equilibrium*. (In an instantaneous equilibrium, n is an exogenous variable and q_1 and q_2 are functions of n.) In the long run, n adjusts so that $\pi_j=\pi=\mu L/[(1-\alpha)\sigma+\alpha-\mu]$ (and hence $q_j=1$) for any active firm. For the time being we assume that capital owners allocate their capital according to *current* nominal differences in rewards according to the following ad-hoc law of motion for n:

(1-7) $$\dot{n} = \gamma n(1-n)(\pi_1 - \pi_2) = \gamma n(1-n)(q_1 - q_2)\pi$$

where γ is a strictly positive parameter and the second equality follows from the definition of π. The long run equilibrium is attained whenever $\dot{n}$ is zero. Three cases can occur: n=0 (in which case $q_2=1$), n=1 (in which case $q_1=1$), and $0<n<1$ (and hence $q_1=q_2=1$). The first two cases are usually referred to as 'core-periphery' equilibria and the third as interior or 'dispersed' equilibria. By the symmetry of the model, the symmetric equilibrium n=½ always exists. More generally denote an interior long run equilibrium as n^0.

To assess the stability of these equilibria the NEG typically resorts on the following informal methods. Staring from any long run equilibrium, the allocation of capital is hit by an exogenous perturbation. For the interior equilibria (in particular the

---

[22] The use of the 'q' notation in this static model is deliberate: in a straightforward dynamic extension of the model, π would play the role of the replacement cost of capital.

symmetric equilibrium n=½), one evaluates the sign of the change in the nominal profit gap, viz. π_1-π_2. If the displaced unit of capital increases the profit in the receiving region, then the symmetric equilibrium is unstable.

For the agglomerated (or core-periphery) equilibrium, one checks whether the perturbation creates a nominal profit in the periphery that is higher than the nominal profit in the core. If this is the case then this equilibrium is unstable.

Mathematically, these two tests can be written as:

(1-8) $$\left.\frac{d(\pi_1 - \pi_2)}{dn}\right|_{n=n^0} < 0, \quad (\pi_1 - \pi_2)\big|_{n=1} > 0$$

The equilibrium under consideration is stable when the relevant inequality holds. It is unstable otherwise.

Before describing the forces that compete to make long run equilibria stable or unstable (the topic of Section 1.3), I briefly justify the use of the law of motion for capital (1-7) and the informal stability tests in (1-8) following the method introduced by Baldwin (2001) for the truly dynamic version of the CP model.

Dynamic optimal allocation

Assume that the instantaneous utility function is as in (1-1) and that inter-temporal preferences are represented by:

(1-9) $$\bar{V}(t) = \int_{s \geq t} e^{-\rho(s-t)} V(s) ds$$

where ρ is the subjective discount rate.

Assume further that the management of the portfolios is entrusted to a-spatial funds that maximise the shareholder's nominal value. These funds maximises the capital earning of their customers less some adjustment cost.[23] For simplicity, we assume that re-location costs are quadratic to the flow of capital and proportional to the sending and receiving regions capital stocks. Portfolio value maximisation thus requires each fund to

---

[23] We can justify the assumption that these funds maximise nominal returns as follows. First, the market for funds is perfectly competitive: each of these funds is atomistic. Second, each of these funds buys shares of a sub-sample of firms, which mass is negligible. These two assumptions together have two implications. On the one hand a typical capital owner, whose capital endowment is mass-less, will invest in funds that maximise her nominal earning (the location decisions of the firm managed by these funds having a negligible impact on the price index). On the other hand, by free-entry and exit in the market for funds, each active fund will propose a gross return of π.

chose the optimal capital allocation time path to solve the infinite, discounted stream of operating profits net of 'migration' cost:

(1-10) $$\max_{m} \int_{t\geq 0} e^{-\rho(s-t)} \left(n\pi_1 + (1-n)\pi_2 - \frac{\Gamma m^2}{2n(1-n)} \right) dt; \quad m = \dot{n}$$

where $\Gamma>0$ is a parameter. This allows for almost any sort of migration behaviour (see Baldwin 2001).

Taking m as the control variable, the current-value Hamiltonian for this problem is the term in the brackets in (1-10) plus $m\lambda$, where λ is the co-state variable of this problem; λ captures the asset value of capital 'migration'. The necessary conditions for an optimum are

(1-11) $$\forall t: \quad m = \frac{\lambda n(1-n)}{\Gamma}, \quad \dot{\lambda} = \rho\lambda - (\pi_1 - \pi_2)$$

and the transversality condition $\lim_{t\to\infty} e^{-\rho t}\lambda m=0$. Standard differential calculus implies that the migration equation and the asset-value of migration respectively satisfy:

(1-12) $$\dot{n} = \frac{n(1-n)\lambda}{\Gamma}, \quad \dot{\lambda} = \rho\lambda - (\pi_1 - \pi_2)$$

If expectations are static, fund managements assume that the current gap will persist forever (Baldwin (2001). In this case, choosing parameter values such that $\Gamma=\gamma\rho$ and solving (1-12) for λ yields $\lambda=(\pi_1-\pi_2)/\rho$ and:

(1-13) $$\dot{n} = \gamma n(1-n)(\pi_1 - \pi_2)$$

which is identical to (1-7). To sum up, we have:

> Result 1-1. The law of motion (1-7) is consistent with optimal behaviour from forward-looking fund managers with quadratic migration.

Still building on the work of Baldwin (2001), it is easy to show that the following couple of results also hold:[24]

> Result 1-2. The 'informal stability tests' in (1-8) correspond exactly to the formal, local stability of the model.

---

[24] This is taken almost verbatim from Baldwin's (2001) paper.

Result 1-3. (a) The model is globally stable. (b) Informal methods to determine the where the system ends up is confirmed by the formal 'Liaponov's Direct Method'.

Point (a) above says that the system drives the system to anyone of the steady states, regardless of initial conditions. The informal methods point (b) refers to are the following. By the definition of a stable interior steady state (1-8) and by continuity of the short-run equilibrium as a function of n, it must be that stable and unstable equilibria alternate on the n scale. The informal methods claims that if n is at some n' between n^0 and n'', where n^0 is part of a stable long run equilibrium and n'' is part of an unstable long run equilibrium given the value of the parameters, then the model will converge to n^0.

With Result 1-1 at hand, I follow the standard but implicit practice in the NEG literature in assuming that expectations are static and adjustment costs of (capital) migration are quadratic.[25] Also, as a consequence of Result 1-2 and Result 1-3, I follow the tradition in the NEG literature and keep the discussion of the stability properties of the models of this chapter and next ones at a heuristic level.

The description of the model is now complete: the long run equilibria consist of the values of n in the interval [0,1] that solve (1-3)-(1-7) for $\dot{n} = 0$.

1.3. Agglomeration and dispersion forces

The model as described in the previous section features both agglomeration and dispersion forces. These forces are of the same nature in all models reviewed in the introductory chapter. The dispersion forces usually present in the NEG models are also present in the 'New trade models'. What sets the NEG apart in the literature is the presence of agglomeration forces. These forces are usually of two sorts –the 'backward' and 'forward linkages'. We start with the former linkage, which can be thought of as a demand linkage.

Backward linkage

This linkage is best illustrated in expression (1-5) and (1-6) keeping the Δ's equal and constant. Start from a 'long-run' configuration in which firms are active in

---

[25] Baldwin also carries the analysis for the case of forward-looking agents.

both regions so that $q_1=q_2=1$. Now imagine a firm originally located in 2 relocates in region 1, i.e. $dn>0$. Since this firm buys intermediates from other firms, this increases (decreases) expenditures on manufacturing goods that emanates form region 1 (region 2), viz. $de_1>0$ ($de_2<0$). Since manufacturing firms sell their output with a positive mark-up and transportation costs are strictly positive (viz. $\phi<1$) this *expenditure shifting* gives rise to *profit shifting*. Mathematically, this can be seen using (1-6): obviously q_1 increases if e_1 increases and e_2 decreases, keeping Δ_1 and Δ_2 is constant. The opposite is true for q_2. From (1-7), it is clear that this profit shifting gives rise in turn to *production shifting*, viz. $dn>0$, so the cycle repeats.

The next agglomeration force is the 'forward linkage', which can be thought of as a cost linkage.

Forward linkage

To illustrate the cost linkage, we use the same set of equations as before and make a similar thought experiment. However, we now keep the e's constant as well as the Δ's at the denominator of (1-6). As it turns out, the Δ's play two conceptually distinct roles, so it is useful to distinguish them in order to fully understand the mechanisms at work. We are here interested in the Δ's that enter the numerator of (1-6) at the power α. Using (1-4), we claim:

Result 1-4. Δ_1 is maximised at $n=1$ (by symmetry, Δ_2 is maximised at $n=0$). Moreover, Δ_1 increases relative to Δ_2 as n increases.

Proof. The proof of the first claim is easy. Remember that $\Delta_1 \in (0,1]$, so Δ_1 cannot be larger than one. Next, it is immediate from (1-4) that $n=1$ implies $\Delta_1=1$. Turn now to the second claim. Define Δ as Δ_1/Δ_2 and η as $n/(1-n)$. Clearly, Δ and η are respectively increasing in Δ_1 (decreasing in Δ_2) and n. Using (1-4) and the new notation, the latter definition of the price indices can be written as $\Delta^{\alpha}\eta=(\Delta-\phi)/(1-\Delta\phi)$. We shall see in Chapter 2 that $\phi<\Delta<1/\phi$ always hold. Simple algebra then reveals that $\partial\Delta/\partial\eta>0$ and which implies$\partial\Delta/\partial n>0$ by definition of η. *QED.*

The interpretation of this result is straightforward if we remember that Δ_1 is negatively related to the price index of region 1, G_1 (and hence to the marginal cost of producing in 1): it says that the more numerous are the firms that locate in 1, the lower is the price of intermediates in 1 relative to 2. This directly follows from the assumption that trade/transportation costs are positive.

Since intermediate inputs enter with a share α in the cost function, and because f.o.b. prices are proportional to marginal costs, q_1 is proportional to Δ_1^{α}. Clearly, if n increases it becomes relatively cheaper to produce in 1 and more expensive to produce in 2. As a result, profitability in 1 (in 2) rises (decreases). In other words, this *production shifting* gives rise to *cost shifting* and, in turn, to *profit shifting*. By (1-7) $dn>0$ and the cycle repeats.

Together, the forward and the backward linkages are referred to as 'vertical' (or output-input) linkages, as they both would be inexistent if firms did not buy some of each other's output (which is the case if $\alpha=0$). In other words when $\alpha=0$ cumulative agglomeration forces are absent. Clearly, if the vertical linkages were the only forces at work production would always be located in a single region. However, there also exists a dispersion force that offsets these agglomeration forces –the 'market crowding effect'.

Market crowding

To illustrate this effect as clearly as possible, we impose $\alpha=0$ to turn off the agglomeration forces. This implies that the e's as constant. For simplicity, take $s=½$ so that $e_1=e_2=½$. The objects of interest now are the Δ's at the denominators of q_1 and q_2 in (1-6). These can be thought of as market shares.[26] Indeed, on the numerators are the expenditure shares of each region; a given firm gets only a fraction of this and this fraction is lower, the larger is the mass of competitors.

Clearly, $d\Delta_1/dn>0$ and $d\Delta_2/dn<0$ imply that the relocation of a firm from 2 to 1 increases both the domestic and foreign market shares of firms that stay put in 2 and decreases the market shares of firms in 1. Since profits are increasing in market shares, this initial *production shifting* implies a *negative profit shifting*; by (1-7) $dn<0$ and hence this counterweights the initial increase in n. If this force dominates the

---

[26] For the general case $\alpha \geq o$, the term $\Delta_1^{\alpha}/\Delta_1$ in the expression for q_1 corresponds to the share that a typical firm in 1 enjoys on the domestic market. Conversely, the term $\phi\Delta_1^{\alpha}/\Delta_2$ corresponds to the share a typical firm located in 2 enjoys on market 1.

agglomeration forces surveyed above, then any exogenous shock of the form dn>0 or dn<0 is self-correcting. Note that this force is always present –even when $\alpha=0$. This contrasts with the agglomeration forces discussed under the two previous headings.

Turn now to the formal analysis of the interplay of these forces and ask when the dispersion force dominates the agglomeration forces, and vice-versa. I pursue this analysis under two headings. On the one hand, the model with $\alpha>0$ is a NEG model: generically, it displays multiple equilibria and the production structure of ex-ante identical regions (i.e. s=½) typically diverges when trade/transportation costs are low enough (this is the topic of Section 1.4). The application of Chapter 4 will be using the resulting model. On the other hand, Section 1.5 conveys the analysis for $\alpha=0$. This simplified version will be used in the application of Chapter 3.

1.4. Symmetric regions and vertical linkages: The FCVL model

Following the tradition of the NEG, my primary interest here is to discuss how regions that share identical tastes, endowments, and technology might endogenously diverge in terms of production structure and real incomes. Therefore, we impose s=½ in the remainder of the section. As we shall stress, the resulting model features the same qualitative results of the CPVL model of Fujita et al. (1999, chapter 14.2).

(Un)stability of the dispersed equilibrium

We define a dispersed equilibrium as the configuration in which n=½. With symmetric regions such an equilibrium always exists as can be seen from (1-4)-(1-7). It might not always be stable in the sense of (1-8), though. Here the thought experiment is, if a firm moves from region 2 into region 1, in which case would the gap π_1-π_2 thus created be positive and hence, by (1-7), widening? In which case would that gap be negative and hence the perturbation be self-correcting? Formally, answering this question is equivalent to signing dq_1/dn evaluated at n=½.[27]

As we know from the symmetry of the model, n=½ implies $e_1=e_2$, $\Delta_1=\Delta_2$, and $q_1=q_2$. We therefore denote common variables with the nought subscript. From (1-4)-(1-6), we find that $q_0=1$, $e_0=½$, and $\Delta_0^{1-\alpha}=(1+\phi)/2$. Taking first derivatives at the

---

[27] Or, which is equivalent by the symmetry of the model when s=½, to signing $d(q_1-q_2)/dn$ at n=½.

symmetric equilibrium, defining Z as $(1-\phi)/(1+\phi)$, and using $de_0 \equiv de_1 = -de_2$, $d\Delta_0 \equiv d\Delta_1 = -d\Delta_2$, and $dq_0 \equiv dq_1 = -dq_2$, we obtain:

$$\text{(1-14)} \qquad \begin{bmatrix} 0 & 1-\alpha Z & 0 \\ 2 & 0 & -\alpha\beta \\ -2Z & Z-\alpha & 1 \end{bmatrix} \begin{bmatrix} de_0 \\ d\Delta_0 / \Delta_0 \\ dq_0 \end{bmatrix} = 2 \begin{bmatrix} Z \\ \alpha\beta \\ 0 \end{bmatrix} dn$$

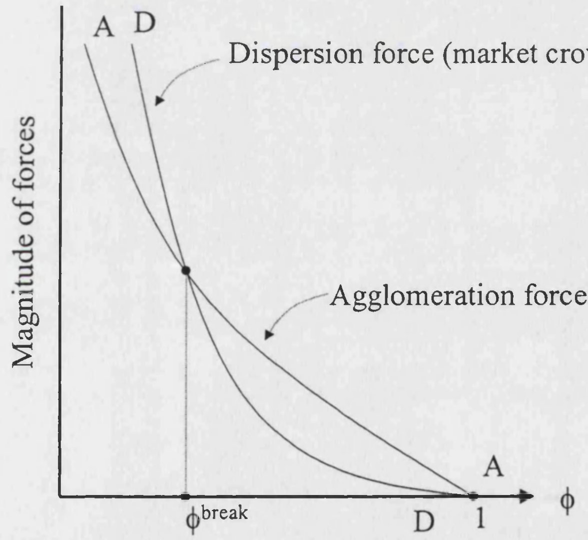
Using Cramer's rule and solving $dq_0/dn=0$ for $Z>0$ gives the corresponding break point, defined as:

$$\text{(1-15)} \qquad \phi^{break} \equiv \frac{(1-\alpha)(1-\alpha\beta)}{(1+\alpha)(1+\alpha\beta)}$$

which is strictly smaller than unity by inspection. To sum-up we can write:

Result 1-5. The symmetric equilibrium is stable for all ϕ below ϕ^{break}.

Figure 1-1: Dispersion and agglomeration forces at the symmetric equilibrium



This suggests a picture along the lines of Figure 1-1, which depicts the magnitude of the forces discussed in Section 1.3 at the symmetric equilibrium. The degree of trade free-ness ϕ is depicted on the horizontal axis; when $\phi=0$ trade in manufacturing goods is prohibitive. When $\phi=1$ trade in M is perfectly free, so the two regions are perfectly integrated and location no longer matters. AA plots the agglomeration forces and DD plots the dispersion forces. It can be shown that both agglomeration and dispersion forces decrease with ϕ (Baldwin et al. 2001, 2002), so both AA and DD are downward sloping. Result 1-5 suggests that agglomeration forces dominate dispersion forces if ϕ is large enough and that the two curves cross only once. In other words, when ϕ is low dispersion forces take the upper hand. When ϕ increases, both dispersion and agglomeration forces decrease but the former decreases faster; eventually the later dominates. Formally, this is so whenever $\phi>\phi^{break}$.

The symmetric equilibrium is unstable for lower values of ϕ if agglomeration forces are stronger. Ceteris paribus, agglomeration forces are increasing in α and β. α captures the strength of vertical linkages among firms (when α is large, firms buy a lot of each others' output as intermediate inputs). To a large β corresponds a low elasticity of substitution σ; the less different varieties are substitutes, the larger is the market power of each firm and hence the more sensitive profits are to changes in the conditions firms operate.

Using (1-8) again, we now ask a rather different question, namely, if all firms are set in one region, does any of them have any incentive to deviate?

Sustainability of the concentrated equilibrium

Here the question is, is a core-periphery pattern sustainable? To answer this question, we check under which conditions n=1 and $q_2 \leq 1$ hold simultaneously. No firm wants to leave 1 if the shadow profit in 2 is inferior to π. Substituting n=1 into (1-4)-(1-6), we find that this condition holds whenever $\phi \in [\phi^{sust}, 1)$, where ϕ^{sust} is implicitly defined as the smallest root of the following polynomial (1 is the unique other root):

(1-16) $$2(\phi^{sust})^{1-\alpha} - [1+\alpha\beta](\phi^{sust})^2 - [1-\alpha\beta] = 0$$

In words, whenever trade costs are low enough, if it already happens that all firms have agglomerated in either region, then none has any incentive to leave the core and start producing in the periphery. To sum-up, we have:

Result 1-6. The Core-periphery equilibrium is said to be sustainable if ϕ is larger that ϕ^{sust}.

It can be shown that $\phi^{sust}<\phi^{break}$. I show this formally in Chapter 2.

Other equilibria and global stability

Thus far we have shown that long run equilibria in which $n \in \{0, ½, 1\}$ exist and under which conditions they are stable. Now the following natural question may arise. Do there exist other equilibria, and what are their dynamic properties if they do?

Simulations undertaken by several researchers provide the following answer to this question. Generically, there are five equilibria: the dispersed equilibrium (n=½), the concentrated equilibria (n=0,1), and two asymmetric, interior equilibria n' and n" with n'+'n"=1. Moreover, whenever they exist, these equilibria are unstable in the sense that $dq_1/dn>0$ and $dq_2/dn<0$ at n=n',n". The analytical proof of these claims is a formidable task that is undertaken in Chapter 2. The methodology in this chapter works by showing that a whole class of models can be written in a more natural state space. The current model is no exception to this. An appendix at the end of this chapter writes the model of this section in this 'natural' state space.

Comparison with the break and sustain points of the CPVL model

The reader familiar with the NEG literature already knows that break and sustain points in the CP and CPVL models are isomorphic, as are the corresponding point of the FCVL model of this section. In particular, the break and sustain points of the CPVL model solve:

(1-17)
$$\phi_{CPVL}^{break} = \frac{(1-\alpha)(\beta-\alpha)}{(1+\alpha)(\beta+\alpha)}, \quad 2(\phi_{CPVL}^{sust})^{1-\frac{\alpha}{\sigma-1}} - [1+\alpha](\phi_{CPVL}^{sust})^2 - [1-\alpha] = 0$$

The model requires the so-called 'no black hole conditions' $\beta>\alpha$ to hold. Without these, the break points would be negative, implying that the symmetric equilibrium is never stable. The similarity between the two models is striking as both the functional forms for the cost functions and the mechanism driving agglomeration are different. In particular, agglomeration stems from labour mobility between the manufacturing and the background sectors within each region the CPVL model. By contrast, international capital mobility is the driving mechanism in the FCVL model.

The similarities among different NEG models run deep and wide, as we shall see extensively in Chapter 2.

1.5. Asymmetric regions in the absence of linkages: The FC model

In this section we withdraw the restriction s=½ that we imposed earlier. We also allow for asymmetries in transportation costs. The problem, however, is that the model becomes so highly intractable that virtually no analytical solutions can be derived. The source of the problem is the implicit and simultaneous definition of the Δ's in (1-4) whenever $\alpha>0$. Is it to say that nothing can be said in the case of asymmetric regions?

The answer to this question is negative. It turns out that all the complications wash away if we no longer assume that firms buy each other's output as intermediates.[28] Hence, in the remainder of the chapter we assume that vertical linkages are absent, viz. $\alpha=0$. With this parameter restriction, the resulting model collapses to the model used by Flam and Helpman (1987) and Martin and Rogers (1995).

Note an important implication of this assumption. The very existence of agglomeration forces depends on the existence of vertical linkages, as was shown in Section 1.3. Since we assume $\alpha=0$ the market crowding dispersion force alone remains. Hence, the model will display a unique stable equilibrium akin to the symmetric equilibrium in the symmetric-region case. Graphically, this means that the AA curve in Figure 1-1 is flat and is identical to the zero horizontal line. Hence, the break point is 1 in which case does not matter anyway.

As we shall see, the location equilibrium is now a smooth function of ϕ.

Formal analysis of the location equilibrium

Here we consider asymmetric regions and asymmetries in trade cost by allowing region 1's trade free-ness parameter to differ from that of region 2. We refer region 1's as ϕ and region 2's as ϕ^*; when $\phi<\phi^*$ it is cheaper for region 1's firms to export to market 2 than for region 2's firms to export to market 1. Carefully tracing through the impact of this change on the mill pricing of region 1 and region 2 firms and thus on operating profit, we can easily establish the more general version of (1-6) and (1-7):

---

[28] Baldwin et al. (2001, 2002) characterize some aspects of the case of asymmetric regions in the CP model.

(1-18) $$q_1 = \frac{e_1}{\Delta_1} + \frac{\phi^* e_2}{\Delta_2}; \quad q_2 = \frac{e_2}{\Delta_2} + \frac{\phi e_1}{\Delta_1}; \quad \Delta_1 = n + \phi(1-n); \quad \Delta_2 = \phi^* n + (1-n)$$

where e_1=s and e_2=1-s, as can be seen from (1-5). In words, expenditure now consists only on consumer expenditure; given the structure of the model nominal incomes are identical, hence consumer expenditure is proportional to income. Region 1 is endowed with a share s of world endowments, so it owes a share s of world income.

Solving these for q_2=q_2, implies that there is only one interior equilibrium:

(1-19) $$n = s + (2\phi)\frac{s-½}{1-\phi^*} + \frac{\phi^* - \phi}{(1-\phi)(1-\phi^*)} - (1-2\phi)\frac{(s-½)(\phi^* - \phi)}{(1-\phi)(1-\phi^*)}$$

This expression is valid for ϕ's where n lies between zero and unity. For ϕ's outside this range n is either zero or unity as appropriate. For the sake of illustration assume n≥½ and $\phi \geq \phi^*$ unless otherwise specified. We can infer four things from (1-19). First, n is larger than s if s>½ or $\phi^* > \phi$ or both. Market access considerations induce firms to locate in the larger market (market 1 by s>½) or in the market from which it is easiest to reach final consumers located in the other region (market 1 by $\phi < \phi^*$). The market crowding effect will temperate thcsc considerations, so generally 0<n<1.

Second, the sign of the interaction term in (1-19) is ambiguous. When ϕ is large location is extremely sensitive to any asymmetry between the two regions, hence the interaction factor is positive. When ϕ is low, by contrast, the market crowding effect is large and the tendency of firms to locate far apart is strong, so the additional benefit brought about by better market access is decreasing. Put differently, s-½ and ϕ^*-ϕ are the same thing –they both capture the market access advantage of market 1– and (1-19) says that n is increasing and concave (convex) in market access if ϕ is small (large).

Third, (1-19) reveals that n can be larger of smaller than s, but we know for sure that n increases with the gap in market access, viz. ϕ^*-ϕ. To see this, note that the sum of the third and last terms in the right-hand side of (1-19) is proportional to 1-(1-2ϕ)(s-½), which is strictly positive because both (1-2ϕ) and (s-½) are smaller than unity in absolute value.

Finally, it can be shown that n is increasing in (s-½) by the same token. Additionally, assume now that $\phi = \phi^*$. Then n>s if and only if, s>½: the large market attracts a more than proportional share of industries and hence exports the good for which it has an unusually large demand. This is a manifestation of the 'home market effect' (HME) of Krugman (1980). More precisely, the HME is defined as

(1-20) $$\partial \frac{n/(1-n)}{\partial(s/(1-s))} > 1$$

for values of ϕ such that n is an interior solution to (1-19). Note also that the HME is magnified by the degree of trade free-ness ϕ: as location becomes more sensitive for low levels of trade/transportation costs, any difference in income is amplified at the location equilibrium. To sum-up we can write:

> Result 1-7 (HME). Ceteris paribus, the largest region attracts a more than proportional share of firms, viz. $n>s \Leftrightarrow s>\frac{1}{2}$. Moreover, the HME is magnified by the magnitude of trade freeness.

Asymmetric Relative Factor Endowments

The model can also easily handle regions that are asymmetric in terms of their K versus L endowments. Until now, s has denoted region 1's endowment of both labour and capital relative to the world aggregate. By implication, s also denotes region 1's share of income. Indeed, $Y_1=s_L L+s_K \pi K$ which is equal to $s(Y_1+Y_2)$ if $s_L=s_K=s$. Under this heading, we allow s_L and s_K to differ and impose $\phi^*=\phi$. Using the equilibrium expression for π, it is easy to see that the following relationship

(1-21) $$s=(1-\frac{\mu}{\sigma})s_L+\frac{\mu}{\sigma}s_K$$

is an identity at equilibrium. This defines s as a weighted average of s_K and s_L. One interesting question in the context of distinct relative factor endowments is the direction of capital flows, which boils down to the sign of s_n-s_K. If this difference is positive region 1 employs more of the world's capital than its own, so it must be a capital importer. If the difference is negative the region 1 is a net capital exporter.

By Result 1-7 we know that if region 1 is bigger but $s_L=s_K=s$ then region 1 will be a capital importer. An interesting case is when region 1 is both larger and relatively well endowed with capital. Suppose $s_K>s>\frac{1}{2}$ and $\phi^*=\phi$ for simplicity. In this case, the region 1's relative abundance of capital tends to offset the home market effect. In particular, manipulating (1-19) and (1-21), we get:

$$n - s_K = \frac{2\phi}{1-\phi}(s - \frac{1}{2}) - (s_K - s) \quad \textbf{(1-22)}$$

This shows that if region 1's factor endowment is sufficiently skewed towards capital then region 1 may be a capital exporter despite the home-market effect. However, the home-market effect will eventually dominate for sufficiently low trade costs. See Baldwin and Robert-Nicoud (2000) for the general case in which all asymmetries are simultaneously present.

1.6. Welfare analysis

In the chapter thus far the analysis has concentrated on positive issues exclusively. The following sections deal with normative issues. The discussion is organised around the FC and the FCVL models of this chapter. From an equity-perspective one may ask: Who are the gainers and the losers from agglomeration? Which factor owner benefit and which suffer? Which regions are advantaged and which are disadvantaged? From an efficiency-perspective one may wonder: Can the gainers compensate the losers? Does the free working of market forces deliver the optimal degree of agglomeration? If not, is there too much or too little agglomeration for the economy as a whole? The aim of the next couple of sections is to hint at the answers of this kind of questions.

1.7. Equity and Efficiency in the Footloose Capital Model

We start with the most tractable model, the FC model of section 1.5. The analysis borrows extensively from Baldwin et al. (2001).[29]

Thus far we have assumed that capital ownership was uniform across each the populations of each regions. The algebra of the model carries over other factor ownership assumptions. In particular, we could assume that each individual owes either one unit of labour or one unit of capital. Potentially, this gives rise to 'class conflicts.' We take this interpretation of the model here so that we can use the terminology 'worker' and 'capital owner' as a short hand. It should be understood that an individual who owes both labour and capital might see her welfare unaffected even though her wage increases and her capital reward decreases.

---

[29] See also Ottaviano, Tabuchi and Thisse (2002), Ottaviano and Thisse (2002), and Ottaviano (2002).

At the finest level, each factor located in each region may have distinct and frequently conflicting welfare concerns. Conflicts of interest between these various classes of agents arise along two dimensions. The first dimension is spatial: viewing all factors within a region as a group, the question is: How industrial location affects living standards of each region? The second dimension distinguishes between agents according to their factor ownership (labour versus capital).

Pareto welfare analysis

In this section we want to assess: (i) the welfare of each group of individuals at the market equilibrium; (ii) which groups of individuals can be made better off by a *fiat* relocation of M-firms when no inter-group transfers are allowed for ('Pareto welfare analysis'); (iii) how the economy as a whole can be made better off by a *fiat* relocation of M-firms when transfers are available ('global welfare analysis').

Welfare analysis is particularly simple in the FC model. There are four groups: capital owners in regions 1 and 2 and workers in regions 1 and 2. In this section I assume that factor ownerships are degenerate for simplicity: workers own no capital and vice versa. The nominal incomes (i.e. incomes measured in terms of the numéraire) of all four groups are independent of the spatial allocation of industry n and transportation costs ϕ so all welfare effects stem from the cost-of-living effects, i.e. changes in the Δ's. (Remember that the cost of living in region j is equal to $\Delta_j^{\mu/(1-\sigma)}$.) Transportation costs imply that the cost-of-living is lowest in the region with the most industrial firms. Consequently, any change in the location of firms that increases Δ_1 will decrease Δ_2, and vice versa. Therefore, we have:

> Result 1-8. In the FC model conflicts of interests arise on the spatial dimension only. Indeed, any spatial reallocation of capital (industry) benefits one region at the expense the other. Moreover, there is no conflict between capital owners and workers who live in the same region.

This implies that no Pareto improvement is feasible but it is legitimate to ask whether a planner with a utilitarian social welfare function can improve on the decentralized equilibrium. I turn to this issue next.

Global welfare analysis

This sub-section derives the utilitarian social planner's optimum. For simplicity I work with the logarithmic transformation of (1-1). Accordingly, we can write the utilitarian objective function as:

(1-23) $$W = W_0 + \frac{\mu}{\sigma - 1}[(s_L L + s_K)\ln\Delta_1 + ((1-s_L)L + 1 - s_K)\ln\Delta_2]$$

where W_0 collects all the (indirect) utilities derived from the factor rewards. This is constant at equilibrium since factor rewards are constant themselves. Note that the aggregate mass of capital owners is one by the normalisation of section 1.2.

In principle there are many potential sources of inefficiency a planner may want to deal with. First, firms price above marginal cost. Second, capital owners choose where to offer their services without taking into account the impact of their decisions on consumer surpluses in the two regions. Third, they also do not take into account the impact on firms operating profits. Thus the agglomeration and dispersion forces of section 1.3 result in pecuniary externalities.[30]

In the first-best outcome all distortions are removed. As part of this the planner imposes price equal to marginal cost β for both local and export sales, so the local price of a typical variety will be (1-1/σ) and the export price will be τ times this. The resulting Δ_j is just (1-1/σ) times the market one. Of course, marginal cost pricing drives operating profits to zero, so lump-sum transfers are needed to support capital-owners' consumption. Following Baldwin et al. (2001) we can assume that these transfers are set at some exogenously determined level to avoid unenlightening complications.

It is however often admitted that marginal cost pricing cannot be imposed, either because lump-sum transfers from consumers to firms are not available, or because the degree of surveillance necessary to enforce it is impractical. This implies that firms set prices above marginal cost as usual. As a result the second-best planner tackles the last two inefficiencies. Note that under the specific assumptions of the model the first-best objective function is identical to the second-best objective function up to a constant (the welfare gains associated with marginal cost pricing are independent of n). This is due to

[30] Moreover, the effects on consumer surpluses and operating profits are also neglected in the entry decisions by firms so that generically the number of firms operating is sub-optimal. In the present setting, however, this last source of inefficiency is inexistent because the number of firms is determined only by the M-good fixed costs and the aggregate endowments of capital, so entry is optimal.

the conjunction of the following assumptions and results. First, the number of varieties is fixed by capital endowment. Second, transportation costs take the iceberg form. Finally, mill pricing is optimal in Dixit-Stiglitz monopolistic competition. This last point implies that the ratio of imported to locally produced varieties is exactly τ (exogenous) in both the first and second best cases, so above-marginal-cost pricing does not distorted the location decision. As a result the planner maximizes an objective function that equivalent to (1-23) in either case. To sum-up, we have:

> Result 1-9. The first- and second-best geographical distributions of firms coincide.

Therefore, we can use (1-23) to derive the optimum spatial allocation of firms. The first-order condition to the planner's problem is:

(1-24) $$\frac{dW}{dn}=(L+1)\left[\frac{d\ln\Delta_1}{dn}s_{pop}+\frac{d\ln\Delta_2}{dn}(1-s_{pop})\right]=0;\; s_{pop}\equiv\frac{s_L L+s_K}{L+1}$$

where s_{pop} is defined as the share of the world's population living in 1. The second-order condition is satisfied by concavity of (1-23). This shows that the optimising planner must strike a balance between the opposing effects of changing n on individual welfare. Using (1-18), (1-24) can be rewritten as:

(1-25) $$(1-\phi)\left[(1+\phi)(s_{pop}-\frac{1}{2})-(1-\phi)(n-\frac{1}{2})\right]=0$$

This holds either when trade is perfectly free (viz. $\phi=1$) because the planner is indifferent to firm location since this is irrelevant to consumers' welfare. Or, when transportation costs are positive (viz. $\phi<1$), (1-25) reveals that the planner chooses n by balancing two opposing effects. The first effect –the transportation cost saving effect– is $(1+\phi)$ times the region 1's share of world population, which depends on the spatial endowments of workers and capital owners. The second is $(1-\phi)$ times the region 1's share of industry. The second effect is what Baldwin et al. (2001) call the individual welfare effect. By the concavity of utility, the welfare trade off between region 1's residents and region 2's worsens as the division of industry becomes more extreme. Thus the individual welfare effect is a force that favours an even distribution of industry.

Solving (1-25) for n gives the optimal allocation of firms n* where

(1-26) $$n^{*} = s + (2\phi)\frac{s_{pop} - ½}{(1-\phi)}$$

if $s_{pop} \in [\phi/(1+\phi), 1/(1+\phi)]$; outside this range, the planner concentrates all production in the big region. To summarise, we write:

Result 1-10 (Social home-market effect). The socially optimal spatial allocation of industry requires the large region to have a more than proportional share of industry.

By implication Result 1-7 continues to hold qualitatively.

We are now ready to establish whether there is too much or too little spatial concentration of industry in the market equilibrium. All we have to do is to compare n with n*. Taking the difference between (1-19) with $\phi^*=\phi$ and (1-26), we obtain:

(1-27) $$n - n^{*} = \frac{1+\phi}{1-\phi}(s - s_{pop})$$

This shows that any difference between the market and social allocation of industry depends upon the difference between region 1's share of world expenditures (viz. s) and its share of world population (viz. s_{pop}). The reason is that the utilitarian criterion (1-23) puts equal weights on each individual regardless of their incomes and places of residence. By contrast, the market cares about expenditures, which implies that richer individuals count more. This implies:

Result 1-11. The market outcome has too many firms in the region that has the highest per capita income.

Per capita income depends on two things: the region's relative factor endowment and the relative reward of the two factors. To take a natural case, suppose the income of capital owners is higher than that of workers (this requires $L>(\sigma-\mu)/\mu$). In this case, the region relatively well endowed with capital is also richer and there will be too many firms located in that region at the laissez-faire equilibrium. If this region is also the largest, then there is too much agglomeration.

Result 1-12. There are two special cases when the market outcome is optimal, when the two regions are scaled versions of each other (in which case $s_L=s_K$) or when remunerations are equalized across factors, viz. $\pi=1$ (this requires $L=(\sigma-\mu)/\mu$).[31]

The location inefficiency is larger the larger is the factor price and the difference in relative factor endowment differentials. The inefficiency gap n-n* is magnified by ϕ. However, it can be shown that the welfare loss is independent of ϕ, i.e. W(n)-W(n*) is equal to a constant. Since W(n*) is increasing in ϕ, this implies that the welfare loss is greater as a fraction of aggregate welfare when trade is more restricted.

This overall result hides a potential conflict between factor-owner groups. While there are no conflicts between different factor owners within a region, the preferred spatial allocation of industry for the inter-regional coalition of workers will differ from the preferred allocation of the inter-regional coalition of capital owners. By the same token as above,[32] it is straightforward to derive the preferred spatial allocation of firms for the two groups as n^L (the workers' bliss point) and n^K (the capital owners' bliss point), where:

(1-28) $$n^L = s_L + (2\phi)\frac{s_L - \frac{1}{2}}{1-\phi}, \quad n^K = s_K + (2\phi)\frac{s_K - \frac{1}{2}}{1-\phi}$$

which is similar to (1-19) and (1-26). Comparing the first expression in (1-28) with (1-19) and (1-26), we obtain:

Result 1-13. From the workers' (capital owners') point of view, both the market and the planner allocate too many (too few) firms to the capital abundant region.

This result is clearly more relevant in a situation where the two regions belong to the same nation, since it is easier to think of ways in which inter-regional interest groups can be affected when both regions are within the same political system.

---

[31] In Chapter 3, we will assume that the former special case holds so that departures from the optimum will be exclusively due to the political economy process.

[32] The workers' aggregate welfare is like (1-23) with zero weight to capital owners' welfare, and conversely for the capital owners' aggregate welfare.

1.8. Equity and efficiency in the FCVL model

The FC model is the simplest and most analytically tractable model of this chapter. As a result, its welfare analysis is quite simple. In this section, I consider the FCVL model, which is less tractable than the FC model but displays a wider range of features.[33] As for the FC model, I start with the Pareto welfare analysis and then turn to the global welfare analysis.

Pareto welfare analysis

The first thing to note is that nominal rewards are invariant in the FCVL model, exactly like in the FC model. This implies that all the welfare effects occur via the price index. Since increasing the share of industry n hurts capital owners and workers in region 2 alike and benefits capital owners and workers in region 1, there is no way we can Pareto improve upon the market outcome. In other words Result 1-8 holds in the FCVL model as well. Therefore each of the stable multiple equilibria is Pareto efficient.

Equilibria ranking

We can also rank the various equilibria according to each of region 1 and 2's residents welfare. The decentralised equilibrium delivers two types of stable long run equilibria: the dispersed equilibrium (in which case n=½) or a core-periphery outcome (n=0 or n=1). The first thing to note is that anyone's preferred outcome is to live in a region in which the whole of industry clusters. This is unambiguous, for e.g. Δ_1 is maximised at n=1 by Result 1-4. Thus we write:

> Result 1-14. In all cases, capital owners and workers of region 1 (region 2) alike are best off when all firms are clustered in region 1 (region 2).

Take for instance the residents in 1. Result 1-14 says that their preferred outcome is n=1 but is silent on which outcome is their second best: are those people better off if they are in the periphery (n=0) or if firms are evenly spread between the two regions (n=½)? As it turns out, this is not a trivial question.

---

[33] An even wider variety of results arise when we consider embodied factor mobility. See Baldwin et al. (2001) and Ottaviano (2001).

Take the FC model as a benchmark. The FCVL model collapses to the FC model with s=½ when α=0. As we know from section 1.3, α=0 implies that there are no forward or backward linkages –and hence no agglomeration economies. In such a case (1-4) reveals that Δ_1 is strictly increasing in n. This implies:

Result 1-15. When the magnitude of the vertical linkages is small ($\alpha \approx 0$), then residents in 1 rank the possible equilibria as follows: there are best off under the core-periphery pattern n=1; their second best is the dispersed equilibrium n=½; they are least well off under the core-periphery pattern n=0.

Now turn to the case $\alpha>0$. When α is large the magnitude of the agglomeration economies is large. As a result, when all firms cluster in a single region the production costs are very low because no intermediate inputs need be imported. Of course, this hurts people left behind in the periphery unless agglomeration economies are so strong that the cost of importing all varieties is compensated by lower mill prices.

To see this formally, assume without loss of generality that if firms are fully agglomerated then region 1 is the industrial cluster and region 2 the periphery. Then manipulating (1-4) reveals that in the dispersed outcome n=½ we have:

(1-29) $$\Delta_1 = \Delta_2 = \left(\frac{1+\phi}{2}\right)^{1/(1-\alpha)} < 1$$

By contrast Δ_1=1 and $\Delta_2=\phi$ if firms are clustered. Therefore residents in 1 always prefer the core-periphery pattern and residents in 2 do so if, and only if,

(1-30) $$2\phi^{1-\alpha} - (1+\phi)$$

is positive. This expression is negative for low values of ϕ and nil if ϕ=1 in which case location is irrelevant. However, it is easy to see that this expression is increasing at ϕ=0, decreasing at ϕ=1 if and only if α<½, and everywhere concave. As a consequence it has a unique maximum at which (1-30) is non-negative. If α>½ the expression in (1-30) is positive for any ϕ in (ϕ_P,1), where ϕ_P is the unique real root of this polynomial in (0,1). If $\alpha \leq$½ then (1-30) is negative for all admissible value of ϕ (in this case define ϕ_P as the corner solution ϕ_P=1). Hence, we have shown:

Result 1-16. If $\alpha>\frac{1}{2}$ then residents in the periphery are better off under the core-periphery outcome than under the dispersed outcome $n=\frac{1}{2}$ if, and only if, transportation costs are low enough. If $\alpha\leq\frac{1}{2}$ then residents in the periphery are worse off than under $n=\frac{1}{2}$ for all ϕ.

This result is quite intuitive. Agglomeration economies ensure that producer prices are lower in the core-periphery outcome than in the dispersed outcome. Low transportation costs ensure that consumer prices are also lower in the former configuration than in the latter configuration, even for the consumers who have to import the manufacturing goods.

Moreover, it is easy to see that this effect is stronger, the larger the agglomeration economies. Indeed, at the limit $\alpha=1$, $\Delta_2\to\Delta_1=1$, so consumer prices in the periphery are the same as in the core, for all any value of ϕ. More generally,

(1-31) $$\frac{\partial}{\partial\alpha}\left(2\phi^{1-\alpha}-(1+\phi)\right)=-2\phi^{1-\alpha}\ln(\phi)>0$$

which implies that ϕ_P is decreasing in α. In other words,

Result 1-17. The range of ϕ over which everybody benefits from the clustering of industry in either region vis-à-vis the dispersed equilibrium is increasing in the magnitude of agglomeration economies.

This too is rather intuitive.

We can finally address the question of whether the market provides too much or too little agglomeration from the periphery's resident's point of view. To answer this question we rank ϕ_P, ϕ^{break}, and ϕ^{sust}.

Plug (1-15) into (1-16) and (1-30); numerical simulations reveal that the former resulting expression is positive hence $\phi^{sust}<\phi^{break}$.[34] Also, the sign of the latter depends upon the magnitude of α and σ. As is to be expected, the condition (1-30) is less likely to be violated at the break point when the magnitude of the vertical linkages α is large. Also, numerical comparisons show that $\phi^{sust}<\phi_P$ holds for all parameter values.

---

[34] This fact is formally demonstrated in Chapter 2. This implies that for all $\phi\in(\phi^{sust},\phi^{break})$ both the core-periphery and the dispersed outcomes are stable long run equilibria.

Together, these facts imply first, $\phi^{sust}<\phi^{break}$, $\phi^{sust}<\phi_P$ and second, the sign of $\phi_P-\phi^{break}$ is ambiguous. As a consequence, five cases can occur. From the point of view of region 2's residents:

1. If $\phi<\phi^{sust}$ the market outcome (n=½) is the second best;
2. If $\phi>\phi_P,\phi^{break}$ the market outcome (n=1) is the second best;
3. If $\phi_P>\phi>\phi^{break}$ the market outcome (n=1) provides too much agglomeration;
4. If $\phi_P<\phi<\phi^{break}$ the market outcome is the second best if n=1but provides too little agglomeration if n=½;
5. If $\phi^{sust}<\phi<\phi_P$ the market outcome is the second best if n=1but provides too little agglomeration if n=½.

By contrast, anyone's welfare is maximised when firms cluster in one's region. Hence, from the point of view of region 1's residents, three cases can occur:

1. If $\phi>\phi^{break}$ then the decentralised outcome (n=1) delivers the first best;
2. If $\phi<\phi^{sust}$ then the decentralised outcome (n=½) delivers too little agglomeration;
3. If $\phi^{sust}<\phi<\phi^{break}$ then the decentralised outcome delivers the first best if n=1but provides too little agglomeration if n=½.

This concludes our Pareto analysis. Turn now to the global welfare analysis.

Global welfare analysis

Here the question is, can the planner improve upon the decentralized equilibrium? I invoke Result 1-9 (first best and second best outcomes coincide) to focus only on the second-best analysis. Namely the planner will choose n so as to maximize a utilitarian welfare function, not being able to correct for above marginal cost pricing. This assumes that the planner can only correct for the inefficiencies that arise as the result of the spatial allocation of firms.

In the FCVL model we assume that the two regions are equally endowed with labour and capital. Hence, the planner maximizes (1-23) with $s_L=s_K=½$, which is equivalent to maximising

(1-32) $\Omega = \ln(\Delta_1) + \ln(\Delta_2)$

where the Δ's are given by (1-4).

As before we take the FC model as a benchmark. The FCVL model collapses to the FC model with s=½ when α=0. As we know from section 1.3, α=0 implies that there are no forward or backward linkages –and hence no agglomeration economies. In such a case, (1-32) is concave for all n; this implies that the solution to the utilitarian planner's problem is n*=½. In words when there are no vertical linkages the planner chooses to spread industry evenly between two symmetric regions. In the light of Result 1-11, the market and the social optimum coincide because both population and expenditure are spread the same way.

Turn now to the case α>0. By a continuity argument, it must be that Ω is concave in n in the neighbourhood of n=½, at least for a small α. When α increases, however, agglomeration economies start to be of important magnitude and hence the planner may be willing to cluster all industries in a single region so as to lower the production costs on all varieties. By Result 1-16 this may hurt people left behind in the periphery.

To see this formally, note that (1-29) and (1-32) imply that Ω is larger when n=1 than when n=½ if, and only if,

(1-33) $4\phi^{1-\alpha} - (1+\phi)^2$

is positive. This expression is negative for low values of ϕ and nil when ϕ=1 in which case location is irrelevant. However, it is easy to see that , like the left-hand side of (1-16), this expression is increasing at ϕ=0, decreasing at ϕ=1, and everywhere concave. As a consequence it has a unique maximum at which it is positive. Therefore the expression in (1-33) is positive for any ϕ in (ϕ_{USP},1), where ϕ_{USP} is the unique real root of this polynomial in (0,1). The subscript 'USP' stands for utilitarian social planner. Moreover, by the same token as for the analysis of (1-30) and (1-31) it is easy to show that $\partial\phi_{USP}/\partial\alpha<0$. Hence, we have proved the following result:

> Result 1-18. The social planner prefers the core-periphery pattern to the dispersed equilibrium if, and only if, transportation costs are low enough. Moreover, the range of ϕ over which the planner prefers the core-

periphery pattern is increasing in the magnitude of the vertical linkages as parameterised by α.

This result is quite intuitive in light of Result 1-15, Result 1-16, and Result 1-17.

We can also finally address the question of whether the market provides too much or too little agglomeration from a utilitarian point of view. To answer this question, we rank ϕ_{USP}, ϕ^{break}, and ϕ^{sust}. Plug (1-15) into (1-16) and (1-33); numerical simulations reveal that the former resulting expression is positive and that the latter is negative. Together, these imply $\phi^{sust}<\phi_{USP}<\phi^{break}$. Four cases can occur:

1. If $\phi<\phi^{sust}$ the market outcome (n=½) is socially optimal;
2. If $\phi>\phi^{break}$ the market outcome (n=0 or n=1) is optimal;
3. If $\phi^{sust}<\phi<\phi_{USP}$ the market outcome is optimal if n=½ but provides too much agglomeration if n=0 or n=1;
4. If $\phi_{USP}<\phi<\phi^{break}$ the market outcome is optimal if n=0 or n=1but provides too little agglomeration if n=½.

This concludes our welfare analysis.

1.9. Concluding remarks

In this chapter I have developed an alternative NEG model in which agglomeration is driven by the interaction of increasing returns at the firm level, trade/transportation costs, and vertical linkages among firms –that is, firms use each other's output as an input. I have sketch the ways the present model behaves in a very similar fashion to already well-established economic geography models. In particular, it shares the features of the original model developed by Venables (1994, 1996a). However, it is more tractable and hence allows for easier extensions and less reliance on simulations. I have also shown that this models nests the 'footloose capital' model of Flam and Helpman (1987) and Martin and Rogers (1995).

The models of vertical linkages-driven agglomeration of Venables (1996a), Krugman and Venables (1995), and Fujita et al. (1999) are intractable for several reasons. First, the price indices are defined implicitly and simultaneously, as here. In addition, there is only one primary factor, which means that w and π are the same thing; since prices are functions of wages, the q's in (1-6) enter the price indices definitions of

(1-4) as well, adding one more degree of simultaneity –and complexity. As a consequence extensions of the basic model that allow for asymmetries become extremely tedious and one has to rely on simulations to derive the predictions of the model. The model above is simpler and hence extensions along the same lines are more tractable. A good example of this is the paper by Puga and Venables (1997) in which the authors study preferential trading agreements in a NEG framework using the CPVL model. Conveying the same kind of exercise with the model developed in this chapter provides more analytical solutions.[35]

The model of this chapter will be used in the applications of Chapters **4** and **5** to two specific issues. Chapter 2 completely characterizes the set of equilibria for a whole class of NEG models, including the FCVL model of Section 1.4.

Appendix

In this appendix I rewrite the model in (1-4)-(1-6) in the so-called natural state space of Chapter 2. Note first by that $nq_1+(1-n)q_2=1$ for all n, in any instantaneous equilibrium (this holds by the definitions of π and the q's). Consequently define $\eta \equiv nq_1/[(1-n)q_2]$. Also, define the following ratios: $q \equiv q_1/q_2$, $\Delta \equiv \Delta_1/\Delta_2$, and $e \equiv e_1/e_2$. Finally, it proves useful to denote the ratio of expenditures in the special case all firms cluster in a single region by χ; to get a solution for χ, substitute n=0 in (1-5) in the definition of e to get $\chi=(1-\alpha\beta)/(1+\alpha\beta)$. With all these definitions, the instantaneous equilibrium can be written as:

(1-34) $$e=\frac{\eta+\chi}{1+\eta\chi}; \quad \eta\Delta^{\alpha}=q\frac{\Delta-\phi}{1-\Delta\phi}; \quad q=\Delta^{\alpha}\frac{Y+\Delta\phi}{Y\phi+\Delta}$$

As we shall see in Chapter 2, this system is much simpler to deal with when it comes to show that it has no more than three long run equilibria (defined as (1-34) with q=1).

[35] For an application see Baldwin et al. (2001).

Chapter 2. THE STRUCTURE OF SIMPLE 'NEW ECONOMIC GEOGRAPHY' MODELS

2.1. Introduction

The beauty of the neoclassical trade theory stems for a good part on its ability to provide strong results relying on few assumptions and robust to the choice of functional form; see for instance the two-sector, two factor model of Jones (1965). By contrast, the so-called 'New Economic Geography' (NEG hereafter), initiated by Krugman's (1991a,b) seminal contributions, seems to rely on very specific functional forms and yet to be highly intractable. NEG models come in basically three categories:[36]

- Migration-based models
- Models based on vertical linkages
- Models based on factor accumulation

In models based on factor migration, agglomeration is a result of the interaction of the 'Home-Market effect' (Krugman 1980) and factor spatial mobility. Krugman's (1991b) famous Core-Periphery (CP hereafter) model assumes that the factor used intensively in the imperfectly competitive sector is mobile between regions and moves according to regional differences in real wages. In models based on vertical-linkages among firms, like the FCVL model of Chapter 1, agglomeration comes as a result of the interaction of the home-market effect and input-output linkages among firms. Finally, agglomeration can also result as the interaction between the home-market effect and the accumulation of some factors, as in Baldwin (1999).

Beyond assuming different agglomeration mechanisms, these contributions use different functional forms (albeit the differences are small). Yet, these models are surprisingly isomorphic when one rewrites them in a more natural state-cum-parameter space, as I shall demonstrate in this chapter. Consequently, the predictions of the NEG models are robust to important changes in the particular agglomeration mechanisms.

---

[36] This topology is not hermetic, as e.g. Puga's (1999) model combine ingredients of several of these categories.

This implies that, very much like the neoclassical trade model, the NEG paradigm is well suited to be brought to the data and its theoretical predictions to be tested.

Another ambition of this chapter is to characterize fully the set of steady states of a wide class of NEG models. To be more precise, the aim here is to fully characterize the set of equilibria of the symmetric two-region models of Table 1-1.[37] As strange as it may seem, this is yet an unfinished business. This is odd because the multiplicity of equilibria is one of the central features of the NEG paradigm. Numerical simulations of these models consistently display the following features:

- There are no more than five steady-states (and no more than three interior steady-states);
- Among them the symmetric one always exists,[38] but is not always stable;
- When they exist, the two asymmetric interior steady states are always unstable.[39]

In Krugman's terminology, these properties are illustrated by the famous 'tomahawk bifurcation diagram' of Figure 2-2.

Importantly, no formal proof of these results has ever been provided despite many attempts. However, many important statements of practical importance made in the literature –hypes, in the words of Neary (2001)– rely on these features.[40] The present chapter provides such proofs.

In short, the purpose of this chapter is twofold. For one, it fills an important gap in providing analytical proofs to some key results in this strand. On top of that, it contributes to a deeper understanding of the core mechanics of NEG models by showing that many such models are isomorphic even though they assume different agglomeration mechanisms; choose some different functional forms, or both.

---

[37] Two reservations are in order here. First, the full characterization of the CPVL model is a bit more involved for reasons that will be clear later. Hence, the conclusions reached in this chapter are more tentative than conclusive in this case. Second, the analytical method developed here applies to the models of Table 1 strictly speaking. It would need to be extended, possibly at prohibitive cost in term of effort, to the various extensions of these models (like the extension of the FCVL model conveyed in Chapter 4).
[38] This is obvious given the symmetry of the model.
[39] That is, the number of equilibria is odd. Of course, this is not particular to this model. Indeed, it is a very general result: for instance, see the Index Theorem in Mas-Colell, Whinston and Green (1995, p. 593).
[40] Think, for instance, of Krugman and Venables's (1995) 'History of the world, Part I' in which they conjecture that the very same process that has brought huge income inequalities among countries, globalisation, might yield income convergence in the future. See also Chapter 4 on this issue.

2.2. A natural state space for migration-based models

Our main task in this section is to set up the natural state space for the migration-based CP and 'Footloose-Entrepreneur' models alike. The Footloose-Entrepreneur (FE hereafter) model from Forslid and Ottaviano (2001) is very similar to the CP model, but an apparently small modification in the functional form of a cost function makes the model more tractable whilst retaining all of its qualitative features.

Writing these models in their natural state space is a necessary intermediate step for the central proofs of Section 2.3. Here I only sketch the structure of the models, for they are very similar to the model of Chapter 1. I am using the notation of Fujita, Krugman, and Venables (1999) so as to facilitate the comparison between the CP model as laid down in their chapter 5 and the FE model.

The common structure

Common to virtually all NEG models are two regions (indexed by j=1,2) and two sectors. The background sector A produces a homogenous good under constant returns to scale (CRS) in a competitive environment using unskilled workers only; its output is freely traded; consumers spend a share $1-\mu$ of their expenditure on A. The manufacturing sector M produces a differentiated product under increasing returns to scale (IRS) in a monopolistically competitive environment a-la Dixit and Stiglitz (1977). Denote the elasticity of substitution between any two varieties by $\sigma>1$. The mathematical expression describing tastes is given by (1-1). Shipping this good into the other region involves 'iceberg' transportation costs: T>1 units need to be shipped so that 1 units arrives at destination; the rest, T>1, melts in transit. Consumers spend a share $0<\mu<1$ of their expenditure on the composite M. Further, in models based on factor migration, both regions are equally endowed with $L_U/2$ unskilled workers. These two regions also share L skilled workers, with λ $(1-\lambda)$ of them living in region 1 (region 2); λ is the variable of interest; it is endogenous in the 'long run'.

The Core-Periphery model

In the CP model, each factor is specific to a different sector: the background (manufacturing) sector uses unskilled (skilled) workers only. In particular, the cost function of the typical M-firm takes the following form: $C(x)=(F+\rho x)w$, where x is the firm output, F is the fixed labour requirement, ρ is the variable labour requirement, and w is the (skilled) labour wage; when necessary, w will be indexed by a subscript 1 or 2 so as to distinguish between regions. Tastes are as specified in Chapter 1.

As usual, monopolistic pricing implies that the producer price p of any variety involves a mark-up, viz. $p(1-1/\sigma)=\rho w$. Free-entry eliminates pure profits so $px-C(x)=0$, which implies $\rho x=(\sigma-1)F$ irrespective of the wage. This has two consequences. First, the equilibrium firm size is constant, viz. $x=(\sigma-1)F/\rho$. Second, skilled worker wages have a dual nature here. On the one hand, these wages are a cost –in particular, higher wages translate into higher variable costs and hence higher prices. On the other hand, w captures the monopoly rents. Indeed, pure profits are eliminated at equilibrium, hence the fixed factor –skilled labour again– captures the rents that accrue from monopoly pricing. This dual nature of skilled labour is the cause of much of the analytical intractability of the CP model.

Normalizations

Following FKV, let us make the following normalizations. First, we take A as the numéraire. Second, we choose units so that (i) $F=\mu/\rho$; (ii) $L=\mu$ and $L_U=1-\mu$; (iii) the labour-output requirement in A is 1, which implies $w_U=1$ in each region; (iv) $\rho=1-1\sigma$ so that the M-sector f.o.b. prices are all equal to unity.

Instantaneous equilibrium

With all these assumptions at hand, we can solve the model treating λ as a parameter to get the so-called 'instantaneous equilibrium'. Such an equilibrium is defined for any value of λ as a situation where all markets clear and trade is balanced. Using subscripts to indicate regions, the instantaneous equilibrium of the CP model is characterized by the following equations (see FKV, p. 65):

(2-1)
$$\begin{aligned}
&Y_1=\mu\lambda w_1+\frac{1-\mu}{2}, \quad Y_2=\mu(1-\lambda)w_2+\frac{1-\mu}{2}\\
&G_1^{1-\sigma}=\lambda w_1{}^{1-\sigma}+(1-\lambda)(Tw_2)^{1-\sigma}, \quad G_2^{1-\sigma}=\lambda(Tw_1)^{1-\sigma}+(1-\lambda)w_2{}^{1-\sigma}\\
&w_1^{\sigma}=\frac{Y_1}{G_1^{1-\sigma}}+T^{1-\sigma}\frac{Y_2}{G_2^{1-\sigma}}, \quad w_2^{\sigma}=T^{1-\sigma}\frac{Y_1}{G_1^{1-\sigma}}+\frac{Y_2}{G_2^{1-\sigma}}
\end{aligned}$$

The economic signification of (2-1) is as follows. Income in region j, Y_j, is defined as the sum of skilled workers' income (whose wage is w_j) and unskilled workers' income (whose wage is unity by choice of numéraire). G_1 and G_2 are the true CES price indices.

The third line provides the free-entry-and-exit conditions (also called the 'wage equations'): firms in region *j* break even if they pay the wage w_j.

We claim that the following holds for all λ:

(2-2) $$\lambda w_1 + (1-\lambda) w_2 = 1$$

To see this, multiply the wage equations in (2-1) by $w_j^{1-\sigma}$ and the result follows readily by using the income and price index equations. This property stems from the Cobb-Douglas functional form of the upper-tier utility function.

Define the real (skilled labour) wage in region j as ω_j, viz. $\omega_j \equiv \omega_j G_j^{-\mu}$. Hence, the system in (2-1) can be viewed as providing an implicit solution for $\omega_1 - \omega_2$ as a function of λ; accordingly, we write (2-1) as $F(\omega_1 - \omega_2) = 0$ for short. Skilled labour is assumed to move from one region to the other so as to eliminate differences in real wages according to the following law of motion:

(2-3) $$\dot{\lambda} = \gamma \lambda (1-\lambda)(\omega_1 - \omega_2)$$

where $\gamma > 0$ is a parameter. Clearly, this means that expectations are static (see Chapter 1).[41] A long run (or steady-state) equilibrium is defined as a value of λ that solves (2-1) and such that (2-3) holds for $\dot{\lambda} = 0$. We are interested in the number of *interior* solutions, i.e. the number of λ's such that $\omega_1 = \omega_2$. Call the set of these λ's L_0, with λ_0 being the typical element of L_0. That is,

(2-4) $$L_0 \equiv \{\lambda \in [0,1] : F(0,\lambda) = 0\}$$

Using (2-1), it is straightforward to show two things. First, $\lambda = \frac{1}{2} \in L_0$ is always true. Also, $\lambda_0 \in L_0$ if, and only if, $1-\lambda_0 \in L_0$. These two facts hold by virtue of the symmetry of the model.

Define the parameter ϕ as $\phi = T^{1-\sigma} \in [0,1)$; ϕ is decreasing in T. We now move on by using the definition of ω_j (j=1,2) and plugging it into (2-1):

---

[41] Baldwin (2001), however, shows that this is merely an assumption of convenience. Indeed, allowing for rational expectations and sufficiently large quadratic migration costs does not alter the 'break' and 'sustain' points defined in section 2.3 below. Yet, expectations can be self-fulfilling when migration costs are low enough, as in Matsuyama (1991) or Krugman (1991b).

(2-5)
$$Y_1 = \mu\lambda w_1 + \frac{1-\mu}{2}, \quad Y_2 = \mu(1-\lambda)w_2 + \frac{1-\mu}{2}$$
$$G_1^{1-\sigma} = \lambda w_1(\omega_1 G_1^{\mu})^{-\sigma} + \phi(1-\lambda)w_2(\omega_2 G_2^{\mu})^{-\sigma},$$
$$G_2^{1-\sigma} = \phi\lambda w_1(\omega_1 G_1^{\mu})^{-\sigma} + (1-\lambda)w_2(\omega_2 G_2^{\mu})^{-\sigma}$$
$$(\omega_1 G_1^{\mu})^{\sigma} = \frac{Y_1}{G_1^{1-\sigma}} + \phi\frac{Y_2}{G_2^{1-\sigma}}, \quad (\omega_2 G_2^{\mu})^{\sigma} = \phi\frac{Y_1}{G_1^{1-\sigma}} + \frac{Y_2}{G_2^{1-\sigma}}$$

We can transform (2-5) further to simplify both the resolution of the problem and the algebra. First, by virtue of (2-2) we can define the parameter η as the ratio of the mobile factor's expenditure in 1 to mobile expenditure in 2, namely $\eta \equiv \lambda w_1/(1-\lambda)w_2$. Next, define θ as the following collection of parameters $\theta=\mu/\rho$ (as we shall see θ can be interpreted as measure of the forward linkage). Third, define the analogue of ϕ for the price index, $\Delta_j \equiv Gj^{1-\sigma}$, with $d\Delta_j/dG_j<0$.

Finally, the entire problem can be rewritten in terms of ratios because the two regions are the mirror image of one another (in this two-region world, only the relative size of any variable matter). Accordingly, define Δ, Y and ω as $\Delta\equiv\Delta_1/\Delta_2$, $Y\equiv Y_1/Y_2$, and $\omega\equiv\omega_1/\omega_2$. When all mobile workers settle in a single region (viz. $\lambda\in\{0,1\}$), this region is dubbed as 'the core' and its GDP is equal to $L+L_U/2$ by (2-2). Conversely, the GDP in the 'periphery' is equal to $L_U/2$. Let us then define χ as the ratio of these two, viz.

(2-6)
$$\chi = \frac{L_U}{2L + L_U} < 1$$

Using these ratios as well as the definitions of η and Δ_j into (2-5), we obtain the following system:

(2-7)
$$Y = \frac{\eta + \chi}{1 + \eta\chi}; \quad \eta\Delta^{\theta} = q\frac{\Delta - \phi}{1 - \Delta\phi}; \quad q = \Delta^{\theta}\frac{Y + \Delta\phi}{Y\phi + \Delta}$$

where the 'natural' benchmark q is defined as ω^{σ} (the model reaches an interior steady state whenever q=1). The model in (1-34) is identical to the model in (2-1), as I will show shortly. The endogenous variables Y, Δ, and q are functions of the 'natural' parameters ϕ, θ, and χ as well as the state variable η. Following Fujita et al. (1999) I impose the 'no-black-hole' condition $\theta<1$ (FKV, p. 59). Without this the symmetric equilibrium is never stable. If $\theta\geq1$ then both the break and sustain points defined below are equal to zero.

Backward linkage, forward linkage, and market crowding

As it turns out the expressions in (1-34) illustrate the agglomeration and dispersion forces of the model neatly. These forces are described at length in Chapter 1 so I will be brief here.

Start with the backward linkage, an agglomeration force that stems from the fact that some expenditure is mobile. It is obvious from the first expression in (1-34) that incomes are increasing in the local share of skilled workers' expenditure. In other words Y_1 is increasing in λw_1.[42] This in turn is transmitted into the relative real wages via $\partial q/\partial Y>0$ in the last expression of (1-34). Obviously firms care about the value of demand so λw_1 and $(1-\lambda)w_2$ are the proper yardsticks to asses relative demands when tastes are homothetic, as here. All things equal, an inflow of mobile expenditure in 1 increases demand for local firms' output; this, in turn, increases the wage at which firms can break even, attracting further mobile workers, and the cycle repeats. Also χ is inversely related to the magnitude of the backward (or demand) linkage: when no expenditure is mobile then $\chi=1$, in which case Y is invariant in η.

The second expression in (1-34) provides an implicit definition for Δ and illustrates the nature of the 'forward linkage' in this model. To boost intuition, assume real wages are equalized, so that q=1. It can be shown fairly easily that $1>\theta\geq 0$ implies $\partial\Delta/\partial\eta>0$, namely, the price index in 1 falls relatively to the price index in 2 when the share of mobile expenditure in 1 relative to 2 increases. This is the source of the forward linkage: the more firms there are in 1, the cheaper it is for workers to live there, ceteris paribus, so this induces more workers to settle in 1, and the cycle repeats. Interestingly Baldwin's (1999) model displays no forward linkage so $\theta=0$ in his model..

The last expression in (1-34) also captures the 'market crowding' dispersion force. To identify this dispersion force switch off the forward and backward linkages, namely impose $\theta=0$ and $\chi=1$. This restriction is sufficient to ensure that $\partial q/\partial\Delta<0$. What this says is the following: Δ is also a measure of the degree of competition in market 1 relative to market since Δ is larger than 1 if there are more firms in 1 than in 2, ceteris paribus. Since Δ is itself increasing in η, this implies that as mobile workers move from

---

[42] As we shall see, Y_1 it may or may not increase in λ (see Result 2-1 below on this).

2 to 1, the increased competition in market 1 has a negative impact on manufacturing wages, acting as dispersion force.

Write now (1-34) in a compact form as f(q,η)=0. With this notation, an interior steady state is defined as a value for η such that f(1,η)=0. By analogy with (2-3), define N_0 as:

(2-8) $$N_0 \equiv \{\eta \geq 0 : f(1,\eta) = 0\}$$

We denote the typical element of N_0 by η_0. Remark that it is readily verified using (1-34) that η=1 is always an element of N_0. We can infer the exact number of interior-steady states $\#L_0$ of the problem (2-3) from $\#N_0$ if the two are related, e.g. if $\#N_0=\#L_0$. Result 2-2 below shows that there indeed exists a one-to-one mapping from N_0 to L_0. Before turning to this important result, I need the following lemmas –the proofs of which, as well as the proof of other intermediate results, are relegated to an appendix at the end of the chapter.

Lemma 1. Define the function $\mathfrak{M} : [0,1] \to \mathbb{R}_+$ where $\eta = \mathfrak{M}(\lambda)$. Then $\mathfrak{M}$ is surjection (a.k.a. onto).

Proof. See appendix.

Lemma 2. $\mathfrak{M}'(\cdot) > 0$ at the symmetric steady state λ=½.

Proof. See appendix.

Result 2-1 below says that expenditure of skilled workers in, say, region 1, increases with the share of such workers there, even taking the (potentially) depressing effect of λ on w_1. Note that this is not trivial as the mapping from w_1 to λ is not one-to-one (not an injection).[43]

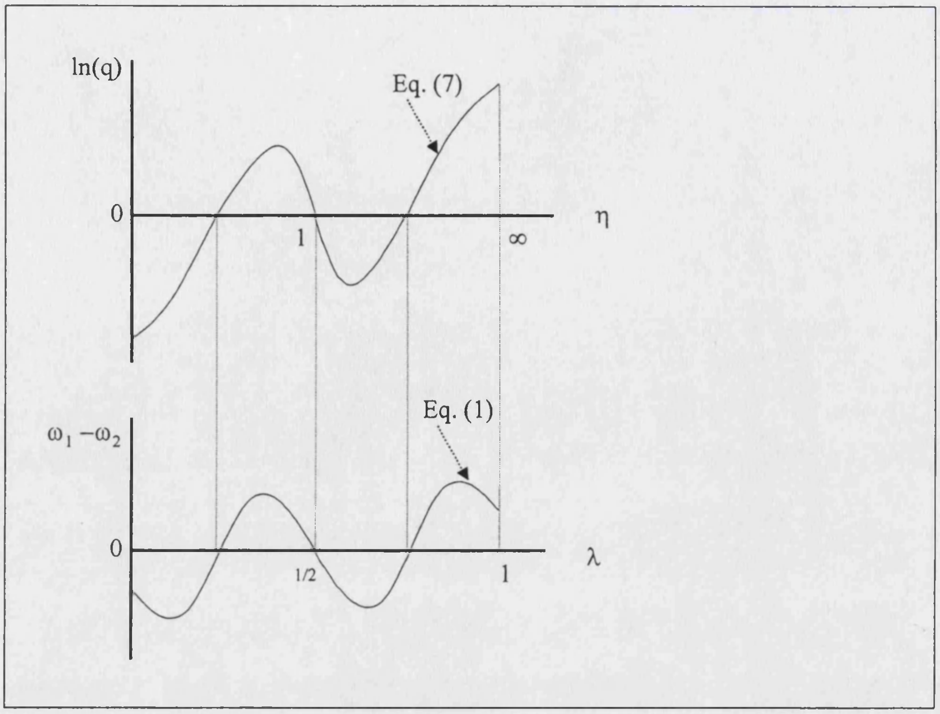
Result 2-1. (**a**) $\mathfrak{M}$ is a bijection; (**b**) η is increasing in λ.

Proof. See appendix.

---

[43] This claim can be made rigorously fairly easily. Since this is nowhere needed for our purposes, however, we omit the proof.

Figure 2-1: Instantaneous Equilibrium



Turn to Figure 2-1. This figure plots ln(q) as a function of η in the upper panel, viz. (1-34), and ω_1-ω_2 as a function of λ in the lower panel, viz. (2-1). Result 2-1 above says that to each λ on the horizontal axis in the lower panel corresponds exactly one η on the horizontal axis of the upper panel. The figure also prefigures Result 2-2 which shows that the number of solutions for λ to the problem (2-4) is identical to the number of solutions to the problem (2-8), viz. $\#N_0=\#L_0$. Result 2-2 is essentially a corollary of Result 2-1. It is fundamental because it is much easier to characterize q in (1-34) than ω_1-ω_2 in (2-1).

Result 2-2. Let $\mathfrak{M}_0 : L_0 \to N_0$ be the mapping such that $\eta_0 = \mathfrak{M}_0(\lambda_0)$. Then $\mathfrak{M}_0$ is a bijection, implying $\#N_0=\#L_0$.

Proof. $\mathfrak{M}_0$ is identical to $\mathfrak{M}$, except for its range and domain that are subsets of those of $\mathfrak{M}$. Therefore, to each solution for η in (2-8) corresponds exactly one value for λ (and reciprocally). This must obviously correspond to a solution in (2-4). Indeed, $\lambda_0 \in L_0$ implies $\omega_1=\omega_2$, or $q=(\omega_1/\omega_2)^{\sigma}=1$ so $\mathfrak{M}(\lambda_0) \in N_0$. A parallel argument shows that for all η_0 in N_0 there exists $\lambda_0 \in L_0$ such that $\lambda_0 \in \mathfrak{M}^{-1}(\eta_0) = \mathfrak{M}_0^{-1}(\eta_0)$; by Result 2-1 this λ_0 is unique. *QED.*

Hence it remains to show that the curve q-1 plot against η crosses the horizontal axis at most thrice, a case Figure 1 illustrates. A sufficient condition for this to be true is that the curve q-1 admits at most two flat points when plotted against λ, or η. It turns out that the simplest way of showing this is to invoke the alternative model of migration-driven agglomeration, the FE model.

The Footloose-Entrepreneur model

In the FE model, skilled workers are specific to the M-sector, but unskilled workers are employed in both sectors; specifically, the cost function takes the same functional form as in Chapter 1, namely $C(x)=Fw+\rho x w_U$, where w_U is the unskilled labour wage rate and is equal to unity by our choices of units and normalizations.

As usual, monopolistic pricing implies that the producer price p of any variety involves a mark-up, viz. $p(1-1/\sigma)=\rho w_U$. Free-entry eliminates pure profits so $px-C(x)=0$. Using the fact $w_U=1$ this implies $\rho x=(\sigma-1)Fw$. This has two consequences. First, this gives us an expression in both x and w, so we need one more expression to be able to solve for x and w, hence x and w are determined simultaneously. Remember that in the CP model, the size of firms is fixed and the variable of adjustment is the number of firms; in the FE model, exactly the opposite is true, as we shall see.

Second, unskilled workers' wage enter the variable cost only. By contrast, skilled worker captures the monopoly rents; in effect, w is the operating profit that accrues from monopoly pricing.[44] For this reason, we can see the skilled workers as 'entrepreneurs'. By contrast, in the CP model skilled workers held a dual role: they were workers and entrepreneurs at the same time. This somewhat innocuous change makes the FE model much more tractable than the CP model.

---

[44] In the FCVL and FC models 'capital' plays the role of 'skilled labour' here. We are using the former terminology when the specific factor is disembodied, as in Chapters 1, 3, and 4. We use the latter terminology when it is embodied, as here.

Instantaneous equilibrium

We keep all the normalizations of the CP model, but one: we take L=μ/σ and L_U=(σ-μ)/σ; this way (2-2) holds as before. The instantaneous equations can then be written as follows:

$$\textbf{(2-9)}\qquad \begin{aligned} &Y_1=\frac{\mu\lambda w_1}{\sigma}+\frac{\sigma-\mu}{2\sigma},\quad Y_2=\frac{\mu(1-\lambda)w_2}{\sigma}+\frac{\sigma-\mu}{2\sigma}\\ &G_1^{1-\sigma}=\lambda+(1-\lambda)\phi,\quad G_2^{1-\sigma}=\lambda\phi+(1-\lambda)\\ &w_1=\frac{Y_1}{G_1^{1-\sigma}}+\phi\frac{Y_2}{G_2^{1-\sigma}},\quad w_2=\phi\frac{Y_1}{G_1^{1-\sigma}}+\frac{Y_2}{G_2^{1-\sigma}}\end{aligned}$$

This is a simpler model in that the price-indices now depends on λ and ϕ only –in particular, they depend on no 'short-run' endogenous variable; this way explicit solutions for ω_1 and ω_2 are available.

Since (2-2) holds in the FE model as well, we can define η as before. Together with the definition of ω_j, this implies $\omega_1/\omega_2=(G_1/G_2)^{-\mu}\eta(1-\lambda)/\lambda$. We solve for λ/(1-λ) and plug the result into (2-9); we also use the definition of Δ_j and the ratio notation to get:

$$\textbf{(2-10)}\qquad Y=\frac{\eta+\chi}{1+\eta\chi};\quad \eta\Delta^{\theta}=q\frac{\Delta-\phi}{1-\Delta\phi};\quad q=\Delta^{\theta}\frac{Y+\Delta\phi}{Y\phi+\Delta}$$

In this expression q is now defined as ω and θ as μ/(σ-1); χ is still defined by (2-6) and the no-black-hole condition θ<1 is unchanged. Clearly, (2-10) and (1-34) are identical, hence the CP and FE models are <u>isomorphic</u>. Consequently, we can use Figure 2-1 for the FE model, too, in which case the bottom and upper panels correspond to (2-9) and (2-10), respectively. We define the set of the interior long run solutions to (2-10) for η and λ as N_0^{FE} and L_0^{FE}, respectively:

$$\textbf{(2-11)}\qquad N_0^{FE}\equiv\{\eta>0:f(1,\eta)=0\},\quad L_0^{FE}\equiv\{\lambda>0:F(0,\lambda)=0\}$$

these are the equivalent of N_0 in (2-8) and L_0 in (2-4).

The intuition for the proof of the main result in this chapter (Result 2-4 below) goes as follows. Since (2-10) is identical to (1-34), the sets N_0 and N_0^{FE} are identical as well, though generically the sets L_0 and L_0^{FE} are not. (In particular, asymmetric steady states differ). What is important, however, is that the number of solutions to the two models is identical. A final intermediate result is needed before we can turn to the proof proper:

Result 2-3. The FE model admits at most 3 interior steady states, viz. $\#L_0^{FE} \leq 3$.

Proof. See appendix.

We have at last everything at hand to prove the main result of this chapter, namely that $\#L_0$ is no larger than 3 for the CP model as well.

2.3. On the number of steady-states and stability analysis

In this section we show three results. First, the CP model admits at most three interior steady states. Second, asymmetric interior steady states are always unstable. Finally, location displays hysteresis in a well-defined sense. We turn to the issue of determining the generic number of steady states first.

On the number of steady-states

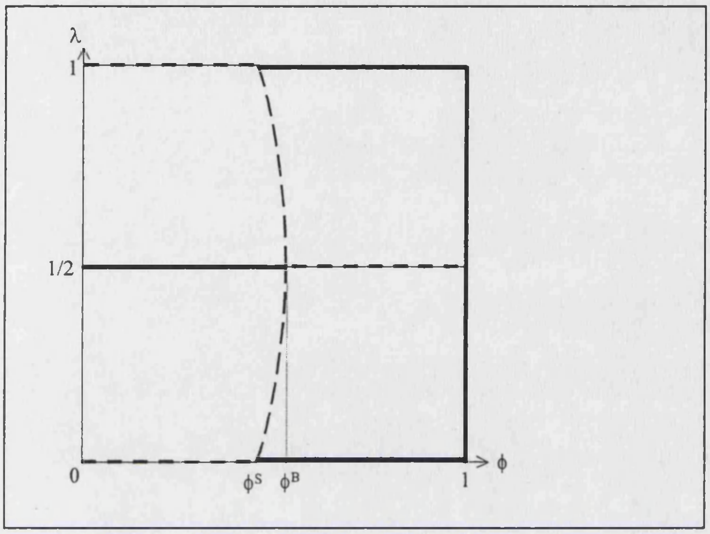
Result 2-4. The CP model admits at most 3 interior steady states.

Proof. Start with the FE model. Using Result 2-2 and Result 2-3 we have $\#N_0^{FE} \leq 3$. Since (2-10) is identical to (1-34), it must be that $N_0 = N_0^{FE}$. Using Result 2-2 again we find $\#L_0 = \#N_0 = \#N_0^{FE} = \#L_0^{FE}$. This implies $\#L_0 \leq 3$ and establishes the result. *QED.*

Figure 2-2 plots the famous 'tomahawk bifurcation diagram'. It plots ϕ on the horizontal axis against λ and shows the stable steady states in plain lines and the unstable ones in dotted lines. We infer from the figure and the analysis in Chapter 1 that when trade/transportation costs are high ($\phi < \phi^{sust}$) the only stable steady state is the symmetric one. In particular, the core-periphery structure is said not to be 'sustainable' (hence the name of the threshold ϕ^{sust}). When trade integration is deep enough ($\phi > \phi^{break}$) the only stable configuration is the core-periphery structure; the stability of the symmetric steady state is said to be 'broken' (hence the name of the threshold ϕ^{break}). For intermediate values of ϕ both the dispersed and the core-periphery outcomes are stable;

there are also two interior, asymmetric steady states. These are always unstable because $\phi^{sust}<\phi^{break}$ in this model (see Result 2-5 below). This form of multiplicity of equilibria also implies that location is path-dependent.[45] In order to establish that the bifurcation diagram has the shape drawn in Figure 2, we need to rank the 'break' and 'sustain' points, as defined in the sequel. We turn to this issue next. As an aside, it should be noted that doing so using the 'natural' state-space, viz. (1-34) or (2-10), is much simpler than it is using (2-1) or (2-9).

Figure 2-2: The Tomahawk Diagram



Hysteresis

As is well known, two local stability tests are usually applied in the NEG literature.[46] A first question we may ask is, 'is the core-periphery structure *sustainable*?' To answer this question we solve for the lowest value of ϕ in [0,1] such that it is just profitable to produce in the periphery. That is, we imposeη=0 (or $\eta=\infty$) and q=1 in (1-34) or (2-10) and solve for ϕ; this gives us the following implicit definition of the 'sustain point' ϕ^{sust}:

(2-12) $$(1+\chi)(\phi^{sust})^{1-\theta}-(\phi^{sust})^{2}-\chi=0$$

[45] To see this, fix ϕ in (ϕ^{sust}, ϕ^{break}) for some time t+Δt. Also, assume that at time t ϕ was either below ϕ^{sust} –in which case $\lambda(t)=½$– or greater than ϕ^{break} –in which case $\lambda(t) \in \{0,1\}$. Clearly, $\lambda(t+\Delta t) = \lambda(t)$ by (2-3). In other words, location displays hysteresis.

[46] Generally, these test are also valid to asses the global stability of the system. See Chapter 1.

This expression holds for both the CP and the FE models (remember, they are isomorphic). The corner steady states are stable whenever $\phi>\phi^{sust}$.

We may also ask an alternative question: When is the symmetric equilibrium stable? To answer this question, we sign $d\omega/d\lambda$ at $\lambda=½$. Since q is strictly increasing in ω and η is increasing in λ by Result 2-1 this is equivalent to signing $dq/d\eta$ around $\eta=1$. This is <u>much</u> simpler than dealing with the usual equations in (2-1), as we did for the FCVL of Chapter 1. To see this, note first that $\eta=1$ implies that all endogenous ratios are equal to unity by symmetry of the model, viz. $Y=\Delta=q=1$. This is handy because in that case $dlnY=dY$, $dln\Delta=d\Delta$, and $dlnq=dq$. Next, using the previous fact, it is easy to show that total differentiation of the system in (1-34) can be written in matrix form as:

(2-13)
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \theta - Z^{-1} & -1 \\ -Z & Z-\theta & 1 \end{bmatrix} \begin{bmatrix} dY \\ d\Delta \\ dq \end{bmatrix} = \begin{bmatrix} (1-\chi)/(1+\chi) \\ -1 \\ 0 \end{bmatrix} d\eta$$

where Z is defined as $Z\equiv(1-\phi)/(1+\phi)$. Using Cramer's rule, we find that $dq/d\eta$ equals zero if and only if $Z(\theta - Z^{-1})(1-\chi)(1+\chi)^{-1} + Z - \theta = 0$. Solving for $Z>0$ ($\phi<1$) gives the solution for the 'break point'; the corresponding value for ϕ is defined as:

(2-14)
$$\phi^{break} \equiv \chi \frac{1-\theta}{1+\theta}$$

which is in (0,1) by the no black-hole condition and (2-6). The symmetric steady state is unstable whenever $\phi>\phi^{break}$. Interestingly, when forward linkages are absent (like in Baldwin 1999) the break and sustain points coincide, viz. $\theta=0$ implies $\phi^{break}=\phi^{sust}$.

Using the definitions of χ and θ, (2-12) and (2-14) are of course equivalent to Eq. (5.17) and Eq. (5.28) in Fujita et al. (1999, pp. 70, 74) and the corresponding points in Forslid (1999) and Ottaviano (2001). It remains to describe the stability properties of the asymmetric interior steady states, namely, of $\lambda_0 \in L_0\backslash(½)$ or $\eta_0 \in N_0\backslash(1)$. We turn to this issue next.

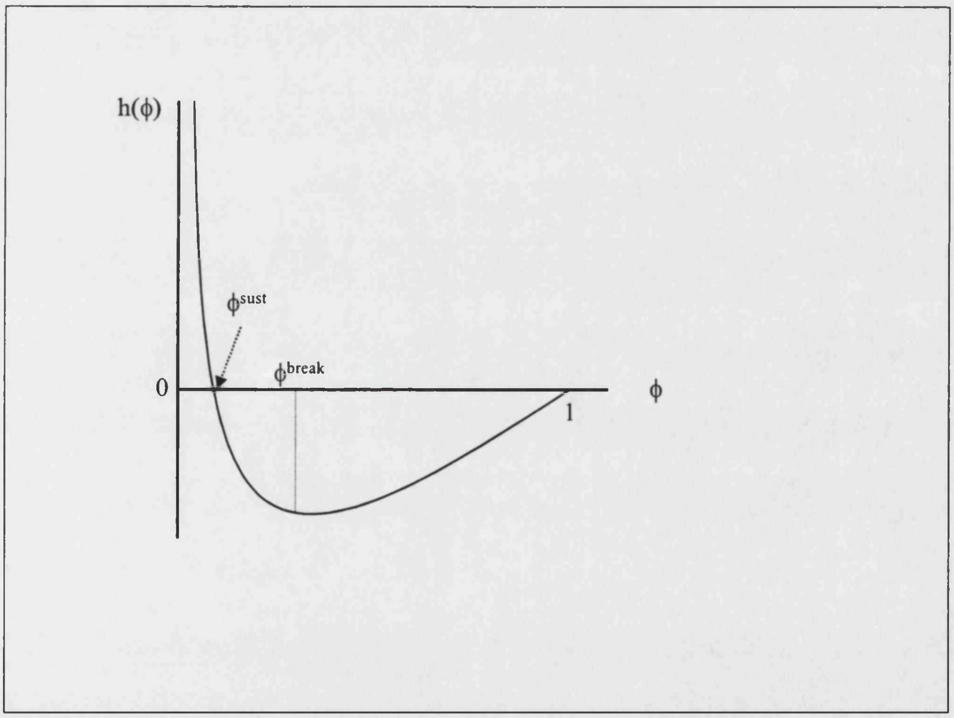
Stability of the asymmetric interior steady-states

We claim that whenever they exist such steady states are unstable. To this aim, a look at Figure 2 confirms that it is sufficient to show that $\phi^{sust}<\phi^{break}$ holds.

Result 2-5. If they exist, asymmetric equilibria are unstable.

Proof. We have to show that $\phi^{sust}<\phi^{break}$.[47] First, we characterize how the function $h(\phi)\equiv\chi+\phi^2-(1+\chi)^{1-\theta}$, which is just a transformation of the expression in (2-12), changes with ϕ. This function is of interest since ϕ^{sust} is one of its roots. With some work we can show three facts: **(a)** h(1)=0 and h'(1)>0; **(b)** h(0)>0 and h'(0)<0; **(c)** h(.)has a unique minimum. This is illustrated in Figure 2-3, in which we plot h as a function of ϕ. As can be seen on the graph, the previous properties of h taken together imply that h has a unique root between zero and unity. Next, we show that $h(\phi^{break})<0$, which is only possible if $\phi^{sust}<\phi^{break}$, given the shape of h(.). To this end, observe that $h(\phi^{break})$is a function of χ and θ. Call this new function $g(\chi,\theta)$ and note that the partial of g with respect to χ is positive (not immediate, but true) and g(1,.) is zero.[48] The point of all this is that the upper bound of g, and thus the upper bound of $h(\phi^{break})$, is zero. We know, therefore, that for permissible values of χ and θ, $\phi^{sust}<\phi^{break}$ as claimed. QED.

Figure 2-3: Proving the $\phi^{break}>\phi^{sust}$



This completes the proof and implies that the asymmetric interior steady states are unstable.

---

[47] Here we only sketch the proof, since Neary (2001) contains a detailed proof building on this.
[48] To make this last point is easy: $\chi=1$ implies $\mu=0$ (and hence and $\theta=0$) in any model; from (2-14) this in turn implies $\phi^{break}=1$ and we know that h(1)=0; the rest follows from the definition of g(.).

2.4. Concluding remarks

In this chapter I have shown that the original Core-Periphery (CP) model and the alternative Footloose-Entrepreneur (FE) model can be entirely characterized by the same set of equations in the appropriate state-cum- parameter space; in other words, they are isomorphic –see expressions (1-34) and (2-10). In particular, the natural state variable of these models is the mobile expenditure. This implies that the relevant variable to look at in empirical studies based on these models is the spatial distribution of expenditure or income (rather than population) of mobile factors. Because (1-34) and (2-10) are identical, it is sufficient to describe the properties of either to know the stability properties of both the CP and FE models. In both of these models agglomeration is driven by labour mobility.

Agglomeration mechanisms other than migration have been put forward, too. On the one hand are the models based on factor accumulation; among them, the 'Constructed-Capital' (CC) model of Baldwin (1999) is isomorphic to the CP and FE models so (1-34) and Result 2-4 hold for these as well (the CC model is a special case in which forward linkages are absent). On the other hand are the models based on input-output, or 'vertical', linkages among firms. The FCVL model introduced in Chapter 1 is isomorphic to the CP and FE models, too, and, as such shares their dynamic properties (see the Appendix to Chapter 1).

It should be possible to develop an extension of the proof of this chapter for the 'Core-Periphery Vertical-Linkages' (CPVL for short) model of Krugman and Venables (1995) and Fujita et al. (1999, section 14.2). In these models, factor owners are immobile across regions but labour moves from one sector to the other so that *nominal* wages are equalized within each region across sectors. The definition of the relevant parameters χ and θ and law of motion in (2-3) must be changed accordingly.

For the method developed here to be applied to the CPVL model as well, however, we need a supplementary trick because this model has two state variables, not one. We can do as if this was not the case and use a relation similar to (2-2) and show that this approximates the CPVL model when $q \neq 1$ (namely, outside state-states) but that the model is *exactly* characterized by (1-34) otherwise. What remains to be shown formally is that approximating the model this way does not eliminate (nor adds) any new steady state, nor does it change the stability analysis. This would imply that the

CPVL model is isomorphic to the CP model at the steady-states only; this true only as an approximation elsewhere. This task is left for further research.

It is my hope that the present chapter is helpful in understanding why most of the NEG models that build on the Dixit-Stiglitz monopolistic competition are so strikingly similar and how the agglomeration and dispersion forces interact in such a way that no more than five steady-states ever exist. Common functional forms are necessary for the isomorphism result. However, it is extremely remarkable that they are sufficient as well, since the mechanisms driving agglomeration differ from one model to the next, as does the interpretation of the parameters.

Appendix

Proof of Lemma 1. We have to show that $\forall \eta \in [0,\infty)$ there exists a $\lambda \in [0,1]$ for which $\eta = \mathfrak{M}(\lambda)$. Start with $\lambda = 0$; by (2-2) we know that $w_2 = 1$ and by definition of η, viz. $\eta \equiv \lambda w_1/((1-\lambda)w_2)$, this implies $\eta = 0$. By symmetry, we know that $\lambda = 1$ implies $\eta = \infty$. In other words, $\mathfrak{M}$ maps the least (greatest) element of the domain into the least (greatest) element of the range. Consequently, if $\mathfrak{M}(\cdot)$ is continuous then to any η in the range $[0,\infty)$ corresponds at least one λ in the domain of $\mathfrak{M}(\cdot)$. Next, if we can show that w_1 and w_2 are continuous in λ, then η is continuous in λ, too. By (2-2), $w_1(\lambda)$ is continuous if, and only if, $w_1(\lambda)/w_2(\lambda)$ is continuous, as long as both w_1 and w_2 are strictly positive. Define $\Lambda \equiv \lambda/(1-\lambda)$ and $w \equiv w_1/w_2$, and rewrite (2-1) using the ratio notation to get:

(2-15) $$Y = \frac{\Lambda w + \chi}{1 + \Lambda w \chi}; \quad \Delta = \frac{\Lambda w^{1-\sigma} + \phi}{\phi \Lambda w^{1-\sigma} + 1}; \quad w^{\sigma} = \frac{Y + \Delta\phi}{Y\phi + \Delta}$$

Three things can readily be noted from (2-15). First, whenever $\phi > 0$, $\lambda = \Lambda = 0$ implies $w^{\sigma} = (\chi/\phi + \phi)/(\chi + 1)$, a real, finite, and positive number. Also, w is finite and positive, viz. $w \notin \{0,\infty\}$, for all Λ. Assume not, i.e. set $w = 0$ or $w = \infty$ in (2-15); this implies

$0=(\chi+1)/(\chi/\phi+\phi)$ in the third expression whenever $\phi>0$, a contradiction. Finally, no denominator in the three expressions of (2-15) nor $w^{1-\sigma}$ is ever non-positive. These facts imply that each endogenous variable in (2-15), namely Y,Δ and w, is a continuous function of Λ (and of λ in turn). We can back up this result for $w_1(\lambda)$ and $w_2(\lambda)$ to claim that these two functions are continuous, too. As we saw, this in turn implies that $\mathfrak{M}(\cdot)$ is continuous, and, together with the fact that $\mathfrak{M}(0)=0$ and $\mathfrak{M}(1)=\infty$, that $\mathfrak{M}(\cdot)$ is a surjection. *QED.*

Proof of Lemma 2. Differentiate (2-15) at the symmetric steady-state to get:

(2-16)
$$\begin{bmatrix} 1 & 0 & -(1-\chi)/(1+\chi) \\ 0 & 1 & (\sigma-1)Z \\ -Z & Z & \sigma \end{bmatrix}\begin{bmatrix} dY \\ d\Delta \\ dw \end{bmatrix}=\begin{bmatrix} (1-\chi)/(1+\chi) \\ Z \\ 0 \end{bmatrix}d\Lambda$$

Solve for $dw/d\Lambda$ and plug this into $d\eta\equiv dw+d\Lambda$ to get:

(2-17)
$$\left.\frac{d\eta}{d\lambda}\right|_{\lambda=1/2}=\phi\frac{\sigma(1+\chi)}{\phi^2+(2\sigma-1)(1+\chi)\phi+\chi}$$

this is positive by inspection, as was to be shown. *QED.*

Proof of Result 2-1. (**a**) To each η corresponds at least one λ, by Lemma 1. So we are left to show that to each η corresponds at most one λ (namely, $\mathfrak{M}$ is an injection, too). The proof is by contradiction. Assume that there are two tuples $\{\lambda,w_1,w_2,Y_1,Y_2,G_1,G_2\}$ and $\{\widetilde{\lambda},\widetilde{w_1},\widetilde{w_2},\widetilde{Y_1},\widetilde{Y_2},\widetilde{G_1},\widetilde{G_2}\}$ that solve (2-8) and such that $\eta=\lambda w_1/((1-\lambda)w_2)$ and $\widetilde{\eta}=\widetilde{\lambda}\widetilde{w_1}/((1-\widetilde{\lambda})\widetilde{w_2})$. Let $\widetilde{\Delta}$ and $\widetilde{Y}$ be the equivalent to Δ and Y in (1-34) for the second tuple. Assume $\widetilde{\eta}=\eta$ and $\widetilde{\lambda}\neq\lambda$. This implies $(w_1/w_2)\neq(\widetilde{w_1}/\widetilde{w_2})$ by (2-2) and $Y=\widetilde{Y}$ by the first expression in (1-34). Next, using the wage equations in (2-15), we obtain:

(2-18) $$\Delta\frac{w-\phi}{1-\phi w}=\tilde{\Delta}\frac{\tilde{w}-\phi}{1-\phi\tilde{w}},\quad w\equiv\left(\frac{w_1}{w_2}\right)^{\sigma},\quad \tilde{w}\equiv\left(\frac{\widetilde{w_1}}{w_2}\right)^{\sigma}$$

From the relative price index in (1-34), we find:

(2-19) $$\Delta=\frac{\eta+\phi w}{\phi\eta+w},\quad \tilde{\Delta}=\frac{\eta+\phi\tilde{w}}{\phi\eta+\tilde{w}}$$

Substituting for Δ and δ into (2-18) implies $(w-\tilde{w})(1+\eta)(\eta+\tilde{w}w)(1-\phi^2)=0$. Since $\phi<1$, this holds if, and only if, $w=\tilde{w}$. This contradicts $(w_1/w_2)\neq(\widetilde{w_1}/\widetilde{w_2})$. Hence, to each η corresponds only one λ. (**b**) The proof immediately follows from Lemma 2 and (a) above. *QED.*

We note that the corresponding lemmas and result hold for the FE and FCVL models.

Proof of Result 2-3.[49] From (2-9), it is possible to get analytical solutions for w_1 and w_2. Using the definition of ω_j and the ratio notation, we get:

(2-20) $$\omega=\frac{\lambda\phi+(1-\lambda)\Phi}{(1-\lambda)\phi+\lambda\Phi}\left[\frac{\lambda+(1-\lambda)\phi}{\lambda\phi+(1-\lambda)}\right]^{\theta}$$

where $\Phi\equiv(\chi+\phi^2)(\sigma+\mu)/(2\sigma)$ is a collection of parameters. Simple algebra reveals that ω generically admits two flat points. Indeed, the numerator of $d\ln\omega/d\lambda$ is a second-order polynomial in λ. Indeed, $d\ln\omega/d\lambda$ is equal to:

(2-21) $$\frac{\theta(1-\phi^2)[\lambda\phi+(1-\lambda)\Phi][\lambda\Phi+(1-\lambda)\phi]-(\Phi^2-\phi^2)[\lambda\phi+(1-\lambda)][\lambda+(1-\lambda)\phi]}{[\lambda\phi+(1-\lambda)\Phi][\lambda\Phi+(1-\lambda)\phi][\lambda\phi+(1-\lambda)][\lambda+(1-\lambda)\phi]}$$

[49] Forslid and Ottaviano (2001) independently developed the proof for this result.

This implies that $d\omega/d\lambda$ equals zero at mot twice which, in turn, implies that the set in (2-4) admits at most three zeroes for the FO model.

Hence we write $\# L_0^{FO} \leq 3$. *QED.*

Chapter 3. THE 'VOTE-MARKET EFFECT', OR, THE POLITICAL ECONOMICS OF INDUSTRY LOCATION

3.1. Introduction

The de-location process associated with trade integration has been a major concern for European policy makers for decades. It is reflected, for instance, in the quadrupling of cohesion spending as a share of the EU budget since 1986, and in the important level of spending by member states on their disadvantaged regions such Germany's Eastern Länder and Italy's Mezzogiorno. Much of this spending is explicitly aimed at preventing, delaying or even reversing the agglomeration of economic activity in favoured regions.

The aim of the present chapter is to address the issue of agglomeration during a process of regional integration in a framework where regional policy is determined by political economy forces. More precisely, taking a laissez-faire equilibrium as a benchmark, we show how politics and economic integration interact in both directions to speed up or slow down the agglomeration process that results from integration.

State intervention at the regional level could take any form, from infrastructure spending to tax reduction and so forth. To be concrete we focus on a location-specific subsidy that reduces the fixed cost a firm faces when setting up production in the subsidised region. In our simple model, the fixed cost consists of only capital, so the location subsidy ends up as a subsidy to capital (and, in equilibrium, to the level of production). The interaction between the two regions at the political level determines the direction and the amount of the regional subsidy. We assume lump-sum transfers are available; in this context, the policy instrument chosen is non-distortionary, which implies that the results of our analysis do not depend upon the choice of instrument.

The political economy model we work with is based on electoral competition rather than on a lobbying approach. Sectoral concerns are likely to be transmitted to the decision makers through lobbying activities (see e.g. Becker 1983, Grossman and Helpman 1995, Olson 1965). However, when we look at regional issues, we see regions as spatial entities and not as sectors. Since they are often recognized as distinct entities in the political system, regions are more likely to influence the policy outcome directly, i.e. through elections. Hence, we will not rely on a lobbying approach to characterize the political game. Instead, we use a Hotelling-Downs probabilistic-voting model (Hinich 1977, Ledyard 1984).

Probabilistic-voting models (also referred to as swing-voter models) feature a second, political, dimension, in addition to the economic-policy dimension that characterizes median-voter models. Voters are endowed with non-policy preferences over the two parties (what we call 'ideology'). The economic-policy dimension is, by assumption, orthogonal to 'ideology'. Candidates know voters' preferences when it comes to the economic issue at hand, but they know only the distribution of voters' preferences in the ideology dimension. As we shall see, this assumption implies a dramatic departure from the median voter's 'dictatorship' (Hinich 1977); in particular, the outcome can be non-majoritarian on some (or even all) dimensions. The key is that the 'swing voter' (the voter whose ballot can be thought of as winning the election) need not be the voter whose preferences reflect the median in the economic dimension because voters also care about ideology. In this set up, a group of voters with particularly uniform ideology are more likely to be swing voters. Knowing this, candidates will craft their policy platforms to cater to the economic interest of this group.[50]

We assume that the population in the urban region is more widely spread out along the ideology dimension than the population in the small, rural one. We make this assumption on the grounds that economic activities and, hence, special interests are more variegated in more urbanized regions than in less urbanized ones (or, equivalently, that regional policy is a less salient issue).[51] As we shall see, this stylised fact will shift the equilibrium policy variable in favour of the economically small region (see Persson and Tabellini 2000, chapter 3).[52]

Indeed, there is some indirect evidence that politically powerful regions are not always the largest ones when it comes to regional policies. As an illustration, take the

---

[50] See Coughlin (1992) for an exhaustive analysis of the probabilistic voting approach to representative democracy and Anderson et al. (1992) for the theory of discrete choice when preferences are random.

[51] This stylised fact - that larger cities tend to be more diversified - is one of many on city diversity and specialisation, as surveyed by Duranton and Puga (2000).

[52] The fact that small groups sometimes possess disproportionate political power is not new in the literature. Alternative explanations exist. For instance, small groups can presumably circumvent the free-rider problem since they are more easily able to organise themselves into pressure groups (Olson 1965). Or the electoral system might incite competing candidates to appeal to narrowly defined and specific groups rather than more broadly (Myerson 1993, Persson and Tabellini 2000, chapter 8).

repartition of the EU 1994-99 budget devoted to the 'Structural Funds'. Almost 72% of this was allocated to 'Objective 1' European regions, namely to regions that are mainly peripheral and with relatively little industrial activities.[53] All the same, people living in 'Objective 1' European regions accounted for approximately 26% of the total European population. In total, aids used to finance objectives with a specifically regional nature (Objectives 1, 2, 5b and, the recently created, 6) accounted for 87% of the total Structural Funds. The same bent can be found also at the national level. In 1999-2000, for example, subsidies granted to firms located in the Mezzogiorno (Southern Italy), where 36% of the total Italian population lives, were twice as much as those given to firms located in the rest of the country.[54]

In addition to accounting for the commonly observed phenomenon of anti-agglomeration policy, the endogenisation of policy has interesting implications for the economic geography literature. Once regional policy is considered as a political issue, economic integration does not necessarily lead to full agglomeration of industries in the larger region, as the orthodox geography model would predict. The location of economic activity will depend both on the economic home-market effect and on what we call by analogy the vote-market effect. As usual, we find that low levels of openness to trade correspond to dispersed outcomes (neither region attracts all firms), but sufficiently high levels of openness result in a core-periphery pattern. One of our novel results concerns the location of the core (orthodox economic geography models predict that it must be in the big region). However, if the economically small region is politically over-represented, the big region attracts the core if and only if its relative economic size overcomes its relative political weakness. Finally, and interestingly, although the equilibrium spatial allocation of industry is never ambiguous, the question of which region gets subsidised has an ambiguous answer, the answer being determined by both economic and political considerations. As we shall see, this is partly due to the fact that agglomeration creates quasi-rents that can be taxed in the core without leading to re-location.

The rest of the chapter is organized as follows. The next section introduces the basic economic model and solves it taking the policy variable as given. Section 3 presents the reduced form of the welfare functions and discusses how utility is affected

---

[53] Formally, 'Objective 1' regions are regions with per capita GDP below 75% of the EU's average.
[54] The sources are Eurostat for the EU data and the Italian Treasury's 'Terzo Rapporto sullo Sviluppo Territoriale, 1999-2000' for the Italian data.

by the policy choice, still exogenously given. In Section 4 the economic model is integrated into a political economy model and the two are solved together. Section 5 discusses the results. The concluding section considers some casual empirics that support our model. Some proofs are relegated to the appendix.

3.2. The augmented FC model

The FC model on which this chapter builds is described at length in Chapter 1, so here we skip the details but stress the additional assumptions. The basic set-up consists of two regions (1 and 2) that belong to the same nation; two factors (labour L and physical capital K) and two sectors, manufacturing M and agriculture A. Individuals have identical preferences, endowments, and technology. Regions differ in their size only, so that region 1 is just the upscale version of region 2. In particular, 1 is endowed with $\Lambda>1$ times as much of both capital and labour as 2. For this reason, we will sometimes refer to Λ as the relative economic strength of region 1 (or, equivalently, as the relative economic weakness of region 2).

Turn to technology. Both labour and capital are used to produce the differentiated good M under increasing return to scale and monopolistic competition. Production of each manufacturing variety involves a one-time fixed cost consisting of one unit of K and a per-unit-of-output cost consisting of β units of L. Sector A produces a homogenous good under constant return to scale and perfect competition using one units of labour per unit of output. Labour is the only input. This good is also chosen as the numéraire. Labour is perfectly mobile across sectors, but immobile across regions.

Physical capital can move freely between regions, but capital owners cannot, so all K-reward is repatriated to the country of origin. Industrial and agricultural goods are traded. Trade in A is costless. Industrial trade is impeded by frictional (i.e., 'iceberg') import and barriers and transportation costs such that $\tau \geq 1$ units of a good must be shipped in order to sell one unit abroad. In this chapter τ is mostly interpreted as technical trade barriers (τ generated no tariff revenue). Accordingly, we refer to regional integration as a gradual fall in τ.

Preferences of the representative consumer comprise the usual Cobb-Douglas nest of a CES aggregate of industrial varieties and consumption of the A-good. More precisely, in this chapter we take a logarithmic functional form, viz. $U=\mu\ln(M)+(1-$

μ)lnA, where M is the CES composite of all manufacturing varieties and A is the amount of good A consumed.

Each region's representative consumer owns the entire region's L and K and her income (and expenditure) equals wL+ρK, where ρ is the private return on capital; there is no tariff revenue with iceberg barriers. This consumer pays lump-sum tax T so as to finance the central government's government regional policy, hence her disposable income is wL+ρK-T. There are no savings in this static model, so private expenditure equals disposable income. Therefore, we can write:

(3-1) $$E_1 = L_1 + \rho K_1 - L_1 T$$

and E_2 is isomorphic.

We assume that state intervention consists of a subsidy to firm's fixed costs. As such subsidies are independent of output and, given that the one-time fixed cost consists of one unit of K, they actually represent a subsidy to capital.[55] Let π_1 (π_2) be the before-subsidy capital reward, equal to operating profits by free-entry, to entrepreneurs producing in region 1 (region 2). Also, define θ as π_1/π_2, with $\theta>0$. When rural production is subsidized, $\theta>1$ and θ-1 is the ad-valorem subsidy given to capital owners who produce in that region, regardless of the region they reside in. Under these circumstances, $\theta\pi_2$ is the capital reward for producing in 2 and π_1 is the capital reward for producing in 1. When $\theta<1$, $1/\theta$-1 is the ad-valorem subsidy given to capital owners producing in region 1. In this case, capital owners that set up their firms in 1 get π_1/θ, those setting up their firms in 2 get π_2. Clearly $\theta=1$ means that capital is not subsidized anywhere and the lower is $|\theta-1|$, the smaller is the subsidy.

In what follows the location subsidy is determined implicitly as the result of a political game between the two regions. The fact that at most one region is subsidized at a time is without loss of generality. This is because, in New Economic Geography (and related) models, only relative sizes matter (see Chapter 2). Moreover, this is always an equilibrium even in the more general case in which both regions can be subsidized.[56]

Competition for K drives operating profits up to the level where pure profits are eliminated, so K's reward in 1 is the operating profit of a typical region 1-based firm. A

---

[55] By free-entry, though, this will be equivalent at equilibrium to a subsidy to the value of sales (estimated at f.o.b. prices) and hence to the quantity of output.
[56] We can also use perturbations to the game so that this would be the unique equilibrium.

similar condition holds in 2. When production takes place in both regions, perfect capital mobility equalizes the after tax/subsidies rewards to capital across regions, i.e. $\rho_1=\rho_2$ in equilibrium. Due to subsidies, however, before tax/subsidies rewards π_1 and π_2 can differ. In particular, when setting a firm in region 2 is subsidized (viz. $\theta>1$), the prevailing equilibrium private return on capital ρ is equal to $\pi_1=\theta\pi_2$. When $\theta<1$, namely when firms in 1 are subsidized, $\rho=\pi_1/\theta=\pi_2$.

Since manufacturing firms need one unit of capital per variety, and capital is inter-regionally mobile (even though its reward is repatriated), capital's full-employment condition requires the total mass of firms to be equal to the aggregate mass of capital, which we normalise, to unity. Hence the location equilibrium is entirely determined by the proportion n of firms that settle in 1. As usual, it is even more convenient to work with ratios: θ can be interpreted as the ratio of subsidies granted to firms in 2 to the subsidies granted to firms in 1. Similarly, we write η as the mass of firms that settle in 1 over the mass of firms that settle in 2, viz. $\eta=n/(1-n)$. In the same spirit, Λ represents both the relative population sizes and relative nominal incomes as the result of three assumptions taken together. First, capital ownership is uniform across the country (made of regions 1 and 2). Second, free trade in A and free capital mobility ensure that nominal rewards are equalized throughout. Finally, everybody pays the same per-capita tax T in this economy.

T is endogenous in the model. We assume the government has a unique role –to set the regional policy. Hence T is a function of θ. Imagine $\theta>1$ so that firms in 2 are subsidised. Therefore, the government budget is balanced if, and only if,

(3-2) $$T(L_1+L_2)=(\theta-1)(1-n)\pi_2$$

where the left-hand side is the government revenue and the right-hand side represents the ad-valorem subsidy paid on the operating profit that the (1-n) firms that operate in region 2 obtain. As always, an equilibrium condition is that expenditures on M's output is equal to the value of this output, viz. $\mu(E_1+E_2)=np_1x_1+(1-n)p_2x_2$. As we saw in Chapter 1, it is a regular feature of the DS monopolistic competition framework that $\pi_1=p_1x_1/\sigma$ and $\pi_2=p_2x_2/\sigma$, where the p's are producer prices and the x's are typical outputs. Using this, we can plug (3-2) into (3-1) to get an expression for ρ (as a function of θ and the parameters of the model and aggregate endowments), viz. $\rho[n+(1-n)/\theta](\sigma-\mu)=\mu(L_1+L_2)$. This is not interesting in itself. What is more important is that incomes do not depend on θor on ρ. Indeed, plug this expression for ρ in (3-2) and (3-1) to get:

(3-3) $$E_1 = L_1 \frac{\sigma}{\sigma - \mu}, \quad E_1 = L_2 \frac{\sigma}{\sigma - \mu}$$

which is identical to the case without subsidy and hence ensures that the following result holds:

Result 3-1. Relative expenditures are a function of relative endowments only, viz. $E_1/E_2=\Lambda$.

To get an intuition for this result, note that each individual is both a taxpayer and a capital owner. Since taxes are collected in a lump sum fashion, there is no efficiency loss associated with the introduction of the policy. This ensures that what each consumer gives with the left hand as a taxpayer will be paid back to her right hand as a capital owner. Of course, if capital ownership is not uniform, the policy has re-distributive effects. In any case, though, this has no aggregate effect.

We now have everything at hand to solve for the equilibrium location of the model.

Equilibrium location

To solve for η, we use the same technique as in Chapter 1. For convenience, we rewrite here the equilibrium expressions for the operating profits. As in Chapter 1 define π as the average profit prevailing in this nation, viz. $\pi=n\pi_1+(1-n)\pi_2$. Accordingly, define q_j as π_j/π and e_j as $E_j/(E_1+E_2)$ (j=1,2). Note that e_j is proportional to the population in j, so that $e_1/e_2=\Lambda$. With these definitions, the equilibrium solutions for the operating profits can be written as:

(3-4) $$q_1 = \frac{e_1}{n + \phi(1-n)} + \phi \frac{e_2}{\phi n + (1-n)}$$

The solution for q_2 is isomorphic.

With capital mobility, the number of varieties produced in a region may differ from the region's capital stock, so we also have to determine the equilibrium location of manufactures. To close the model, we invoke the non-arbitrage condition on capital markets. The novelty here is that returns on capital may be taxed on a regional basis, hence free capital mobility ensures that $\pi_1/\pi_2=\theta$ whenever firms are active in both regions (with the obvious Kuhn-Tucker conditions more generally).

The solution to this problem, as expressed for η, is:

(3-5) $$\eta = \frac{(1-\theta\phi)\Lambda - \phi(\theta-\phi)}{-(1-\theta\phi)\Lambda\phi + (\theta-\phi)}$$

This expression holds for admissible values of η, namely when parameters are such that $0<\eta<+\infty$. Outside this parameter space, η equals zero or $+\infty$ in an obvious manner. This expression is more informative than what might be inferred at first sight. By inspection, region 1's relative share of firms η is increasing in 1's relative size Λ and is decreasing in region 1's relative cost of capital θ. Moreover, η is larger than Λ if $\Lambda>1>\theta$ (not immediate, but true). These inequalities illustrate effects that will be recurrent in the sequel.

Expression (3-5) is the fulcrum of our analysis, so it is worth studying it in the absence of subsidies, i.e. when $\theta=1$. In this case the equilibrium η becomes:

(3-6) $$\eta\big|_{\theta=1} = \frac{\Lambda-\phi}{1-\Lambda\phi}$$

When 1 is larger, making trade freer ($d\phi>0$) results in a de-location of firms to the big region. This is consistent with the 'New Trade' models a-la Krugman (1980). In particular, the Home-Market effect (HME) manifests itself as:

(3-7) $$\left.\frac{\partial\eta}{\partial\Lambda}\right|_{\theta=1} = \frac{1-\phi^2}{(1-\Lambda\phi)^2} > 1; \quad \lim_{\phi\to 1/\Lambda} \eta = +\infty$$

The first expression above says that a larger region will get a more than proportional, larger share of industry. It holds whenever $\phi<1/\Lambda$. On the other hand, when $\phi\geq 1/\Lambda$ the second expression in (3-7) says that all firms cluster in 1, viz. $\eta=+\infty$ or $n=1$. In words, all firms cluster in the larger region when trade costs are low enough (ϕ sufficiently close to unity), yet strictly positive (ϕ strictly lower than unity). Note also the magnification effect of trade liberalisation of the first term, viz. $\partial^2\eta/\partial\Lambda\partial\phi>0$. These results are a re-statement of Result 2-7.

In order to isolate the effect of a subsidy on the firm allocation share, we now take the opposite simplifying case and calculate the equilibrium η for $\Lambda=1$ and take the first derivative of this with respect to $1/\theta$. Using (3-5), it is easy to show that $\partial\eta/\partial(1/\theta)$

is larger than one when Λ=1. This says that one additional unit of subsidy given to 1 leads to a more than proportional change in de-location towards 1. We call this property the Home-Subsidy effect by analogy with the HME. Trade liberalisation magnifies this effect as well, viz. $\partial^2\eta/\partial(1/\theta)\partial\phi>0$. To sum-up:

> Result 3-2. Ceteris paribus, the region that has the larger income or the region that is subsidised has an equilibrium share of firm that is more than proportionally larger than its relative income or than its relative subsidy. These biases are magnified by low values of ϕ.

Both the home market and the home subsidy effects will be used in the following sections to help boost intuition.

Finally, equilibrium A-sector output is determined as a residual.

3.3. The 'subsidy effect' on location

The purpose of this chapter is to determine θ, and ultimately η, as the outcome of antagonist political forces. Bet let us pose a minute and study first the effects of θ on η and on welfare, as these are important to understand the role of the subsidies of the political game to be introduced in the next section.

Start with the following definitions:

(3-8) $$\underline{\theta} \equiv \max\{\theta : n=1\}, \quad \overline{\theta} \equiv \min\{\theta : n=0\}, \quad \Theta \equiv [\underline{\theta}, \overline{\theta}]$$

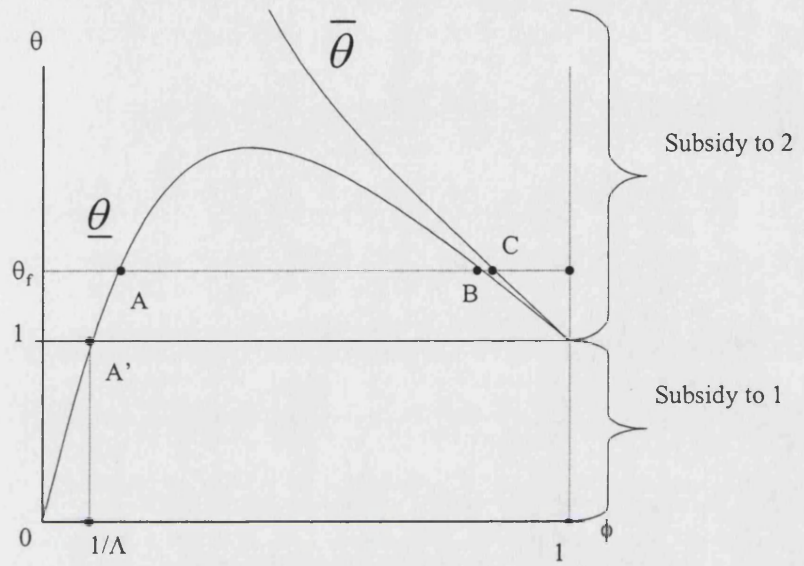
where n is the location equilibrium given by (3-5), with $n=\eta/(1+\eta)$.

Figure 3-1 plots $\overline{\theta}$ and $\underline{\theta}$ as a function of trade freeness ϕ for a given value of Λ. The former shows the minimum level of subsidization necessary to attract all firms in the small region 2. The latter shows the minimum level of subsidy (possibly negative) to firms located in region 1 that is necessary to concentrate industry there. Θ is the vertical interval between the two curves. Substituting η=0 and $\eta=+\infty$ in (3-5), these two parameters and the interval they form are defined respectively as:

(3-9) $$\overline{\theta} \equiv \frac{\Lambda+\phi^2}{(1+\Lambda)\phi}; \quad \underline{\theta}^{-1} \equiv \frac{1+\Lambda\phi^2}{(1+\Lambda)\phi}; \quad \Theta \equiv [\underline{\theta}, \overline{\theta}]$$

As is possible to see from the top curve, the freer is trade, the smaller is the subsidy required to attract all the firms to region 2 because liberalization amplifies the effect of a subsidy. This is because location becomes ever more sensitive to regional disparities as ϕ increases, be they in size, in the cost of capital (the variable affected by the subsidy), or anything else. Note that, since $\overline{\theta} > 1$ holds $\forall\phi$ the minimum level of subsidy to industries in region 2 that makes this small region the core is always positive: a positive regional policy is necessary in order to offset the tendency of the small region to lose firms as ϕ rises.

Figure 3-1. Subsidies and equilibrium location



The converse is not necessarily true, as a look at $\underline{\theta}$ (the bottom curve) shows: the relationship between the minimum level of subsidies to firms located in region 1 necessary to keep all the firms in the same region and the level of trade integration is bell-shaped. The reason is that the HME works in favour of the large region, so that a small tax on firms there (or, equivalently, a small subsidy on potential firms in 2) is ineffective in making any firm move to the smaller region for any $\phi>1/\Lambda$. The fact that taxing capital in the Core does not necessarily lead it to relocate comes in sharp contrast to the classical results on tax competition.

It is also instructive to consider the effectiveness of a given level of subsidy to region 2's firms along the integration path. In particular consider what are the location effects when ϕ varies with the level of θ fixed at some arbitrary θ_f in the diagram. When

ϕ is relatively close to zero there is some economic activity in both regions since $\theta \in \Theta$. As the two regions become more integrated, the HME starts dominating the subsidy effect. From point A onwards, the relative strength of the HME is so reinforced by an ongoing integration process that this level of subsidy to industrial activity in 2 is completely ineffective and this small region becomes the periphery region ($\eta \to \infty$) in spite of the subsidy on offer. As ϕ continues to increase, the relative strength of the HME decreases and eventually, when point B is reached, some of the firms start leaving region 1 (which is hence no longer a core). Things get even worse for the larger region as transportation costs fall further: to the right of point C the core is in 2. Again, if we take the model literally, this shows how effective regional policy is when ϕ is close to unity: without any subsidy (θ=1), the core would be (and remain) in 1 from point A' onwards.

Lastly note that an increase in Λ makes the HME become stronger. As it is possible to see from (3-9), a higher value of Λ makes both $\underline{\theta}$ and $\overline{\theta}$ shift upward. The upward movement of $\overline{\theta}$ indicates that, given ϕ, in order to compensate for the fact that the big market attracts firms. As a result the minimum subsidy to region 2's firms needed to keep the core there has to be higher. Likewise, the upward shift of $\underline{\theta}$ implies that the minimum subsidy level needed to ensure that the core is in region 1 is now lower. Conversely, the range of trade freeness for which a small subsidy offered to location in 2 is still compatible with the core remaining in region 1 is wider.

The 'subsidy effect' on welfare

Since equilibrium nominal incomes are function of the parameters only, in particular they are invariant in η or θ, the welfare of the representative individual is function of the price index prevailing in the region she lives in. Mathematically, the assumed functional forms give us the following expression for the indirect utility functions:

(3-10)
$$V_1(\theta;\phi) = \ln\left(\frac{\mu}{\sigma-\mu}\right) + \frac{\mu}{\sigma-1}\ln\left[\frac{\eta(\theta)+\phi}{\eta(\theta)+1}\right]$$

for region 1's representative consumer and

(3-11)
$$V_2(\theta;\phi) = \ln\left(\frac{\mu}{\sigma-\mu}\right) + \frac{\mu}{\sigma-1}\ln\left[\frac{\eta(\theta)\phi+1}{\eta(\theta)+1}\right]$$

for region 2's representative consumer, with η taken from (3-5). The first term in the right-hand side of each expression above is the (log of) per-capita income and the second term is the true price index for each representative consumer. As can easily been inferred from these expressions, welfare is monotonically increasing in the ratio of firms that locate in the agent's region as this person would then save on transportation costs. See Chapter 1 for details.

This implies that $\partial V_1/\partial\eta>0$ and $\partial V_2/\partial\eta<0$. Since η itself is decreasing in θ, we have the obvious relationship between indirect utilities and the subsidy given to region 2: $\partial V_1/\partial\theta<0$ and $\partial V_2/\partial\theta>0$. Of course, (3-10) and (3-11) hold for values of η in $(0,+\infty)$ or, which is the same thing by (3-8), for values of θ in Θ. Hence, we have:

> Result 3-3. $\underline{\theta}$ and $\overline{\theta}$ are the bliss points of any individual in the regions 1 and 2, respectively.

With this analysis at hand, we now turn to the political process that shall determine θ and η.

3.4. The voting model

The political environment is as follows. Both regions belong to the same constituency. All voters, whether living in 1 (as L_1 of them do) or in 2 (as L_1 of them do), chose a candidate from the same set of candidates. This set is exogenously given for simplicity.

The political game belongs to the Hotelling-Ledyard class of models and makes the following assumptions (see Osborne, 1995):

1. The policy space Θ as defined in (3-8) is one-dimensional;
2. The set of candidates {A,B} is fixed and finite;
3. Each candidate is 'Downsian' in that she cares only about winning office and is assumed to maximize her expected number of votes;
4. The number of citizens, whose preferences are monotonic on Θ, is finite and equal to L_1+L_2;
5. Candidates simultaneously choose a position on Θ (their 'platform');

6. Having observed the candidates' platform, voters decide whether to vote or not and, if so, for which candidate. Voting is costless.

Additionally, our formulation of the voting model follows Lindbeck and Weibull (1984). Candidates differ not only on the policy issue (i.e. in their platform), but also on a second dimension, orthogonal to the policy space –call it ideology or party membership. The ideology of a candidate is not part of her platform because she cannot credibly change it, by assumption.[57]

Voters derive utility both from consumption and how much their own ideology matches the winning candidate's own, assumed to be unrelated to consumption for simplicity. Hence, their utility function has, in fact, two components. The materialistic component (V_1 and V_2 as derived in (3-10) and (3-11) above) is directly affected by the subsidy policy and is known to both candidates.

The other component of people's utility is derived from other policies proposed by the competing parties or by attributes specific to the candidates. Importantly, the candidates know only the distribution of the voters along this dimension. Hence, in effect, voters face a discrete choice between two candidates that they perceive as different, even if the latter choose the same platform θ. This parallels the discrete choice theory of product differentiation, in which firms cannot observe all the variables affecting consumer choices. (See Anderson, de Palma and Thisse (1992) for a contained treatment of this theory.) Hence, even if each consumer or voter chooses a single option (to buy the product or not, to vote for one candidate or the other), the outside observer (a firm or a candidate) sees utility as a random variable reflecting unobservable preferences. We adopt this interpretation of the model, though other interpretations are also possible (see Anderson et al. 1992).

At the time they simultaneously announce their platform candidates know only the distribution of voters along the ideological dimension. If candidate A and B share the same platform, a given voter prefers the first of the two if candidate A's ideology is closer to hers.

---

[57] This dimension is assumed to merge all the issues voters might care about –that is, all but regional policy. It includes the parties' manifesto on moral issues like abortion or the death penalty, as well as on long standing economic positions like commitment to free-trade, joining the Euro, and so on. The key assumption here is that on all these issues the candidates are already pre-committed, whilst they are (credibly) campaigning on the regional policy issue only.

If, on the other hand, candidate B proposes a platform on the policy issue that suits this voter better than A's, the voter trades off her ideological preference against her policy one. To what extent she is willing to do so depends upon the strength of her partisanship, a variable that is unknown to anybody but her. This assumption will ensure that each candidate's best response correspondence will be a smooth, continuous function.[58]

As far as the voting rule is concerned, we abstract from entry issues and assume that two candidates, belonging to two distinct parties, compete for office. Each candidate $i \in \{A,B\}$ proposes a value θ_i in Θ, that is, suggests to what extend she wishes to subsidize region 2. Voters cast a ballot for one of the candidates, according to their idiosyncratic preference and to the candidates' platforms. The elected candidate sticks to her policy platform once in office; her promise is credible in game theoretical terms, since candidates do not care about the policy outcome.

Voters' payoffs are as follows. As an illustration take a region 1 voter, voter j. If candidate B is elected then voter j's utility (both materialistic and derived from her 'ideology') is assumed to be equal to $V(\theta_B)+\varepsilon_j/2$. $V(\cdot)$ is the materialistic utility and measures the voter's economic welfare derived from the implementation of candidate B's economic platform, whereas ε_j is voter j's idiosyncratic ideological bias towards party B and measures the utility she derives from B's political leadership; this term is negative if voter j is ideologically closer to candidate *A*, and equal to zero if she is ideologically neutral and cares only about economic policy.

Conversely, if candidate A is elected voter j enjoys utility $V(\theta_A)-\varepsilon_j/2$. This voter is indifferent between candidates A and B for given platforms θ_A and θ_B if, and only if,

(3-12) $$V(\theta_A)-V(\theta_B)=\varepsilon_j$$

A voter for which (3-12) holds votes for candidate *A* with probability ½, but would vote for the same candidate with probability one or with probability zero if this expression held with strict inequality.

---

[58] Anticipating the results, the assumptions just listed will ensure the convergence of the candidates' platform to a unique equilibrium policy. This outcome is a Condorcet winner, i.e. a policy that beats in probability any other feasible policy in a pair-wise vote.

Given θ_A and θ_B, this voter is ideologically neutral. We call such a voter a 'swing voter'. Hereafter we refer to the value of her idiosyncratic parameter as ε_S (subscript 'S' for 'swing'). Similar expressions hold in region 2, therefore

(3-13) $$\varepsilon_{S,1} = V_2(\theta_A) - V_2(\theta_B), \quad \varepsilon_{S,2} = V_2(\theta_A) - V_2(\theta_B)$$

Timing of the game

The timing of the elections is as follows. First, both candidate A and candidate B announce their platforms simultaneously and non-cooperatively, knowing the preferences of voters over θ and the probability density function of $\varepsilon_{j,1}$ and $\varepsilon_{j,2}$. Second, uncertainty is resolved and voting takes place. Finally, the elected candidate implements the platform she announced in the first stage.

We now introduce an important assumption. In the two regions all ε's are drawn from a continuously differentiable, symmetric, and with mean zero cumulative distribution function $F_1(\varepsilon)$ in region 1 and $F_2(\varepsilon)$ in region 2. These c.d.f.'s are known to anybody (in other terms there is no aggregate uncertainty). We assume that $F(\varepsilon)$and $F_2(\varepsilon)$ belong to the same family.[59]

Our working assumption is that social and economic activities are more variegated and heterogeneous in the big region. Indeed, it is a stylised fact that larger cities tend to be more diversified (Duranton and Puga 2000). Mathematically this translates into a higher dispersion of the cumulative distribution around the mean in the big region. Hence, if ψ^2 is the variance of $F_1(\varepsilon)$ and γ^2 the variance of $F_2(\varepsilon)$, we assume $\psi>\gamma$.[60] There is a technical issue here: to be rigorous, double-sided uncertainty is needed for the equilibrium to exist. That is, some of the attributes of the candidates are unknown to the voters at the time candidates choose their platforms. As a consequence, the mean of both $F_1(\varepsilon)$ and $F_2(\varepsilon)$ is itself a random variable. We take it to be symmetrically distributed around 0. The algebra below is unaffected by other

---

[59] Invoking a CLT, we could take normal distributions.

[60] Another way of interpreting this assumption is as follows: Assume that $F_1(\varepsilon)$ and $F_2(\varepsilon)$ are identical with variance 1, but that the spatial distribution of industry is a more salient issue to those voters left behind in the periphery than to those living in the core, as it presumably is. That is, assume now that the welfare of individual j in 1 or 2 is $(1-\psi')V_1(\theta)+\psi'\varepsilon_j$ or $(1-\gamma')V_2(\theta)+\gamma'\varepsilon_j$, respectively. The assumption that regional policy is more salient in 2 is equivalent to assuming $\psi>\gamma$. In aggregate terms, the two interpretations are equivalent. (Indeed, with $\psi'=\psi/(1-\psi)$ and $\gamma'=\gamma/(1+\gamma)$ all the analytical results below are strictly identical.) In what follows, we use the terminology of the interpretation developed in the text.

parameters of the distribution of the common mean (like its variance that we assume to be finite) so we leave this issue in the background from now on.

It can be shown that $F_1(\varepsilon)$ and $F_2(\varepsilon)$ together with $V_1(\theta)$ and $V_2(\theta)$ fulfil the sufficient conditions for a Nash equilibrium in the platform setting game to exist because they are quasi-concave in θ. To see this, note that the V's are concave with respect to η and that η is a monotonic function of θ; together, these facts ensure the result.[61] In particular, it is sufficient for the V's to be concave in n – which they are – given our assumptions on the F's.

Given these assumptions, candidate A's expected vote share is equal to

(3-14) $$s_A = \frac{\Lambda}{\Lambda+1}F_1(\varepsilon_{S,1}) + \frac{1}{\Lambda+1}F_2(\varepsilon_{S,2}) = \frac{1}{2} + \frac{\Omega(\theta_A,\theta_B)}{2}$$

where

(3-15) $$\Omega(\theta_A,\theta_B) \equiv \frac{\Lambda}{\Lambda+1}F_1\left[V_1(\theta_A) - V_1(\theta_B)\right] + \frac{1}{\Lambda+1}F_2\left[V_2(\theta_A) - V_2(\theta_B)\right]$$

Candidate *A*'s probability of winning the election is increasing in $\Omega(\theta_A,\theta_B)$. Candidate B's expected vote share is equal to $1\text{-}s_A$.

If the ε's were uniformly distributed Ω would look like a weighted social welfare function, with the weights to V_1 and V_2 being proportional to the 'economic' sizes $\Lambda/(\Lambda+1)$ and $1/(\Lambda+1)$, respectively, and inversely proportional to the standard deviations of the idiosyncratic ideological preferences (ψ and γ). In other words, a region 1 vote is worth Λ/ψ and a Region 2ern vote is worth $1/\gamma$ to the politicians. This is due to the underlying uncertainty in the ε's. The same intuition carries over for more general distributions F.

To understand the incentives the candidates face when setting their platform, suppose for a while that candidate A is considering to increase her platform from θ_A to θ'_A. A higher θ will boost the equilibrium number of firms in 2 at the expense of 1, causing an increase in 2's voters' materialistic utility and a reduction in 1's voters' materialistic utility. This shifts the identity of the swing voters in both regions, as can be seen from expression (3-13): as a result of her unilateral deviation, candidate A gains some votes in 2 and looses others in 1. Candidate A has no-longer any incentive to

[61] See e.g. Theorem 2 in Lindberg and Weibull (1987) for details.

deviate when the number of votes lost in the later is equal to the number of votes gained in the former. This brings us to the solution of the candidates' problem.

The Nash-equilibrium of the platform setting game

Candidate A maximizes her expected number of votes given by (3-14), announcing a certain policy θ_A taking θ_B as given. Candidate B instead will seek to minimizes s^A choosing θ_B for a given θ_A. Hence, $\theta_A^{NE}=\arg\max\{\Omega(\theta_A,\theta_B^{NE}): \theta_A\in\Theta\}$ and $\theta_B^{NE}=\arg\min\{\Omega(\theta_A^{NE},\theta_B): \theta_B\in\Theta\}$. In words, the objectives of candidates *A* and *B* are perfectly symmetric and so they face the same optimisation problem: at equilibrium, they both find it optimal to announce the same level of subsidies, call this θ_{NE} for short ('NE' for Nash equilibrium).

Using (3-10), (3-11), (3-5), and (3-15) the equilibrium policy announcement θ_{NE} is the solution to the following first order condition of this problem for both candidates and reduces to:

(3-16) $$\left[\Lambda f_1(0)\frac{\partial V}{\partial n}(\theta_{NE})+f_2(0)\frac{\partial V^*}{\partial n}(\theta_{NE})\right]\frac{\partial n}{\partial \theta}(\theta_{NE})=0$$

The second order condition is satisfied by the quasi-concavity of the V's with respect to θ. Notice that, due to standard statistical properties of probability distribution functions belonging to the same family, we have $f_2(0)/ f_1(0)=\psi/\gamma$. Define m as the ratio of the two standard deviations, viz.

(3-17) $$m\equiv\frac{\psi}{\gamma}=\frac{f_2(0)}{f_1(0)}>1$$

We can interpret m as the relative 'political strength' of region 2.

In (3-16) $\Lambda f_1(0)$ and $f_2(0)$ represent the mass of swing voters in regions 1 and 2, respectively (up to a factor)– namely, the mass of those voters that are marginally indifferent between the two candidates at equilibrium. This mass is increasing with the size of the region electorate (L_1 or L_2) and inversely related to the spread of the population along the political dimension (ψ or γ). Finally, solving (3-16) for θ_{NE} gives:

(3-18) $$\theta_{NE}=1+\frac{(m-1)(1-\phi)}{1+\phi m}$$

if $\theta_{NE} \in \Theta$, or θ_{NE} is equal to either boundary of Θ in an obvious manner otherwise.[62] Since $\phi<1$, this is larger than unity under the assumption $\psi>\gamma$.[63] To get the political equilibrium location, plug (3-18) into (3-5) to get:

(3-19) $$\eta_{NE} = \frac{\Lambda - m\phi}{m - \Lambda\phi}$$

Simple derivations give the expected signs for the following partial derivatives: $\partial\eta_{NE}/\partial\Lambda>0$ and $\partial\eta_{NE}/\partial m<0$, namely, the share of firms in 1 increases with its size and decreases in the 2's political strength. To sum-up, we have shown:

> Result 3-4. At the political equilibrium each region's share of industry is increasing in its size and in its ideological homogeneity. The equilibrium subsidy is increasing in a region's ideological homogeneity.

3.5. The vote-market and the net-market effects

Many interesting results steam from equations (3-18) and (3-19). Since the ultimate concern of voters is to attract economic activities in the region where they live, let us focus first on η_{NE}. As it is clear from (3-19), once we introduce the political dimension as a determinant of the policy decision, the equilibrium industry share does not depend solely on the economic forces at work anymore. Indeed, the equilibrium share of industry of, say, region 1 is increasing with its (relative) expenditure size Λ – essentially an economic parameter – and decreasing in its ideological heterogeneity m – a socio-economic parameter that reflects the relative salience of the regional policy issue for those living in region 2.

The first of these two effects is well known and is an alternative formulation of the standard home-market effect (HME). The second one is new and is dubbed here as the vote-market effect (VME) by analogy. Recall that in the standard model (viz. m=1) the HME is defined as $\partial\eta/\partial\Lambda>1$ whenever $\Lambda>1$. Likewise, when the two regions have the same size ($\Lambda=1$), the VME is defined as $\partial\eta/\partial(1/\theta)>1$ if, and only if, $1/\theta>1$.

---

[62] Note that the utilitarian optimum is the laissez-faire outcome. Indeed, the utilitarian planner puts equal weights on anybody's welfare as does the candidate facing population with identical ideological spreads. This is reminiscent of Result 2-12.

[63] The attentive reader might have noticed that apparently regional size does not matter in determining the equilibrium subsidy; we get back to this in Section 3.5.

In order to determine which dominates we define a third effect shortly: the net-market effect (NME). To this aim, we need to introduce a new variable,

$$\text{(3-20)} \qquad \Lambda_{Swing} \equiv \frac{\Lambda}{m} \equiv \frac{L_1 / L_2}{\psi / \gamma}$$

representing the overall relative force of the two regions. Particularly, Λ_{Swing} is the ratio of the mass of swing voters in 1 to the mass of swing voters in 2. Note that Λ_{Swing} can be lower or larger than unity. Using this definition into (3-19), it is easy to check the following result:

$$\text{(3-21)} \qquad \frac{\partial \eta_{NE}}{\partial \Lambda_{\text{Swing}}} = \frac{1-\phi^2}{(1-\phi\Lambda_{\text{Swing}})^2} > 1 \Leftrightarrow \Lambda_{\text{Swing}} > 1$$

In other words, we have:

> Result 3-5. (Net-market effect) At equilibrium the region that has more swing voters will get a more than proportionally larger share of industries.

To put it differently, in equilibrium, the big region (1) will end up attracting a more than proportional number of firms only if its economic strength (measured by its relative expenditure Λ) more than compensates its political weakness due to the higher dispersion of its population over the political dimension (1>m). Conversely the economically small region can attract a more than proportional number of firms if it has enough political power, i.e. if it has a sufficiently large mass of swing voters. Thus, when the VME is added to the basic model, predictions may differ from those induced by the standard economic HME and the political game may qualitatively reverse the laissez-faire outcome.

A final point deserves attention here. When trade barriers are sufficiently low, but still positive, our model too features a core-periphery outcome (unless $\Lambda_{\text{Swing}}=1$). In particular, all the economic activities concentrate in the large region ($\eta_{NE}=+\infty$) whenever $\phi>1/\Lambda_{\text{Swing}}$, while region 2 becomes the core ($\eta_{NE}=+\infty$) whenever $\phi>\Lambda_{\text{Swing}}$. Note that the two are mutually exclusive because $\phi\in[0,1)$.

The novelty of this analysis is that, when the political game is given the deserved attention, the definite prediction of the traditional models on the big region becoming the core does not necessarily hold anymore: it is not necessarily the large region that attracts all economic activities in the end.

In other words, the political environment does matter in shaping the equilibrium geography.

Having analysed the equilibrium location of economic activities, we can now turn to the analysis of the equilibrium subsidy level which delivers η_{NE}, namely θ_{NE} given by (3-18).

The Equilibrium Subsidy Level: Does Size Matter?

As expected, interior solutions for θ_{NE} are increasing in m – see (3-18). Candidates want to attract swing voters and the less dispersed group has a larger mass of such voters (ceteris paribus). Hence, the wider is the difference in the homogeneity degree of the two regions, the higher is the subsidy level the relatively more homogeneous region receives. Besides, θ_{NE} is larger than unity and, hence, region 2 is subsidised because $1>m$. This departure from a majoritarian result, which is due to the fact that regional policy is more salient to a minority of citizens, is one of the sources of non-majoritarian outcomes discussed in Besley and Coate (2000).

The attentive reader might have noticed that apparently regional size does not matter in determining the equilibrium subsidy θ_{NE}. Indeed, the economic weight Λ does not appear in (3-18) and the region that gets the subsidy is the more homogeneous one, independently from its size. As we shall see, this is clearly a knife-edge result that depends on the logarithmic transformation of the aggregate consumption index in the utility function U. More generally, the relative size of the two populations matter, and the effect of an increase of Λ on θ_{NE} is ambiguous.

This is best understood from the dual nature of Λ. For one, this represents the ratio of electorates and hence, since candidates try to get as many votes as possible, a larger Λ implies a lower θ, ceteris paribus. But Λ also represents the ratio of expenditures and, given (3-5), a larger Λ implies a larger η. Hence, for a constant political equilibrium η_{NE} in (3-19), a larger Λ must be compensated by a larger θ so that the solution to the economic relationship (3-5) is unchanged. In other words, if the result of the political process decides for some location equilibrium η_{NE} a larger subsidy θ will be needed do accomplish this if Λ is larger.

This point is easily made mathematically. Let us detangle the two natures of Λ and write Λ_V when we talk about electorate sizes (the subscript 'V' for voters) and Λ_C when we talk about economic sizes (the subscript 'C' for consumers). Hence, since (3-5) is an economic equilibrium relationship, we write:

(3-22)
$$\eta = \frac{(1-\theta\phi)\Lambda_C - \phi(\theta-\phi)}{-(1-\theta\phi)\Lambda_C\phi + (\theta-\phi)}$$

Conversely, the Λ's in Section 3.4 above clearly represent electorate sizes. Hence, we rewrite (3-19) as:

(3-23)
$$\eta_{NE} = \frac{\Lambda_V - m\phi}{m - \Lambda_V\phi}$$

Obviously, the two η's must be consistent, so θ_{NE} solves $\eta=\eta_{NE}$. Using (3-5), the solution to this problem is:

(3-24)
$$\theta_{NE} = 1 + (1-\phi)\frac{(m-1)\Lambda_C + m(\Lambda_C - \Lambda_V)}{(1+\phi m)\Lambda_C - (\Lambda_C - \Lambda_V)}$$

Obviously, the solution to (3-24) is identical to the solution to (3-18) if and only if there are as many electors as consumers (or, more generally, when the participation rates are the same in the two regions).

For reasons explained earlier, (3-24) reveals that:

> Result 3-6. The equilibrium subsidy to region 2 is increasing in Λ_C and decreasing in Λ_V.

Interestingly (3-23) tells us that only political variables matter in the determination of the η_{NE}. Indeed, η_{NE} is entirely determined by tastes, participation rates, and ideological heterogeneities. Conversely, (3-24) suggest that economic variables matter only for the determination of the level of the instrument needed to accomplish the equilibrium policy.

A final remark is in order here. Different participation rates have the same qualitative effects on the equilibrium location and subsidies than different ideological heterogeneities. To fix ideas take m=1 but imagine instead that in the larger region the participation rate is m' times lower than in the small region –possibly because there are larger foreign populations in big cities and foreigners are forbidden to vote by law. This assumption implies $\Lambda_V=\Lambda_C/m'$.

Then it is readily seen from both (3-23) and (3-24) that the solutions in (3-18) and (3-19) are identical, with Λ replaced by Λ_C and m replaced by m'. In short, we can either assume that large regions are socially more heterogeneous or that participation rates are lower there than in small regions. Both of these are reasonable assumptions that yield the same result: at the political equilibrium, small regions get a larger of industry than in the laissez-faire equilibrium and when trade costs are low enough it is possible that the industrial cluster end up in the economically disadvantaged region.

To summarize, the political equilibrium of this section game has removed the original definite prediction made in the original models like Krugman (1980) regarding the identity of the core. Taken at its face value, the analysis conducted in this chapter has shown that it is no longer obvious that the large, rich region or trading partner will eventually attract all the industrial activity.

3.6. Concluding comments

In the late 1950's, the per-capita income gap between Belgium's main regions, Flanders and Wallony, was particularly wide (see Bismans 1988). Flanders was behind and it was widely expected that it would take it more than 20 years to close the gap. In fact, it took only six years for Flanders' per capita income to reach the level of Wallony. What happened? Early in the 1960's, the 'Loi d'expansion régionale' (literally, the 'law of regional expansion') entered into force. This law was designed to attract investment, mainly to Flanders. Some of this investment was clearly diverted from Wallony and the law had the effect of accelerating the catch-up process. But why did this law get passed in the first place?

The explanation put forth by our model focuses on the difference in the political environments prevailing in the two regions. Wallony was deeply divided politically with a very conservative right, a strong left-wing party, and a tense class struggle. By contrast, the political parties and the unions in Flanders were more moderate. Moreover, Wallony experienced an influx of immigrants and, consequently, its society was more heterogeneous and variegated than the Flemish society. These differences still remain today, and now it is Wallony that is behind.

This historical example illustrates how political groups that focus only on a few issues are particularly attractive for politicians. In particular, we have argued that a more homogenous electoral base is better able to capture political gains through a voting process and we have applied this property to an economic geography issue showing that political forces can dramatically change the market outcome.

Using a very simple model of economic geography enriched to allow for endogenously determined regional policy, this chapter has analysed the impact regional policy on the spatial allocation of industry. In spite of the framework's simplicity, interesting results emerged. The relative size of the regions has an ambiguous effect on the equilibrium subsidy. On the one hand, if a larger fraction of the population –hence, of voters– lives in a given region, the equilibrium subsidy to the other region tends to be reduced as more voters favour a lower value of subsidies. On the other hand, due to the home-market effect, an increase in a region's size increases its equilibrium share of industry - and hence, its real income - for any given subsidy level. This aspect of the home market effect allows all political candidates to raise the subsidy to the other region without altering individuals' welfare. In essence, the very fact the economically big region is big means that its people are willing to accept larger 'real taxes' (in the form of a loss in their economic welfare). The net effect of the relative population size is thus ambiguous.

The effectiveness with which regional policy slows down the agglomeration process depends on the relative size of the two populations. For a given amount of regional aid, the regional policy is less effective in attracting industrial activity to the small region if the size-disparity of the two regions is larger, again due to the home-market effect. Thus, the political factor determines the amount of aid and the economic factor establishes its effectiveness. Indeed, if the small region is much more homogenous than the large region, politically determined regional policy may even reverse the spatial outcome predicted by orthodox economic geography theory. That is, the core can end up in the small region when trade/transport costs are sufficiently low since the agglomeration forces that favour the big region become very weak as trade gets freer, allowing the small region's political advantage to overcomes its economic weakness.

Even if it is difficult to find cases in which regional policy reversed the expected regional specialization patterns, our theoretical findings help exposit the internal logic of the complex interplay of openness, regional policy and the observed spatial allocation of industry. Poor and depressed regions may not attract all the economic activities and become the core in spite of the help they get in the form of regional aids as happened in the case of Flanders. Nevertheless we should expect countries (or constituencies) marked by prominent 'rural-versus-urban' divides to have lower levels of agglomeration at any point on the integration path, ceteris paribus. We note, however, that adding labour mobility –in the form of, say, a rural exodus– might increase the heterogeneity of urban areas (assuming it takes one generation for the newcomers to adapt fully to the new life-style). To the extent that this favours a more aggressive anti-agglomeration regional policy, such a mechanism would tend to favour a more spatially dispersed outcome for industry.

Chapter 4. GEOGRAPHY AND COMMUNICATION COSTS

4.1. Introduction

This chapter looks at how reductions in communication costs – together with variations in transportation costs – affect the spatial distribution of economic activity. 'Communication costs' are understood here as encompassing the cost of coordination and of conveying information between the head of the production unit and workers providing routinised back-office work. Unlike transportation costs proper communication costs are immaterial.

Low communication costs allow a firm to physically separate different activities that used to be performed in a single location. Jones and Kierzkowski (1990) dub this as the 'fragmentation' of vertically integrated production processes. The actual production of the final output is performed by the 'front office'; routinised 'back office' tasks involve include data entry or processing, database management. More elaborate back office tasks include financial and accounting services, processing assurance claims, and the development of computer software (Wilson, 1995).

This chapter analyses the spatial separation of front office and back office of the routinised kind. Duranton and Puga (2001) analyse the later kind of task unbundling.

Unbundling of back and front office production is widespread. Over 200,000 people are thought of as being engaged in back office work in the US. The example of Citibank comes to mind: its headquarters are based in New York and the data processing is carried out in South Dakota. A firm can even get multinational and outsource back office tasks. Such offshore back office is quantitatively not as important: production in the US employs about 30,000 offshore back office workers. In both cases the mechanisms are the same: firms commonly relocate back office production to rural or suburban areas where cheaper and more skilled part-time workers are available, using communication facilities to communicate with the management maintained in the more costly urban core or 'central business district'.

As Baldwin and Martin (1999) point out, 'on the microeconomic side of globalisation, FDI flows are the only thing that has changed sharply in the last twenty years (p. 17).' From 1985 on, foreign direct investment surged well above domestic investment: up to 1985, both series were fluctuating around 1980 numbers. Eleven years later, domestic investment has merely doubled, whilst FDI has been multiplied by a factor 6.5. Part of this pattern certainly comprises some offshore back office production –though this does provide only indirect support for the phenomenon this chapter aims to take account for. Duranton and Puga (2001) review some indirect evidence of another sort.[64] For one, firms increasingly outsource some tasks that used to be carried within the boundaries of the firm. Moreover, there are a growing number of multi-location firms in the US. This last trend presumably reflects the spatial allocation of different stages of the production process according to their comparative advantage (understood in a broad sense, including access to a variety of specific intermediates).

Some anecdotic evidence illustrates how low communication costs and offshore back office production are linked. In the seventies, some US firms would occasionally send some batch work to the Caribbean for processing (Wilson, 1995). The shipment would take two weeks each way. Nowadays, American Airlines assembles the accounting material and ticket coupons in Dallas; uses its own carriages to send it to Barbados where its subsidiary processes 800 thousands AA tickets a day; and sends the data by satellite to Tulsa. Clearly, for this to work both transportation and communication costs must be low (in the example: air cargo and satellite transmission, respectively) relative to the well-documented extra managing cost incurred by multi-plant firms.

The agents these multinationals deal with are their subsidiaries in the Caribbean, various Asian countries or Ireland, i.e. in countries that have undertaken huge investments in communication technology. AA and other firms go multinational in search of low production costs; in particular labour intensive tasks are conveyed in labour abundant countries. The literature on multinationals has identified other purposes for establishing plants in foreign countries (see Markusen 1995 for a survey). For instance, firms might establish foreign subsidiaries to serve the host market so as to avoid the trade barriers and transportation costs associated with producing the good at home and trade it (this is known as the 'jump the tariff argument').

---

[64] I came across this paper while writing the very final draft of this chapter. This is unfortunate because, though the analysis there is clearly relevant to this chapter, I had to disregard the connexions between at this late stage.

Globalisation of services is a recent phenomenon. Just as recent are the substantial falls in communication cost, with new opportunities for both offshore and suburban back office production. Over the post-WWII period, the transatlantic phone bill has been reduced by a factor 100 (see Figure 4-1). Also, the annual real cost of a telephone circuit in 1965 was $22,000. In 1980, this figure was down to $800 and in five years later further down to $30 only (Wilson 1995). Finally, current satellite communications cost less than 10% of what they used to do in the mid 1970's.

Data on quantities show that these downward trends for prices were matched by upward trends in quantities. In 1986, there were 0.1 millions transatlantic and 0.041 millions transpacific of voice paths, two figures that rose to respectively 2.022 and 1.889 millions in 1996 (Baldwin and Martin 1999). Over the same time span, the number of Internet hosts surged from .005 to 12.881 millions and is nowadays close to 30 millions.

Figure 4-1: Transportation vs. communication costs, 1940-1990

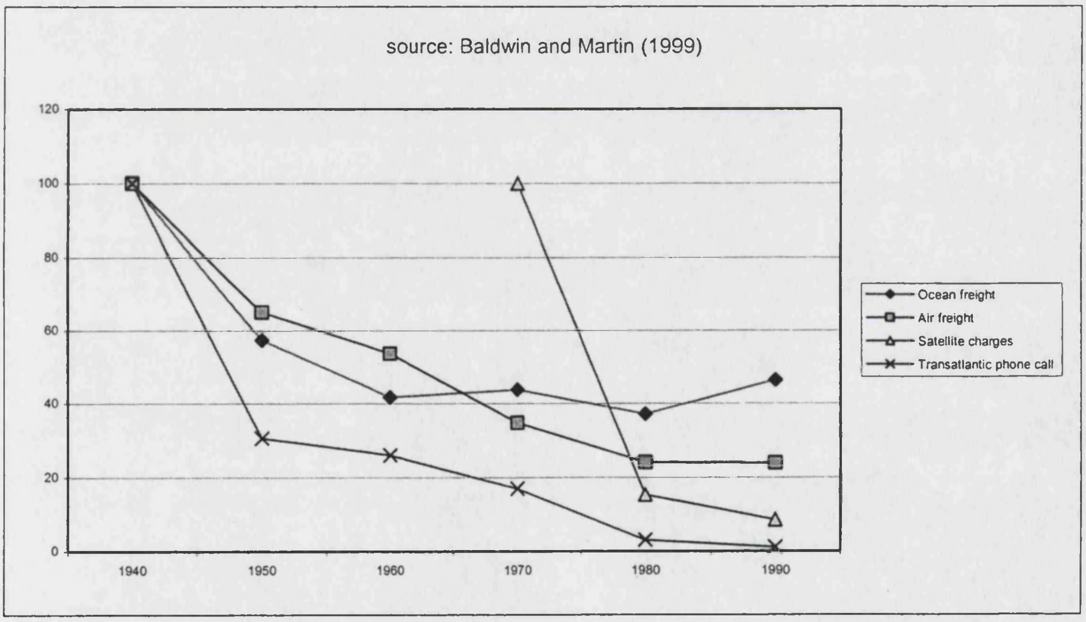


Figure 4-1 also strikes on the relative importance of the downward trend of the series on communication costs versus the series of trade and transportation costs. Transportation costs (over long distances at least) flattened out in the mid-eighties after the Second World War: air and ocean freight costs have decreased by about 40% between 1940 and 1960 – and a further 40% for the former until 1990 (the latter have merely stagnated since then).

Though of different magnitudes, the time patterns of transportation costs and trade barriers, on the one hand, and communication costs, on the other, point in the same direction: the world economy is ever more integrated. These costs are of different nature, however. As Harris (1995) puts it, communication facilitates coordination and the transfer of information, while transportation is a physical transfer of a good across space. How does this translate into firm's location decisions? How are geographical inequalities affected? What are the implications on factor rewards? What does the new economic geography (NEG) suggest these are?

In NEG models in the tradition of Fujita, Krugman and Venables (1999) all forms of costs are melted into one parameter, the iceberg transportation costs. The aim of this chapter is to disentangle communication costs from trade and transportation costs and assess how this framework conveys new insights to help understand the growing importance of these phenomena that have taken some importance recently –primarily the growing importance of unbundled back office work.

The remainder of the chapter is organised in the following way. The next section introduces the model (an extension of the FCVL model of Chapter 1). Sections 4.3 introduces communication costs to the model. Sections 4.4 and 4.5 convey the stability analysis in the extended model. Section 4.6 considers how various integration paths generate radically different specialisation patters whereas section 4.7 evaluate the welfare effects associated with these patterns. Finally, section 4.8 concludes. Some cumbersome algebra is relegated to an appendix.

4.2. The simple extended model

The model developed in the sequel extends the model of Chapter 1 in a simple way. The properties of the resulting model are similar to Fujita et al. (1999, section 14.4). For the time being, I assume that (what I dub as) communication costs are prohibitive, namely, the model is a standard NEG model.

Instantaneous equilibrium

The world is made of two regions or countries, j=1,2, identically endowed with the primary factors. Tastes are driven by (1-1). There are now three primary factors of production (and their respective returns): capital (π), labour (w) and land (r). As before, two goods are produced in this economy, A and M. The manufacturing sector M produces a differentiated good under increasing returns using capital, labour, and intermediates.

The novelty is that the production of the numéraire A now involves a constant returns to scale technology that makes use of both labour and land. I take a Cobb-Douglas functional form such that minimizing costs and perfect competition in the A-sector together yield:

(4-1) $$r_j^{1-\theta} w_j^{\theta} = p_A \equiv 1$$

where $0<\theta<1$.[65] Viewing land as a hidden factor, this is equivalent to saying that there are decreasing returns on L in the A-sector. The assumption $\theta<1$ implies that nominal wages are likely to diverge even if no country ever specializes in M. Hence, we now have to make the dependence of the endogenous variables on w_j explicit. The model of Chapter 1 is the limiting case in which there are constant returns in L, viz. $\theta=1$.

The expressions of Chapter 1 that need be changed are the following. First, monopolistic pricing yields the relation $p_j(1-1/\sigma)= \beta w_j^{1-\alpha} G_j^{\alpha}$ for a typical firm in j. (The term in the right-hand side is the marginal cost.) The normalisation $\beta=1-1/\sigma$ ensures that the following holds:

(4-2) $$\pi_j \equiv \frac{w_j^{1-\alpha} G_j^{\alpha} x_j}{\sigma}$$

As usual I define π as the average operating profit in the world economy, viz. $\pi=n\pi_1+(1-n)\pi_2$.

Second, write the implicit definitions of the price indices as:

(4-3) $$\Delta_1 = n\Delta_1^{\alpha} w_1^{(1-\sigma)(1-\alpha)} + \phi(1-n)\Delta_2^{\alpha} w_2^{(1-\sigma)(1-\alpha)}$$

The expression for Δ_2 is isomorphic.

Third, let E_j denote expenditure in j (we use capital letters so as to refer to the variable itself; in Chapter 1, we used small letters to refer to shares); the structure of the model implies:

(4-4) $$E_j = \mu \frac{Lw_j + \Lambda r_j + \pi}{2} + \alpha(\sigma-1)n\pi + \alpha(\sigma-1)n(\pi_j - \pi); \quad r_j = w_j^{-\theta/1-\theta}$$

---

[65] The parameter θ here is unrelated to the parameter θ in other chapters.

with j=1,2. The first term in the right-hand side of the expression above represents consumer expenditure; L, Λ, and 1 refer to the world endowment of labour, land, and capital, respectively; (4-1) implies $r_j = w_j^{-\theta/(1-\theta)}$. The last two terms refer to the expenditure that firms in j spend on manufactured inputs.

Finally, firms in 1 make the following operating profits:

(4-5)
$$\pi_1 = \frac{1}{\sigma} w_1^{(1-\sigma)(1-\alpha)} \Delta_1^{\alpha} \left(\frac{E_1}{\Delta_1} + \phi \frac{E_2}{\Delta_2} \right)$$

and π_2 is isomorphic.

To close the model, we need an expression for π, the equilibrium average value of operating profits. As before, the value of total output in sector M evaluated at producer prices must equal the value of aggregate expenditure spent on manufactured goods, viz. $np_1x_1+(1-n)p_2x_2=\alpha\beta[np_1x_1+(1-n)p_2x_2]+\mu\Sigma_j[w_jL+r_j\Lambda+\pi]/2$. Because the w's are no-longer constant when n varies, this implies that π is no longer invariant in n either. Alternatively, we can close the model using the labour full-employment conditions. Using (4-1), (4-2) and applying Sheppard's lemma to the cost function, these are:

(4-6)
$$\frac{L}{2} w_j = n_j(1-\alpha)(\sigma-1)\pi_j + \frac{\Lambda}{2} \frac{\theta}{(1-\theta)} w_j^{-\theta/1-\theta}, \quad j=1,2$$

where n_j=n if j=1 and n_j=1-n otherwise. The left-hand side of the expression above is the aggregate wage bill of the workers in country j; the first term in the right-hand side is the wage bill paid by the manufacturing sector; the second term is the wage bill paid by sector A. Note that, for a given π_j, w_j is increasing in n_j. The interpretation for this result is obvious: the more firms settle in j, the higher the demand for labour in 1. Since there are decreasing returns in L in sector A, the supply of labour to the manufacturing sector is imperfectly elastic. Hence, $\partial w_j/\partial n_j>0$ must hold, as claimed.

Together, (4-2)-(4-5) are the counterpart to (2-3)-(2-6) and define the instantaneous equilibrium of the model. Treating n as a parameter, we can solve the system for all factor prices. Then we ask which of these equilibria make sense at all 'in the long run'.

Long-run equilibrium

In the long run, capital is mobile and capital owners open plants where there profits are maximised. Whenever π_1 and π_2 differ, the adjustments follow the law of motion (2-7); accordingly, a long-run equilibrium is defined as an instantaneous equilibrium for which the following hold:

(4-7) $$n_j \geq 0, \quad \pi_j \leq \pi, \quad n_j(\pi_j - \pi) = 0$$

Namely, active firms make no pure profits.

The description of the basic model is now complete. I now briefly review the dispersion and agglomeration forces of the model.

Imperfectly elastic labour supply: a new dispersion force

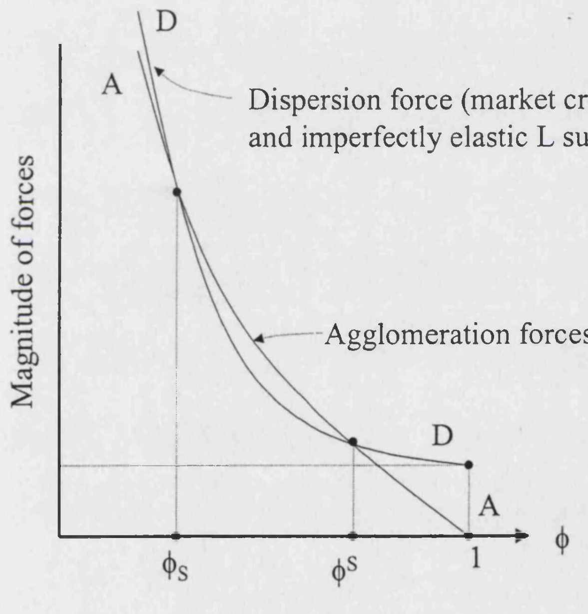
The basic model of this section is very similar to the model of Chapter 1. In particular, the agglomeration forces in both models are identical and stem from the vertical linkages that arise the result of firms buying each other's output as intermediate inputs. The market crowding dispersion force is also present in both models. Ceteris paribus, a firm that locates in a country in which most of its competitors are located has smaller market shares on each market. We saw in that chapter that the magnitude of all these forces is decreasing in trade free-ness ϕ and that the magnitude of the dispersion force decreases faster.

The present model displays yet another dispersion force, a force that stems from the decreasing returns in labour in agriculture. To repeat, these decreasing returns imply that the labour supply curve to M is strictly positive. In turn, this slope does not depend upon ϕ, but θ. This has dramatic consequences for the location equilibrium.

To illustrate the main effect of this new force on the location equilibrium, assume that ϕ is arbitrarily close to 1 and, without loss of generality, that most firms settle in 1, viz. $n>½$. Then we know from the discussion that follows (4-6) that $n>½$ implies $w_1>w_2$ or $\pi_1<\pi_2$ or both. But $n>½$ cannot be part of a long run equilibrium if profits are larger in 2, so it (4-6) implies $w_1>w_2$. Next, $\phi\approx1$ implies $\Delta_1\approx\Delta_2$. Using these and (4-5), $\pi_1\geq\pi_2$ implies $w_1\leq w_2$. This contradicts $w_1>w_2$. As a consequence, $\phi\approx1$ implies $n=½$. The reason is simple, but deep. When ϕ is close to unity, production costs between the two regions cannot differ by much because firms' location decisions are very sensitive to international differences in production costs. At the same time, the cost of intermediates is roughly the same across locations; hence the differences in production costs come from different labour wages must be small, too. With identical and strictly increasing labour supply in both countries, this implies that the labour employment in M must be identical in 1 and 2, too.

Turn to Figure 4-2. Like Figure 1-1, this figure plots the magnitude of the agglomeration and dispersion forces –the curves AA and DD, respectively. The nature of the former and, among the latter, of the market crowding force, is unchanged in the present setting. By contrast to Figure 1-1, Figure 4-2plots an additional dispersion force (the horizontal, dotted line), due to the imperfect elasticity of the labour supply. The more inelastic is the labour supply (viz. the lower θ), the higher is the horizontal dotted line. Figure 4-2 is plot for n=1, a 'core-periphery' pattern. A qualitatively similar figure would emerge for n=½.

Figure 4-2: Dispersion and agglomeration forces at the core-periphery equilibrium



Clearly, if the new dispersion force is not too strong, the curves AA and DD cross twice. For low values of ϕ, the market crowding force is strong enough for the AA curve to lie below the DD curve; the core-periphery outcome is not sustainable. For values of ϕ larger than ϕ_S, the vertical linkages are stronger than the market crowding force. But when ϕ is high enough (larger than ϕ^S on the figure), the strength of the vertical linkages is weaker than the strength of both dispersion forces taken together.

Hence, a core-periphery structure is sustainable only for intermediate values of ϕ. Summing-up:

Result 4-1. Decreasing returns in agriculture introduce a dispersion force that is independent of ϕ. As a result $n=\frac{1}{2}$ is the unique stable long run equilibrium of the model when ϕ is sufficiently close to unity.

To put it simply, in any configuration in which $n\neq\frac{1}{2}$ the presence of decreasing returns in labour in A implies $w_1\neq w_2$. When ϕ is arbitrarily close to 1, this implies that production costs cannot be different in 1 and 2. This in turn implies that $n\neq\frac{1}{2}$ cannot be part of a stable long-run equilibrium.

This brings us to the conditions under which $n=\frac{1}{2}$ is part of a stable long run equilibrium.

Break points

The 'dispersed equilibrium' ($n=\frac{1}{2}$) is always part of a long-run equilibrium by the symmetry of the model, but is not always stable as we saw in the earlier chapters. The break points are defined as the values of ϕ at which the symmetric equilibrium $n=\frac{1}{2}$ is just stable. In this context, 'stability' means that any exogenous shock to n is self-correcting by the law of motion (2-7). In the model of this chapter, there are two break points for ϕ in the unit interval provided that agglomeration forces are large enough relative to the dispersion forces. As we shall see in the analysis of the sustain point below, this means that the agglomeration forces net of the market-crowding force, as captured by a larger μ or a low σ, should be large enough relatively to the extend of decreasing returns in A, as captured by a low θ.

Observe first that $n=\frac{1}{2}$ yields $w_1=w_2$, $r_1=r_2$, $\Delta_1=\Delta_2$ and $\pi_1=\pi_2$. Again, I denote these variables with the nought subscript. Substituting these into (4-3)-(4-6), we can solve for w_0, r_0, Δ_0, and π_0. We find:

(4-8) $$w_0 = \sqrt{\frac{\Lambda \varepsilon_0}{L}}, \quad \pi_0 = \frac{1}{\varepsilon_0 w_0} \frac{\Lambda(\varepsilon_0 - 1)}{(\sigma - 1)(1-\alpha)}$$

Following standard practice in the NEG, I choose units for L and Λ such that $w_0=r_0=\pi_0=1$ at $n=\frac{1}{2}$. From (4-3), it is then easy to see that $\Delta_0^{1-\alpha}=(1+\phi)/2$.

As we shall see, the break points play a minor role in the analysis of this chapter, so I relegate the formal analysis to the appendix of this chapter.

Sustain points

We now turn to the conceptually different issue of the stability of the core-periphery patterns. A long-run core-periphery pattern (n=0 or n=1) is said to be *sustainable* if and only if (4-7) is satisfied at n=1, namely $\pi_1 \geq \pi_2$ and n=1 hold simultaneously. In words, this requires the following: given that all firms are in 1, no individual entrepreneur has any incentive to deviate and start producing in 2. A single-equation implicit solution for the sustain point like (2-10) is no longer available for the general case $\theta<1$. However, the discussion on Figure 4-2 above permits us to assess that there are two sustain points (if any) in the interval (0,1). These are the real roots of a non-integer polynomial.

Two limiting cases are of interest. If $\theta=1$ there are no DRS in agriculture and we know from Chapter 1 that the largest of these roots limits to unity whereas the smaller root –denoted as ϕ^{sust} in Chapters **2** and **3**– is strictly in (0,1). At the other extreme $\theta=0$ in which case agriculture uses land only. As a result labour is supplied inelastically to the manufacturing sector in each region. This in turn implies that if no firm establishes itself in region 2, say, then wages are zero there. This cannot be part of a long-run equilibrium because the shadow-profit π_2 limits to infinity in this case.[66] As a consequence none of the roots belong to (0,1] in this case.

More generally, when $\theta<1$ both roots are smaller than unity if they are real. For the sake of illustration, I develop the case $\theta=\frac{1}{2}$.

Using (4-6), it is easy to see that n=1 implies $w_2=(\Lambda/L)^{\frac{1}{2}}$. From (4-3) we find $\Delta_1=w_1^{\ 1-\sigma}$ and $\Delta_2=\phi\Delta_1$. Plugging this and the previous result in (4-5) and (4-6), and making use of (4-4), we find that the ratio of nominal wages when all firms locate in 1 is equal to:

(4-9) $$\left.\frac{w_1}{w_2}\right|_{n=1} = \frac{1+\psi}{1-\psi} \equiv \varepsilon_0, \quad \psi \equiv \frac{\mu(1-\alpha)(\sigma-1)}{(1-\alpha)(\sigma-1)+1-\mu} < 1$$

---

[66] That is, the marginal productivity of labour is infinite when employment is nil. Since the nominal wage is zero in this case and because the elasticity of demand on the product market is larger than 1, profits are infinite.

Clearly, w_1 is larger than w_2 as labour is scarcer in the country in which the manufacturing industry clusters. Making use of (4-5) again, it is easy to see that the ratio of (shadow) profits in 2 to profits prevailing in 1 is equal to:

(4-10) $$\left.\frac{\pi_2}{\pi_1}\right|_{n=1} = \varepsilon_0^{(\sigma-1)(1-\alpha)}\phi^{\alpha}\left(s_E^1\phi + \frac{1-s_E^1}{\phi}\right)$$

where

(4-11) $$s_E^1 = \frac{1}{2}\left(1+\alpha\beta+\beta(1-\alpha)\mu\psi\right)$$

The s_E's represent the expenditure shares for n=1.Obviously, $\frac{1}{2}<s_E^1<1$ holds in any admissible parameter configuration. Clearly, the core-periphery pattern is 'sustainable' when the expression in (4-10) is less than one.

The interpretation of (4-10) is the following. The first term in the right-hand side, $\varepsilon_0^{(1-\sigma)(1-\alpha)}$, is larger than unity and represents the cost saving realised in the low-wage country 2.[67] By contrast, the term ϕ^{α} represents the extra-cost paid intermediates, all of which must be imported from the industrial core (country 1). Finally, the term in the parenthesis captures both the extra cost of serving market 1 from abroad (s_E^1 is deflated since $\phi<1$) and the gain from serving market 2 from within (s_E^2 is magnified since $1/\phi>1$).

We can now state that the core-periphery configuration is sustainable for any ϕ in $[\phi_S,\phi^S]$, where ϕ_S and ϕ^S are the roots of $h(\phi)=0$, where:

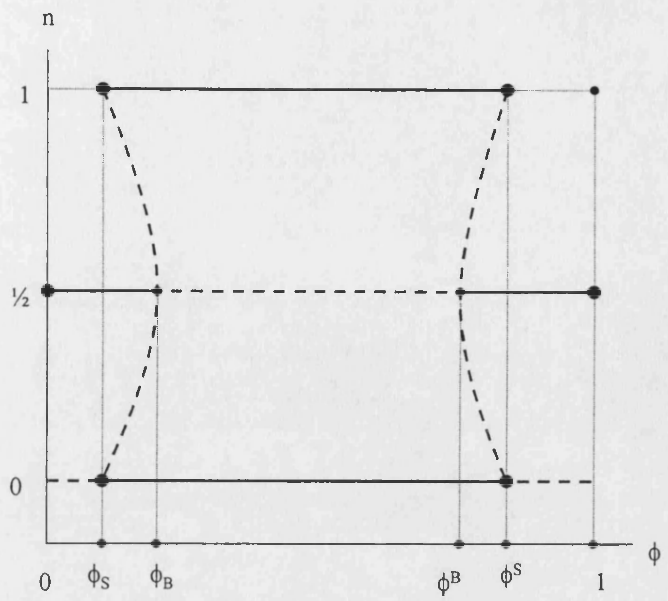
(4-12) $$h(\phi) \equiv \varepsilon_0^{(1-\sigma)(1-\alpha)}\phi^{1-\alpha} - s_E^1\phi^2 - (1-s_E^1)$$

In other words, $\phi\in\{\phi_S,\phi^S\}$ implies that the expression in (4-10) is equal to unity, viz. $\pi_2=\pi_1$.Trivial algebra shows that the polynomial on the left-hand side of the expression above is negative and increasing at $\phi=0$, negative and decreasing at $\phi=1$, and concave everywhere. Hence, h(.) admits a unique maximum. If agglomeration forces are too weak, this maximum is negative (in such a case the curves AA and DD in Figure 4-2 do not intercept). If agglomeration forces are strong enough (μ large and σ low), the two roots of (4-12), namely ϕ_S and ϕ^S, belong strictly to the (0,1) interval, in which case $\partial h/\partial\phi>0$ at ϕ_S and $\partial h/\partial\phi>0$ at ϕ^S. Otherwise these roots are not real.

---

[67] ε_0 limits to infinity in the case $\theta=0$.

Figure 4-3: The 'Double Tomahawk' diagram



Ranking the break and sustain points

We saw in Chapter 1 that the ordering of the break and sustain points was crucial for the dynamics of the model. $\theta=1$ in the model there and as a result there are a unique break point and a unique sustain point in the interval [0,1). We saw that the sustain point came before the break point (see the 'Tomahawk diagram'), which implies two facts. First, no interior, asymmetric equilibrium is ever stable. Second, there is hyseresis in location, in a sense we made precise there. The same holds true here: at least when $\alpha=\mu$ (the case Krugman and Venables (1995), among others, assume) simulations show that the sustain interval $[\phi_S,\phi^S]$ encompasses the break interval $[\phi_B,\phi^B]$, hence $\phi_S<\phi_B<\phi^B<\phi^S$, as shown in Figure 4-3.[68]

This figure plots n against ϕ. As usual, the stable, long-run equilibria of the system are depicted in plain lines whereas the unstable ones are depicted in doted schedules. There is room for hysteresis in this model in the sense that if the system finds itself at n=1 when ϕ is, say, larger than ϕ_B, then this remains a stable long run

---

[68] Details of the calculations can be found on the Maple worksheet vertical_linkages_0523.mws, available upon request. When $\alpha\neq\mu$, the parameter space has many dimensions, hence making sure that the ranking remains unchanged for all parameter combinations is a formidable task. For this reason, I put it aside and *assume* that the combinations of the parameters α, μ, and σ is such that the ranking holds so that Figure 4-3 is always relevant.

equilibrium if ϕ decreases to some value in (ϕ_S,ϕ_B). Conversely, if the system finds itself at n=½ and ϕ is lower than ϕ_S, then this remains a stable long run equilibrium if ϕ increases to the same value in (ϕ_S,ϕ_B). Hence history matters.

4.3. Communication costs and 'multinationals'

I now depart from the standard model. Remember that the analysis thus far essentially assumes that trade in M is costly whilst trade in A is not. Also, capital is the unique factor that can move freely from one country to the other (in the long run), whilst migration of L is prohibitive. The first is merely an assumption of convenience.[69] Removing the second assumption would complicate the analysis because it would require two laws of motion – one for K, one for L. More importantly, labour is not really mobile, especially internationally. However, thanks to good communication infrastructures it is now possible to fragment the production process and different services to the firm can be undertaken in different locations.

The idea in the remainder of this chapter is to disentangle the effects of falling transportation costs (rising ϕ) from those of falling communication costs. The former affects trade in goods (intermediate inputs included) whilst the later affects trade in business services. Presumably, the qualitative effect of these two related phenomena on the location equilibrium are distinct. This is the object of study of this section.

To capture that idea, I now assume that a firm in country 1 can hire some workers living in country 2 (and hence become 'multinational') and incurs a 'communication cost' in doing so (this captures managerial tasks and coordination costs). This firm still manufactures the good in 1, but some services are undertaken by workers abroad. These workers can be thought of as independent subcontractors who sell their services (in which case the firm is not a proper multinational) or as working in a foreign subsidiary. However, the issue of the boundary of the firm is beyond the scope of this chapter.[70]

The possibility to hire workers abroad has an important implication for the dispersion and agglomeration forces. When a firm moves from 2 to 1 (i.e. dn>0) this has both a spatial 'expenditure shifting' effect and a 'cost shifting' effect, as we saw in

---

[69] See Davis (1998) and Fujita *et al.* (1999, chapter 7) for a discussion.
[70] See e.g. Grossman and Helpman (2002) for a framework where the structure of the firm changes endogenously with the market structure.

Chapter 1. This firm now buys more of its intermediates and sells more of its output in 1, which increases relative profits of existing firms in 1. The former effect arises because firms price their output above marginal cost; the latter effect works by reducing other firms' costs.

By contrast, when an entrepreneur in 1 hires a worker in 2, this does not imply an expenditure shifting in this well-defined sense because workers are paid at their marginal product. However, nominal wages in 2 must increase as a result of this firm going multinational and this has a general equilibrium effect on firms' profits.

Modelling communication costs

Take a firm in 1. In the model as described in section 4.2, the implicit assumption is that this firm and the workers it employs communicate freely. By contrast, prohibitive inter-regionally communication costs prevent this firm from hiring a worker in region 2 for doing some 'back-office' job.

This assumption is now relaxed as follows. Any firm in 1 can hire any worker in 1 at unit cost of w_1 and any worker in 2 at unit cost of εw_2, with $\varepsilon \geq 1$ quantifying the magnitude of communication costs. In this spirit, workers are seen as providing a service to the firm. Communication facilities decrease coordination costs between the worker in 2 and the firm in 1. Hence, the manufacturing final output still takes place in country 1, as in section 4.2.

Unlike transportation costs, communicating from one spatial entity to another involves a fixed – as opposed to variable – cost. The nature of this cost can be seen as the time spent on coordination; the more numerous the workers a firm hires abroad, the more time is lost communicating. I assume here that a worker who supplies one unit of L spends $1/\varepsilon$ of his time productively and $1-1/\varepsilon$ of it communicating and coordinating with the entrepreneur. This time lost must be paid at the ongoing wage; hence the communication cost $(\varepsilon-1)$ is multiplied by w_2.[71] Workers hired by an entrepreneur abroad are sometimes referred to as *external* workers below. As a consequence, the cost function (1-2) to firms in 1 has to be rewritten as:

---

[71] Within a region, ε is meant equal unity; across regions ε is assumed to be larger than 1. The key assumption here is that inter-regional communication between agents is more time-consuming than within-region communication. Assuming positive communication costs in the latter case as well would essentially involve a re-scaling of the labour-output coefficient β and/or of the entire labour force L.

(4-13) $$C_1(x_1)=\pi_1+\beta x_1(w_1')^{1-\alpha}G_1^{\alpha};\quad w_1'\equiv\min\{w_1,\varepsilon w_2\}$$

$C_2(x_2)$ is isomorphic. It is immediate that the very (potential) existence of external workers reduces (potential) international nominal wage differences. It is also obvious that if some firms in 1 hire workers in 2, no firm in 2 will ever hire a worker in 1.[72]

I now modify the model of section 4.2 assuming that if any entrepreneur ever hires an external worker, only entrepreneurs in 1 would do so. Accordingly, I concentrate on the case $n\geq½$. This is without loss of generality given the symmetric nature of the model. It happens that the full-employment of labour conditions (4-6) alone have to be modified to incorporate this extension of the model. We also need the no-arbitrage condition $\varepsilon w_2\geq w_1$, as labour is homogenous. Denote m as the proportion of workers firms in 1 hire abroad. Loosely speaking, m can be referred to as the proportion of multinational firms. Given this, the following must hold at all times:

(4-14) $$w_1\leq\varepsilon w_2,\quad m\geq0,\quad (w_1-\varepsilon w_2)m=0$$

The interpretation of the expression above goes as follows. There are either no external workers (the case of section 4.2), and in such a case wages in 1 are below wages in 2 adjusted for communication costs, or wages in 1 have a propensity of being larger than adjusted wages of external workers, and hence it is profitable for some entrepreneurs in 1 to hire workers in 2.

To get an idea of what the possibility of hiring workers from abroad has on the instantaneous and long run equilibria, turn to Figure 4-4. This figure plots the instantaneous equilibrium for some $\phi\notin[\phi_S,\phi^S]$. The horizontal axis measures the proportion of firms established in region 1, n.[73] The vertical axis measures the difference in operating profits between firms operating in 1 and firms operating in 2. In the configuration depicted by the plain schedule the core-periphery outcome is not sustainable ($\pi_1-\pi_2<0$ at n=1) and the dispersed configuration is stable, viz. $d(\pi_1-\pi_2)/dn<0$ at n=½. In other words, the degree of trade free-ness is either low or high, viz either $\phi<\phi_S$ or $\phi>\phi^S$.

The plain schedule represents a typical configuration of section 4.2. In this case, communication costs are prohibitive and firms hire workers on their local labour market

---

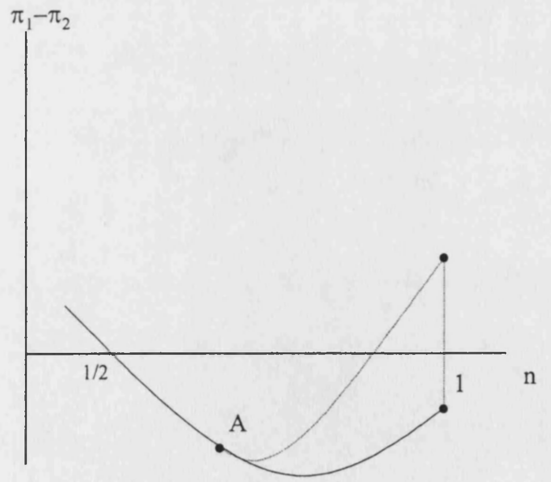
[72] Indeed, a firm in 1 hires a worker in 2 if $\varepsilon w_2\leq w_1$. Conversely, a firm in 2 would hire a worker in 1 if $w_2\geq\varepsilon w_1$. With $\varepsilon>1$, these two conditions are mutually exclusive.

[73] By the symmetry of the model we can disregard the interval [0,½).

only. As we know, employment in the manufacturing sector in region 1 increases when the proportion of firms established in 1 increases. This implies that wages in 1 increase as n increases. Now allow firms to hire workers abroad as well. As can be seen from (4-14), if ε is not too large firms in 1 will start hiring workers in 2 when the wage gap increases beyond some threshold. This happens only if n is substantially larger than ½ (indeed, by the symmetry of the model $w_1 \approx w_2$ when $n \approx ½$).On the graph, this happens from point A onwards. The ability to hire cheap labour increases the profitability of firms in 1 and hence the curve π_1-π_2 rotates anti-clockwise around point A. In the situation depicted, the core-periphery outcome n=1 becomes sustainable. Moreover, m increases as n increases. To sum-up, we can write:

> Result 4-2. The ability of firms established in region 1 to hire workers from region 2 has three effects. First, it potentially increases the relative profitability for firms in 1. Second, m is non-decreasing in n. Finally, a core-periphery pattern that was unsustainable when communication costs were prohibitive might become sustainable as the result of low transportation costs.

Figure 4-4. Hiring workers abroad.



Note, however, that the stability properties of the model are unaffected in the neighbourhood of n=½. Moreover, even when ε is so low that firms in 1 hire workers in 2, we have:

> Result 4-3. Employment in 1 (2) rises (decreases) as n increases. As a result nominal wages in 1 (2) increase (decrease) as n increases. However, this relative effect on employment and wages is reduced as communication costs ε fall.

I formalise all these claims in the sequel.

To close the model, we need to modify the full-employment conditions so as to reflect the possibility that firms in 1 hire workers in 2. Using the free-entry condition (4-2) and the no-arbitrage condition (4-14), full-employment in 1 now requires

(4-15) $$(1-m)n(1-\alpha)(\sigma-1)\pi_1 + \frac{\Lambda}{2}\frac{\theta}{(1-\theta)}w_1^{-\theta/1-\theta} = \frac{L}{2}w_1$$

to hold. Accordingly, full-employment in 2 must now be rewritten as:

(4-16) $$mn(1-\alpha)(\sigma-1)\pi_1 + (1-n)(1-\alpha)(\sigma-1)\pi_2 + \frac{\Lambda}{2}\frac{\theta}{(1-\theta)}w_2^{-\theta/1-\theta} = \frac{L}{2}w_2$$

The first term in the left-hand side in the expression above represent the aggregate wage bill paid to workers in 2 by the multinationals; there are m times n such firms. The second term captures the wage bill of domestic firms. The third term is the wage bill of sector A. All three together must be equal to the total wage bill in this county, the term in the right-hand side of (4-16).

Note that the current model encompasses the model of section 4.2: set m=0 and ε arbitrarily large; the former implies that (4-15) and (4-16) collapse to (4-6) the later means that the no-arbitrage condition is never binding.

We now repeat the stability analysis in the extended model.

4.4. Preliminary results and break point

In this section we characterize some aspects of long run equilibria such that n≥½. Start with the symmetric equilibrium n=½. This is always part of a long run equilibrium given the symmetry of the model. Observe that nominal wages are equal at the symmetric equilibrium, independently of the magnitude of the trade and communication costs: hence n=½ implies m=0, all ϕ and ε. By continuity, it must be that there are no external workers in the neighbourhood of the symmetric equilibrium. Therefore the analysis for the break point in Chapter 1 remains valid provided that ε is strictly larger than unity. Note however that the no-arbitrage condition might be binding at interior, asymmetric equilibria. Such equilibria are always unstable in this model.[74]

Now take n to be 'significantly' larger than ½ so that the no arbitrage condition in (4-14) is binding. It should be obvious that the existence of communication costs implies that most workers who are employed by firms located in 1 are themselves in 1, namely the proportional of 'multinationals' in 1 is smaller than ½ at any long run equilibrium, viz. $m\in[0,½)$. To see this, take the difference of (4-15) and (4-16) and manipulate the terms to get:

(4-17)
$$\begin{aligned}
0 &< \frac{\Lambda}{2}\frac{\theta}{(1-\theta)}w_2^{-\theta/1-\theta}\left(1-\varepsilon^{-\theta/1-\theta}\right)+\frac{L}{2}w_2(\varepsilon-1)\\
&=(1-\alpha)(\sigma-1)[(1-2m)n\pi_1-(1-n)\pi_2]\\
&=(1-\alpha)(\sigma-1)[2n(1-m)-1]\pi_1+(1-\alpha)(\sigma-1)(1-n)(\pi_1-\pi_2)\\
&=(1-\alpha)(\sigma-1)2n[(1-\frac{1}{2n})-m]\pi
\end{aligned}$$

where the inequality stems from the parameter restrictions –in particular $\varepsilon>1$. The first equality follows directly from (4-15) and (4-16). The second equality is just a rearrangement of the previous line. The final equality stems from (4-7) and the definition of π; also, in any long run equilibrium n>0 implies $\pi_1=\pi$; in turn this implies $m<1-1/(2n)$. Since n>½ by assumption, it follows that m<½, as was to be shown.

To get further intuition for this result, assume that all entrepreneurs are in 1 and that half of them hire workers in 1 and the other half hire workers in 2, viz. m=½. As they are all located in the same region, each firm equilibrium size is identical, and hence all entrepreneurs hire the same number (mass) of workers. At equilibrium, they must all pay the same gross wage ($w_1=\varepsilon w_2$). Therefore nominal wages in 2 are lower than in 1. On the other hand, the labour force employed in sector A in each region is the same by the assumption n=1 and m=½. Because there are decreasing returns in labour in this sector, nominal wages must be the same in both regions ($w_1=w_2$), a contradiction if $\varepsilon>1$. In short, we have:

---

[74] This need not be the case if the two countries were not ex-ante identical.

Result 4-4. In the symmetric equilibrium $n=\frac{1}{2}$ costs in both regions are identical, so $m=0$. In any long-run equilibrium such than $n>\frac{1}{2}$ the proportion of workers joining the manufacturing sector is larger in 1, viz. $n(1-m)>mn+1-n$. This in turn implies that less than half of the firms in 1 hire workers from 2, viz. $m<\frac{1}{2}$.

As we saw in section 4.2, a core-periphery pattern n=1 can sometimes emerge as a 'sustainable' long run equilibrium. We now analyse such a configuration when communication costs are non-prohibitive.

4.5. Sustain points

No simple expression exists for the sustain points in the general case $0<\theta<1$, so I impose $\theta=\frac{1}{2}$ as in section 4.2.

To ask whether a core-periphery pattern is sustainable, we set n=1 but impose nothing on m, the proportion of 'multinationals' and, as usual, check whether the shadow profit in 2 is indeed lower than $\pi_1=\pi$. If this is indeed the case the agglomerated equilibrium is stable. If this condition were violated, however, any capital owner would gain by closing her plant in 1 and relocate production in 2. Before proceeding further, note that when communication costs are large enough the solution to the problem is given by

(4-11). In such a case, the difference in nominal wages in each region is maximal and is given by (4-9). Therefore, the no-arbitrage condition is binding only when $\varepsilon<\varepsilon_0$. From now on I assume this is the case, unless otherwise specified. Hence, some firms in 1 will hire workers in 2. Since these workers earn less, it is now less costly to produce in 1. In other words, the dispersion force of producing in the high wage country is weakened by the possibility of going multinational. This has a dramatic effect on the dynamics of the system, and in particular on the sustainability of the core-periphery pattern, as we now show.

The system given by the expenditure definition (4-4), the operating profit (4-5), and the full-employment conditions (4-15) and (4-16) can be solved for the nominal wages, the equilibrium operating profit, and the proportion of 'multinationals'. Using the normalisations that yield $\pi_0=1$ and $w_0=1$ in (4-8), this straightforward exercise gives us the following expressions:

(4-18) $$w_1(\varepsilon)=\sqrt{\varepsilon'}, w_2(\varepsilon)=\frac{1}{\sqrt{\varepsilon'}}, m(\varepsilon)=\frac{(\varepsilon_0-\varepsilon')}{(1+\varepsilon')(\varepsilon_0-1)}, \pi(\varepsilon)=\frac{(1+\varepsilon')}{2\sqrt{\varepsilon'}}$$

where $\varepsilon'\equiv\min\{\varepsilon,\varepsilon_0\}$ and ε_0 is given by (4-9). Standard comparative statics confirm that w_2 (w_1) decrease (increase) with ε, as lower communication costs are synonymous to greater manufacturing employment in 2 (lower manufacturing employment in 1). For the same reason, m decreases with ε. Obviously m=0 when communication are equal to or larger than ε_0 since, in such a case, the discrepancy in nominal labour costs does not compensate for the high transportation costs; by contrast, half of the manufacturing workers are hired from 2 at the limit $\varepsilon=1$ (no communication costs), as is obvious from the expression above.

Equilibrium profits are decreasing in wage costs as is to be expected. However, (4-18) also shows that equilibrium profits are increasing in ε which means that, despite the cost-reducing nature of lower communication costs, capital owner loose when ε decreases. This is not a robust result, for it is the outcome of various countervailing forces. Note first that, with Cobb-Douglas preferences and variable costs a constant fraction of income (μ) and costs (α) is spent on M. With aggregate price elasticity larger than one, these fractions would decrease with ε and as a result π would have a tendency increase as ε falls.

Second, since all firms are located in the same country by assumption, they have the same market shares on both markets 1 and 2 and have the same cost structure. Together with Cobb-Douglas tastes and technology, this implies $\pi=E^w/\sigma$, i.e. operating profits is proportional to the world expenditure on M and does not depend directly on costs at equilibrium. This can be derived using (4-3) and (4-5) using n=1.

Finally, the presence of decreasing returns in agriculture implies that $Lw+\Lambda r$ in (4-4) is a convex function of w, hence E_j is maximised for corner solution employment patterns, namely, when most people work in sector M. Since we convey the analysis for n=1, E^w is maximised for m=0. As a consequence, E^w is increasing in communication costs ε. By opposition, the E_j's are minimal when n=½. This is a general equilibrium effect whose magnitude is small if the sector M's share of GDP is small.

With all this at hand, we follow the same strategy as in section 4.2, namely, we solve for the shadow value of π_2 using (4-3) and (4-5) and ask under which conditions this is smaller than $\pi_1=\pi$ in (4-18). When this is so, the core-periphery equilibrium described by n=1 and (4-18) is sustainable. Computations show:

(4-19) $$\left.\frac{\pi_2}{\pi_1}\right|_{n=1} = \varepsilon^{(\sigma-1)(1-\alpha)}\phi^{\alpha}\left(s_E\phi+\frac{1-s_E}{\phi}\right)$$

where

(4-20) $$s_E=\frac{1}{2}\left(1+\alpha\beta+(1-\alpha)\beta\mu\frac{\varepsilon-1}{\varepsilon+1}\right)$$

Here, s_E is as defined E_1/E^w for n=1 and $\varepsilon<\varepsilon_0$. The expression above is the counterpart to (4-10) and (4-11). Clearly, s_E tends towards something larger than ½ from above as $\varepsilon\to1$. This is obvious: as ε tends towards 1, incomes are equalized since all factor rewards are; however, all firms are clustered in 1 and so all intermediates are sold and bought there, hence expenditure in 1 is always larger than expenditure in 2. Also, $½<s_E$ from (4-11) at the limit $\varepsilon=\varepsilon_0$, which follows from the fact that any communication cost larger than ε_0 is prohibitive. More generally, $\partial s_E/\partial\varepsilon>0$ holds for $\varepsilon\in(1,\varepsilon_0]$. The intuition for this result is as follows. Factor prices diverge less between countries when ε is low; this may or may not increase s_E. Because of the production technologies we have assumed –both sectors A and M use L (the former with decreasing returns) and the other factors (Λ and K) are specific– $1-s_E$ increases in the share of workers in 2 that are employed in the manufacturing sector. By (4-18) this is decreasing in ε, hence $\partial s_E/\partial\varepsilon>0$.

With non-prohibitive communication costs, the core-periphery outcome is sustainable if $\phi\in[\phi_S,\phi^S]$, where ϕ_S and ϕ^S are the roots of $f(\phi)=0$, where:

(4-21) $$f(\phi,\varepsilon)\equiv\varepsilon^{(1-\sigma)(1-\alpha)}\phi^{1-\alpha}-s_E\phi^2-(1-s_E)$$

Accordingly, $\phi\in\{\phi_S,\phi^S\}$ implies that the expression in (4-19) is equal to unity, or $\pi_2=\pi_1\equiv\pi$. The expressions for $f(.,\varepsilon)$ and h(.) in (4-12) are similar, so there is no need to repeat the comparative statics for μ, α, and σ. Note however that f, like h, is negative and increasing at $\phi=0$; negative and decreasing at $\phi=1$; and concave everywhere. Hence, $f(.,\varepsilon)$ admits a unique maximum. Unless otherwise specified, we assume that agglomeration forces are strong enough so that this maximum is positive and the roots of $f(.,\varepsilon)$ are in (0,1). As a consequence, $\partial f/\partial\phi>0$ at ϕ_S and $\partial f/\partial\phi<0$ at ϕ^S.

Communication costs and sustainability of the core-periphery equilibrium

We are now interested in the effects of ε on the sustainability of the core-periphery equilibrium keeping other parameter values constant. The effect of a fall of communication costs, viz. $d\varepsilon<0$, on f(.) is a-priori ambiguous. This is because $d\varepsilon<0$ has two effects. On the one hand, when transportation costs are low it is less expensive to produce in 1, as it is now possible to hire the cheap labour available in 2; this is the direct effect, as captured by the first term in the right-hand side of (4-21). Put another way, when firms can hire workers from any region, a dispersion force is weakened (cost effect): despite the existence of decreasing returns in A, w_1 and w_2 can no longer diverge without bound.

On the other hand, $1-s_E$ is decreasing in ε, hence 2 grows richer relative to 1 as communication costs decrease; by the home-market effect, this makes 2 a relatively better (or less bad) place to locate production. This indirect, general equilibrium effect is best seen using (4-20). When ε decreases, the rewards in 1 and 2 are more alike (and hence their incomes), and so $1-s_E$ increases and s_E decreases. This weakens an agglomeration force (the backward linkage).

Since both an agglomeration force and a dispersion force weaken as ε decreases, the net result of reducing communication costs is ambiguous. Of course, if sector M is relatively small relative to the aggregate economy, viz. $\mu\approx 0$, the indirect effect is negligible and hence the direct effect dominates. Simulations suggest that this is always the case, but since (4-21) is a non-integer polynomial the analytical proof is certainly extremely involved.[75] More precisely, simulations show that $\partial\phi_S/\partial\varepsilon>0$ and $\partial\phi^S/\partial\varepsilon<0$ whenever $0<\phi_S,\phi^S<1$. Following standard practice in the NEG, I am content with the simulations. The implication for this result is that the sustain interval $[\phi_S,\phi^S]$ increases when ε decreases. Namely,

Result 4-5. When communication costs decrease, the core-periphery equilibrium is sustainable over a wider range of trade costs.

---

[75] The Maple worksheet comm._costs_0613.mws contains the simulations and is available upon request.

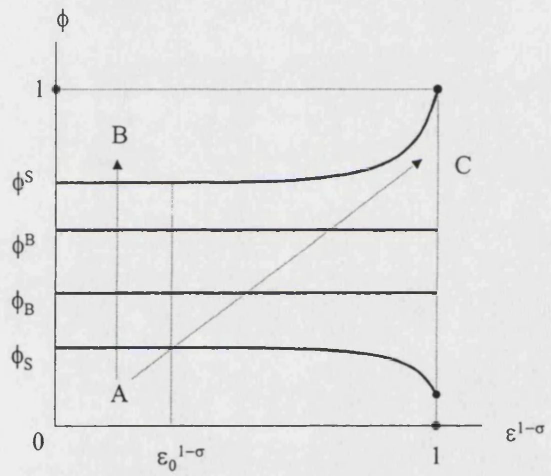
The case of zero communication costs

When communication costs are nil (viz. ε=1), firms can hire workers from any region at no additional cost. Hence wages must be identical or else there would exist arbitrage opportunities. In such conditions, if all firms cluster in the same region (e.g. n=1), exactly one half of the workers employed in the manufacturing sector are hired in 1, that is, m=½. As usual, (4-3) and (4-5) imply $E^w=\sigma\pi$. Using n=1 and m=½ in (4-15) and (4-16) and the former equality implies that the sustain point solves:

(4-22) $$2\phi^{1-\alpha} - [1+\alpha\beta]\phi^2 - [1-\alpha\beta] = 0$$

of which the largest real root is unity and the smallest one, which we defined as ϕ^{sust}, is strictly in the (0,1) interval.

Figure 4-5. The 'Loudspeaker' diagram



Three important facts are of interest here. First, this expression holds regardless of the functional form (4-1) chosen for the production function of sector A.[76] As an immediate corollary, (4-22) does not depend on θ. And, second, this expression for the sustain point is exactly identical to the expression derived in Chapter 1, viz. $\phi_S=\phi^{sust}$. The interpretation of this remarkable result is rather intuitive. The labour supply to sector M is perfectly elastic in both 1 and 2 in Chapter 1. That makes one dispersion

---

[76] Of course, all other functional forms still matter.

force wash away. Here, the possibility to hire workers regardless of where they live is a substitute for this at the margin.

Finally, the analysis here and in the previous subsection imply that the sustain interval looks like depicted in Figure 4-5. On this diagram, the break and sustain points are plot as a function of the communication costs (at the power 1-σ). When either ϕ and $\varepsilon^{1-\sigma}$ are close to zero, transportation and communication costs are both near prohibitive. Conversely, when they limit to unity, trade in both goods and back office services is free. The break points do not depend on ε; consequently, both ϕ_B and ϕ^B are flat on the figure. By contrast, the sustain points are independent of ε only when communication costs are prohibitive, i.e. whenever and $\varepsilon^{1-\sigma}<\varepsilon_0^{1-\sigma}$. When communication costs are low enough, however, arbitrage opportunities on the labour market exist and, as depicted above, ϕ_S increases with ε (decreases with $\varepsilon^{1-\sigma}$) whereas ϕ^S decreases with ε (increases with $\varepsilon^{1-\sigma}$).

4.6. 'History of the world, part II'?

In the recent past, trade barriers and communication costs have fallen dramatically – since the end of WWII, the cumulative number of transatlantic and transpacific voice paths did so by a factor 30, while the factor falls to less than 6 for the air freight costs (see the introduction). How do these two trends differ qualitatively in terms of their effect on spatial distribution of economic activity?

The title of this subsection is inspired by Fujita et al. (1999, chapter 14) and to Krugman and Venables (1995). These authors refer to the rising inequalities between a rich, developed and a poor, undeveloped (and possibly developing) hemispheres as the 'History of the world, part I.' Their model –qualitatively similar to the model of section 4.2– suggests that the same process (falling trade and transportation costs) that may have driven income divergence in the first place may, if furthered, yield convergence in both the industrial structure and in cross-factor income distribution. This scenario is referred to as Part II of the 'History of the world' in the studies quoted earlier. This scenario can be visualized by moving horizontally in Figure 4-3. When ϕ is either low or large, both regions are equally industrialised (n=½) and equally rich. When ϕ takes intermediate values, one region only is industrialised (n=0 or n=1) and the other one lags behind. Note the importance of the assumption of decreasing returns in A for this result to hold.

We also saw that removing communication costs is equivalent to removing decreasing returns in A at the margin. Therefore, with low communication costs, the most likely outcome is that industrial structures will keep diverging when transportation costs and trade barriers fall – that is, a core-periphery pattern prevails. However, nominal incomes might converge at the same time.

Integration in the standard model

Consider the standard case first ($\varepsilon > \varepsilon_0$). This replicates Krugman and Venables (1995) and Fujita et al. (1999, chapter 14). As trade costs decrease, the anti-agglomeration force exerted by the presence of two segmented labour markets with imperfectly elastic labour supplies eventually comes to dominate any other effect. As a result when ϕ is large the sole, stable equilibrium is $n=½$.

If, on the other hand, we assume that such returns are constant ($\theta=1$), then the conclusion is reverted: as ϕ limits to unity, agglomeration forces come to dominate stabilizing forces and the core-periphery outcome is the single stable equilibrium. For $\theta=1$, we have seen that $n \in \{0,1\}$ for all ϕ larger than ϕ^{sust}. But this outcome is an artefact. For any $\theta<1$, the core-periphery outcome is stable for all ϕ in $[\phi_S, \phi^S]$ – two roots that are strictly smaller than unity – so that the graph of the stable equilibria against θ is not lower hemi-continuous at $\theta=1$.

The implicit assumption for this result is that ϕ captures all sorts of transaction costs, be they trade, transportation, or various communication costs. But if one accepts the idea that communication costs affect more intangible transactions not directly related to the manufacturing process, namely, services, whereas trade/transportation costs affect the delivery of the final good to a foreign market as well as the import of intermediates, then a fall in all these forms of transaction costs have potentially qualitatively different effects on the equilibrium location of firms and, consequently, on income divergence.

Integration in the extended model

Consider the extended model ($\varepsilon < \varepsilon_0$) with $\varepsilon \to 1$. As communication costs disappear, the two labour markets become ever more integrated: wages in 1 and 2 can no longer diverge –not because workers can move freely from 1 to 2, but because some services can be provided by people working in a distinct location at no additional cost.

Hence nominal wages in poor and rich countries are expected to converge. Therefore, location does not really matter anymore as far as labour market considerations are concerned. The capital owners now base their location decision exclusively on considerations related to market access and price of intermediate inputs.

Section 4.5 revealed that the break interval $[\phi_B,\phi^B]$ is invariant in ε as long as ε is strictly larger than unity (equivalently, $\varepsilon^{1-\sigma}<1$), and so it is drawn as a horizontal pipe here. On the other hand, agglomeration forces are reinforced by any increase in $\varepsilon^{1-\sigma}$ above $\varepsilon_0^{1-\sigma}$ (i.e. an decrease in communication costs) when agglomeration has already taken place. Below $\varepsilon_0^{1-\sigma}$, communication costs are prohibitive and irrelevant to the sustain interval $[\phi_S,\phi^S]$; hence the "loud-speaker" shape of this interval in figure 3.1. At the limit, we saw above that $\phi_S=\phi^{sust}$ and $\phi^S=1$ when $\varepsilon\to 1$.

4.7. Wage inequalities

It is widely thought that spatial integration of some sort can have a big distributional impact on factor rewards. In trade theory, the point was famously made in the early forties by Stolper and Samuelson (1941). They showed how relative prices of good would influence factor rewards and hence focused on intra-regional, cross-factor distributive effects. Trade integration also hinders inter-regional distributive effects – both in real and in nominal rewards. In this section we look at how falling trade barriers and communication costs affect the distribution of income in the population as a whole.[77]

Before turning to the analysis, it is worth remembering a few facts. In the words of Baldwin and Martin (1999), post-WWII globalisation has been a 'zero sum game' between unskilled and skilled workers in all OECD countries. The former have seen their earnings falling sharply relative to the earnings of the later; in the face of unemployment, this inequality is forcefully present as well. There is no consensus about the extend of the causality from trade openness to these figures but up to 50% of the wage gap may be due to trade and migration. See Friedman (1995) for a critical survey of the empirical evidence on the causal link between rising imports of manufactured goods from developing countries and rising inequalities in real wages in OECD countries.

---

[77] I concentrate the discussion on w and π; this is w.l.o.g because r behaves as the mirror image of w. See expression (4-1).

The least disputed factor at the root of this gap is most certainly skill-biased technological change. Neither this nor migration of unskilled labour is modelled here, but it might still be interesting to consider the impact of transportation versus communication cost reduction in the framework of our model.

Falling communication costs

Potentially two effects arise when labour markets get more and more integrated as communication costs fall. The first is a direct effect: when ε falls, the demand for different factors changes, and so their equilibrium prices. The second effect is indirect: given ϕ, a variation in ε might affect the spatial equilibrium and the shift from a long-run equilibrium to another has its own implications on factor prices. I consider the direct effect first, both at the dispersed and at the concentrated equilibria ($n=\frac{1}{2}$ and 1, respectively). There is no need to consider other instantaneous equilibria because they would not be stable (if they exist at all) and hence would not constitute a long run spatial equilibrium.[78]

Marginal changes in ε

We consider here changes in ε that are small enough so that the nature of the location equilibrium does not change. Start with the dispersed equilibrium. At $n=\frac{1}{2}$ all nominal factor rewards are constant and equal unity by choice of units, so that anybody's real income is just equal to the inverse of the price index, viz. $\Delta^{\mu/(\sigma-1)}$, where $\Delta^{1-\alpha}=\frac{1}{2}(1+\phi)$. Hence Δ can readily be used as a metric for indirect utility. Since nominal wages are the same in both regions $m=0$ for any $\varepsilon>1$. Therefore all rewards, nominal and real, are unaffected by any move in ε.

This absence of conflict falls apart when the economy is on the concentrated equilibrium. There nominal factor rewards diverge according to both the nature of the factor and its spatial location. Here I assume that $n=1$ and $\phi\in(\phi_S,\phi^S)$ for all values of ε under consideration. To account for the evolution of real rewards, we also need an expression for the price indices. From (4-3) and (4-18), we have $\Delta_1=\varepsilon^{(1-\sigma)/2}$ and $\Delta_2=\phi\varepsilon^{(1-}$

---

78 I acknowledge that this fact depends upon the functional form chosen for sector Y production function. Other choices might well bring stable, asymmetric equilibria for intermediate values of ϕ. For instance, if U is of the form A+ln(M) rather than Cobb-Douglas (with A as the quantity consumed of good A and M as the CES aggregate of all varieties of good M), when it exists, an asymmetric, interior equilibrium is stable. Mathematically, the break point comes before the sustain point when returns in labour are non-decreasing in sector A.

$^{\sigma)/2}$. Hence, as intuition would predict, prices decrease as communication costs fall, viz. $\partial\Delta_j/\partial\varepsilon<0$ (j=1,2).

Solving the model for n=1 gives (4-18) and (4-19), i.e. the solutions for m, w_1, w_2, π_1, and π_2, the latest being a shadow reward. Clearly, w_1 decreases and w_2 increases as communication costs fall below ε_0, for the reasons already described in section 4.5. The real wages evolve in the same direction as nominal wages. Define ω_1 as $w_1\Delta_1^{\mu/(\sigma-1)}$ and ω_2 accordingly. We find the following elasticities:

(4-23) $$\frac{\partial\omega_1/\partial\varepsilon}{\omega_1/\varepsilon}=\frac{1-\mu}{2}>0,\quad \frac{\partial\omega_2/\partial\varepsilon}{\omega_2/\varepsilon}=-\phi^{\frac{\mu}{(\sigma-1)(1-\alpha)}}\frac{1+\mu}{2}<0$$

if $\varepsilon<\varepsilon_0$ and 0 otherwise. In other words, the impact of falling wage costs does not compensate for the fall in 1's workers' nominal wage and amplifies the increase in 2's workers 2's nominal reward.

It is worth noting that both wages converge towards unity as communication costs vanish, which is exactly the dispersed equilibrium nominal wage, hence workers are unable to capture any of the agglomeration rents. However, it can be seen that aggregate nominal wages $(Lw_1+Lw_2)/2$ decreases as ε diminish below ε_0 (it remains unchanged otherwise), hence 'globalisation' is not exactly a zero-sum game for the workers. Clearly, when ε is low, competition among workers is fiercer (the aggregate labour supply curve is more elastic) and this hurts them; workers in 1 take more than their fair share of the hit since w_2 does increase as ε increases. However, aggregate real wages $(L\omega_1+L\omega_2)/2$ increases when communication costs fall when ε is small enough. Mathematically,

(4-24) $$\frac{\partial(L\omega_1+L\omega_2)/\partial\varepsilon}{(L\omega_1+L\omega_2)/\varepsilon}=\frac{\varepsilon(1-\mu)-\phi^{\frac{\mu}{(\sigma-1)(1-\alpha)}}(1+\mu)}{2(\phi^{\frac{\mu}{(\sigma-1)(1-\alpha)}}+\varepsilon)}<0$$

if ε and ϕ are close enough to unity. A necessary condition for this elasticity to be negative is $\varepsilon<(1+\mu)/(1-\mu)$; it is easy to check that $(1+\mu)/(1-\mu)$ is larger than ε_0, hence the necessary condition is always fulfilled. If ϕ is large enough, then $\varepsilon<\varepsilon_0$ is also sufficient. In such a case, the favourable cost-reduction effect dominates the unfavourable nominal wage effect when $\varepsilon<\varepsilon_0$. This is intuitive: when $\varepsilon\approx1$, the reduction in w_1 is equal (in percentage terms) to the increase in w_2; since the two populations of workers are equally sized, this is purely a distributive effect and hence the net aggregate effect is nil. By contrast, costs fall uniformly in percentage terms, and hence every worker benefits. So it must be that the net effect of a fall in ε is positive when ε is small enough.

We finally turn to $\varpi_j \equiv \pi\, \Delta_j^{\mu/(\sigma-1)}$, the real capital reward for an owner who lives in j. Recall from (4-17) that equilibrium operating profits are non-decreasing in ε, viz.

(4-25)
$$\frac{\partial \varpi_j / \partial \varepsilon}{\varpi_j / \varepsilon} = \Delta_j^{\frac{\mu}{(1-\alpha)}} \frac{\varepsilon(1-\mu)-(1+\mu)}{2(\varepsilon+1)} < 0$$

if $\varepsilon<\varepsilon_0$ and 0 otherwise. Consequently, there is clearly no class conflict between capital owners and workers but there is an 'international' conflict among workers themselves for all values of ε in $(1,\varepsilon_0)$.

Falling transportation costs

In this sub-section we fix ε and consider the implications on real incomes of falling transportation costs. Turn again to Figure 4-5. For ϕ on $[0, \phi_S)$ or $(\phi^S,1]$, the symmetric equilibrium n=½ is the sole stable one; on $[\phi_S,\phi_B]$ or $[\phi^S,\phi^B]$, there are two more stable (long run) equilibria (h=0 and h=1). On (ϕ_B,ϕ^B), these two are the unique spatial equilibria.

Imagine the economy starts in a situation in which both transportation and communication costs are large, such as at a point like A in the figure. Two facts are worth emphasizing. Clearly, at point A n=½ is the unique stable equilibrium in such circumstances. All nominal rewards are equal to 1 by assumption and the real returns are equal to $\Delta_j^{\mu/(\sigma-a)}<1$, with Δ_j (implicitly) defined as $\Delta_j^{1-\alpha}=½(1+\phi)$, j=1,2

Also, $\varepsilon>\varepsilon_0$ implies m=0 for all n. Now imagine that ϕ increases over time along the arrow that links points A and B on the figure. As long as no relocation ever takes place, which is the case if $\phi<\phi_B$ when agents are myopic, all factor owners earn one monetary unit, and their real returns are increasing in ϕ.

When ϕ increases beyond ϕ_B catastrophic agglomeration takes place in, say, country 1 and hence π, w_1, and Δ_1 incur a discrete, positive jump whereas both w_2 and Δ_2 fall in a discrete manner. The relevant expressions are (4-18) for the nominal rewards (with $\varepsilon'=\varepsilon_0$) and

(4-26)
$$\Delta_1^{1-\alpha} = \varepsilon_0^{(1-\alpha)(1-\sigma)/2}, \quad \Delta_2^{1-\alpha} = \phi\varepsilon_0^{(1-\alpha)(1-\sigma)/2}$$

for the Δ's. Using (4-18), (4-26), and the expression of Δ at n=½, the (discrete) change in welfare for the various factor owners, denoted as δ(.), is found to be:

(4-27)

$$\delta(\omega_1) = \varepsilon_0^{(1-\mu)/2} - \left(\frac{1+\phi_B}{2}\right)^{\frac{\mu}{(1-\alpha)(\sigma-1)}},$$

$$\delta(\omega_2) = \phi_B^{\frac{\mu}{(1-\alpha)(\sigma-1)}} \varepsilon_0^{(-1-\mu)/2} - \left(\frac{1+\phi_B}{2}\right)^{\frac{\mu}{(1-\alpha)(\sigma-1)}},$$

$$\delta(\varpi_1) = \frac{1+\varepsilon_0}{2}\varepsilon_0^{(1-\mu)/2} - \left(\frac{1+\phi_B}{2}\right)^{\frac{\mu}{(1-\alpha)(\sigma-1)}},$$

$$\delta(\varpi_2) = \frac{1+\varepsilon_0}{2}\phi_B^{\frac{\mu}{(1-\alpha)(\sigma-1)}} \varepsilon_0^{(1-\mu)/2} - \left(\frac{1+\phi_B}{2}\right)^{\frac{\mu}{(1-\alpha)(\sigma-1)}}$$

By inspection, $\delta(\omega_1)$ and $\delta(\varpi_1)$ are positive, $\delta(\omega_2)$ is negative, and the sign of $\delta(\varpi_2)$ is ambiguous. These are the result of four effects.

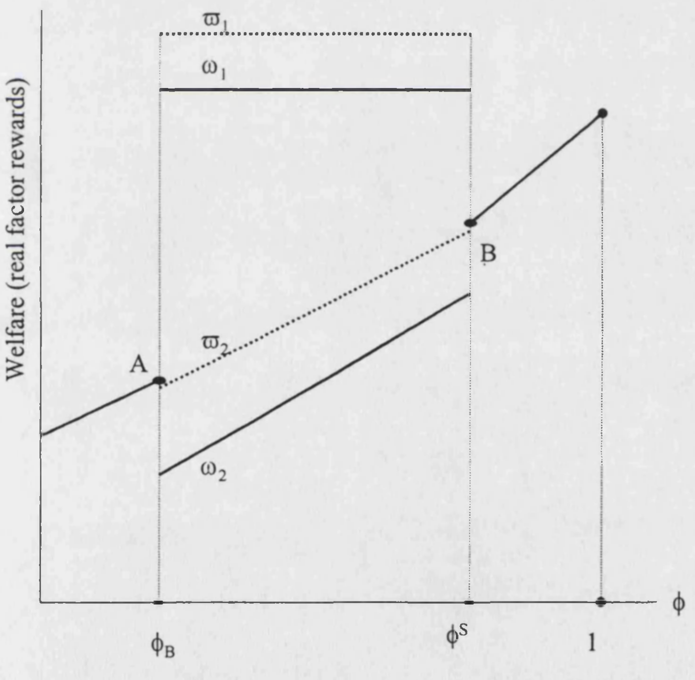
First, when n jumps from ½ to 1, the nominal rewards jump from 1 to a number larger than 1 for all capital owners and for the workers in 1; w_2 falls below unity. This is so because capital owners capture the agglomeration rents and labour become more (less) scarce in country 1 (country 2). Second, since all firms cluster in a single location labour costs increase in the manufacturing sector. All the same, the cost of intermediate inputs falls as firms save on transportation costs for all the inputs that used to be imported but no longer are. The net effect of this on costs, and hence prices, is ambiguous. Finally, as firms cluster in country 1, consumers in 1 save on the transportation costs of the varieties they no-longer import (the opposite is true for 2's consumers). The net effect is a-priori ambiguous for all factors.

Workers in 1 gain because their nominal wage increases. The price of manufacturing good, even if it increases, increases by less for three reasons. First, labour costs represent only a fraction of the variable costs of the firms. Second, the fact that firms fully pass the additional costs onto consumers is an artefact of the model, but this is clearly an upper bound in a more general context. And third, when they join a cluster, firms save on the costs of intermediates; this reduces the price of the final good since this cost reduction is passed onto consumers. Capital owners who live (and consume) in 1 gain for essentially the same reasons. Their nominal return increases by more than w_1, as can be seen in (4-18), so it must be that $\delta(\varpi_1)>0$.

Workers in 2 unambiguously loose for reasons that are broadly symmetric to the increase in ω_1. Their nominal reward falls; and even if the variable production costs of the manufacturing goods falls, the price index of their bundle of these goods increases because they now have to import all units they consume. Finally, capital owners who live in 2 may or may not enjoy a greater utility because both their nominal reward and the price index on the manufacturing goods they consume increase.

This discussion is summarized in Figure 4-6. This figure is the normative counterpart to Figure 4-3. For the sake of clarity, and to be consistent with the thought experiment above, I have only plot the welfare workers and capital owners from both countries enjoy at the symmetric equilibrium for $\phi \leq \phi_B$ and $\phi \geq \phi^S$ and at the concentrated equilibrium n=1 when $\phi \in (\phi_B, \phi^S)$. When ϕ is low, welfare changes according to falling prices in manufacturing goods only. Then real rewards shift as shown (except perhaps for ϖ_2). Real wages and capital rewards are represented with plain and dotted lines, respectively.

Figure 4-6. Trade liberalisation and welfare



As trade liberalization proceeds further or transportation costs keep falling, nominal reward and price indices in 1 remain unchanged. Consumer prices in 2 fall, hence everybody is made no worse-off beyond ϕ_B as long as the core-periphery configuration is sustainable, i.e. as long as $\phi < \phi^S$. When trade/transportation costs are so low that a point like point B in Figure 4-5 is reached, half of the firms move back to 2. Such a 'catastrophic' dispersion has welfare effects that are perfectly symmetric to those described in the previous paragraphs, namely, $\delta(\omega_1)<0$, $\delta(\varpi_1)<0$, $\delta(\omega_2)<0$, and the sign of $\delta(\varpi_2)$ is ambiguous. Since $\phi^S>\phi_B$, the magnitude of these effects is smaller for factor owners located in 1, but ambiguous for factors in 2. When $\phi\approx1$, all factor rewards, both nominal and real, are equal to 1.

Path dependency

So the world economy is at point B where industries are evenly spread between the two countries. Assume now that communication infrastructures improve so that we move horizontally to point C. Since agents are myopic, we know that industries remain evenly spread out as long as the boundary ϕ^B has not been crossed. In other words, even though a core-periphery pattern is sustainable at point C, the manufacturing plants remain dispersed as in B.

Instead, imagine the following, completely different integration path.

Start from point A in Figure 4-5 again. At a point such as A firms are equally active in both countries. Imagine now that transportation and communication costs decrease simultaneously –call this globalisation– so that the integration path looks more like the arrow from A to C. When this arrow crosses ϕ_B then all firms cluster in a single location. Possibly (if $\varepsilon<\varepsilon_0$), some of them hire external workers. When globalisation proceeds further towards point C the curve ϕ^S is never crossed, so the core-periphery pattern remains a stable long-run equilibrium.

In other words, history matters. In the first case, when the world economy globalises in two waves (from A to B first, and then from B to C), then ultimately both economies converge. When transportation costs decrease alongside communication costs (as illustrated by the move from A to C along the diagonal arrow) then divergence prevails.

4.8. Conclusion

This chapter used a standard model of economic geography to assess the effects of falling trade costs as well as communication costs on both the spatial equilibrium and the wage distribution. In particular, a 'double tomahawk' can summarize the dynamics of the model. For high trade costs firms want to be close to final demand; because some factors are inter-regionally immobile (workers and land), so is some final demand; therefore capital (and hence firms) is allocated evenly in the two regions. At the other extreme, for very low trade costs, locating close to final customers is no longer a key issue. Rather, entrepreneurs will save as much as they can on costs and hence locate where cheap labor is. As labor is evenly split between the two regions and labor supply is finitely elastic, these considerations will again yield a dispersed spatial equilibrium.

At intermediate trade costs, however, these two dispersion forces are weaker than the main agglomeration force: if all entrepreneurs are concentrated in region 1, then shipment costs are saved on half of the varieties that used to be imported. Moreover, each other's demand for intermediates is greater when they co-locate.

As the spatial distribution of industrial activity is even, so are real and nominal incomes: all rewards are equalized across regions, and so are the costs of living. When the spatial allocation of firms diverges, so do these variables: the cost of living and the land reward are lower in the core, whereas workers' wage is lower in the periphery.

We then considered what happens when labour markets get integrated as well. We assume that labour markets become ever more integrated because communication costs drop over time. This enables firms in region 1 to buy the services of workers in region 2 at a cost that encompasses any difference in language, the legal system and genuine communication costs. This has the clear, immediate implication that nominal wages can only diverge up to the point at which the no-arbitrage condition between nominal wages binds; spatial wage and land reward inequalities are non-decreasing in those costs.

The second implication is that one dispersion force is now weakened by this mechanism: labour markets are no longer completely distinct, so that agglomeration is now sustainable over a larger range of trade costs – some firms would now simply hire or buy services from (that is: hire) workers living abroad. This has an adverse effect on spatial wage equality, as we already know that spatial agglomeration implies factor rewards and costs of living to diverge. Save for landowners, the two effects go in the same direction, so that real incomes diverge as well.

The net effect on inequality among people living in the same region is generally ambiguous, and for the same reason: given that the economy is on the concentrated equilibrium, falling transportation costs is generally a good thing because nominal rewards usually converge. On the other hand, if it prevents the economy to going back to the dispersed equilibrium (where factor rewards diverge least), then this form of labour market integration is a bad thing on equity grounds.

Finally, we saw that the integration path itself is very important for the north-south duality in both the production structure and real income divergence. In particular, the model suggests that when communication costs fall faster than transportation costs, as the evidence surveyed in the introduction suggests, then it is much more likely that the non industrialised nation remains locked-in the specialisation of labour-intensive products and routinised business services like call centres.

Clearly, this model gives a relatively rich picture of divergence in real incomes – both across factors and across regions – and generates losers and winners that might differ whether labour market de-segmentation or trade integration –or both – takes place. The political economics in such a framework are therefore more involved that in simple 'zero-sum' games, where essentially one factor or one region gains while the other looses.

Appendix

The analysis here complements section 4.4. The 'dispersed equilibrium' is said to be unstable if $d\pi_1/dn>0$ at $n=½$. (By the symmetry of the model, $d\pi_2/dn=-d\pi_1/dn$ at $h=½$.) In words, the dispersed equilibrium is unstable if moving one unit of K from 1 to 2 increases the capital reward in 1 relatively to 2 (remember that capital reward equals operating profit by free-entry). In such a case, agglomeration forces dominate and, by the law of motion (2-7), n increases further. That is, the initial shock is not self-correcting. Accordingly, any long run interior equilibrium n' such that $\pi_1=\pi_2$ is said to be unstable if $d(\pi_1-\pi_2)/dn>0$ at n=n'.

We then differentiate the system around the symmetric equilibrium, as in Chapter 1; we also use the symmetry properties of the model and write $dw_0=dw_1=-dw_2$, $d\pi_0=d\pi 1=-d\pi 2$, etc. See Chapter 1 for details. This way we get a system in $d\ln\Delta_0$, dw_0, and $d\pi_0$, dE_0:

$$\textbf{(4-28)}\quad \begin{bmatrix} 1-\alpha\beta Z & 2\beta[\alpha(\sigma-1)-\mu Z] & Z-\alpha \\ 0 & (\sigma-1)(1-\alpha)Z & 1-\alpha Z \\ -\beta & \beta+(1-\alpha\beta-\mu)\theta/(\mu(1-\theta)) & 0 \end{bmatrix} \begin{bmatrix} d\pi_0 \\ dw_0 \\ d\Delta_0/\Delta_0 \end{bmatrix} = 2\begin{bmatrix} \alpha\beta Z \\ Z \\ \beta \end{bmatrix} dn$$

where $Z\equiv(1-\phi)/(1+\phi)$ and $\beta\equiv1-1/\sigma$ as before, and dn is treated as exogenous. By Cramer's rule, it is easy to get a solution for $d\pi_0/dn$; the break points ϕ_B and $\phi^B<1$ are the zeroes of the resulting second order polynomial in ϕ. In the limiting case $\theta=1$, $\phi_B=\phi^{break}$ in (2-9) and $\phi^B=1$.

With the addition of decreasing returns in A, both ϕ_B and ϕ^B are smaller than unity – when they are real. The reason is, again, that these decreasing returns act as a dispersion force that does not depend on ϕ, so it must be that $d\pi_0/dn<0$ at $n=½$ (namely, the symmetric equilibrium is stable) when ϕ is arbitrarily close to 1. The general solution to (4-28) is not particularly enlightening, even in the special cases $\alpha=\mu$ or $\theta=½$.

However, there are a couple of useful facts that some standard (if tedious) algebra reveals. At the limit $\alpha=1$ (agglomeration forces as measured by strength of vertical linkages are maximal) we have $\phi_B=0$ and $\phi^B=1$, all $\{\sigma,\mu,\theta\}$; this implies that the symmetric equilibrium is never stable. By contrast, $\alpha=0$ implies that the symmetric equilibrium is always stable (the model essentially collapses to the FC model of Chapter 1).

Note an important caveat here: a necessary condition for ϕ_B and ϕ^B to be sufficient statistics for our problem is that the denominator of $d\pi_0/dn$ is different from zero. It turns out, however, that the polynomial that constitutes the denominator generically admits two zeroes. One is always negative and, as such, does not make any economic sense, hence we disregard it entirely. The largest one is sometimes positive (but not always) and smaller than 1, so it matters. Define it as ϕ_0^B. With some effort, we can check that $\alpha=1$ implies $\phi_0^B=0$. However, it can be shown that the following is true:

(4-29) $$\phi_0^B=0 \Leftrightarrow \min\{\max\{0,\phi_B\},\phi^B\}=0$$

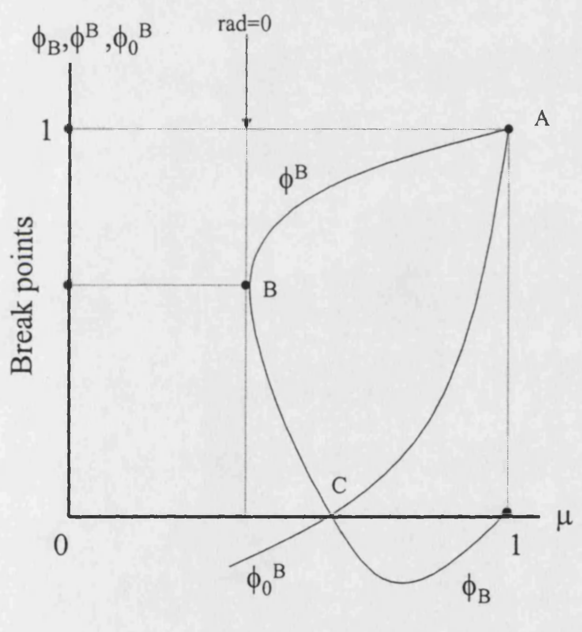
in the simplifying case $\mu=\alpha$ Fujita et al. (1999) and others assume. In words, (4-29) says that, $\phi_0^B=0$ if and only if the smallest between ϕ_B and ϕ^B is nil, provided it belongs to the meaningful range [0,1], are the same. This fact is useful for the graphical analysis we are now conducting.

Figure 4-7 plots ϕ_0^B, ϕ_B and ϕ^B as a function of α, for some $\mu=\alpha$, $\sigma=5.31$, and $\theta=.3$. The first locus, ϕ_0^B, is the curve that links (0,-1) and point A. The second and third loci, ϕ_B and ϕ^B, link point B to point C and A, respectively. At point B the radical of the solution to ϕ_B and ϕ^B is negative, which implies $\phi_B=\phi^B$. Also, ϕ_0^B and ϕ_B cross on the horizontal axis for some α in (0,1). All these properties follow from the observations made in the previous paragraph.

The configuration plot in Figure 4-7 is typical: for low values of α (when agglomeration forces are weak), no real root exists (the radical is negative) and the symmetric equilibrium is always stable. On the other hand, when μ is sufficiently large (i.e. the radical is positive), the symmetric equilibrium is unstable for all combinations of μ and ϕ in the convex set ABC.

When agglomeration forces as measured by σ so weak that $\phi_B<0\ \forall\alpha$, ϕ_B rotates anti-clockwise; it is increasing in μon the relevant range, but negative for low values of α.

Figure 4-7: Break points (parameter values: $\alpha=\mu$, $\sigma=5.31$, $\theta=.3$)



It can be shown that when σ is low enough (agglomeration forces strong enough) the convex set of the combinations (α,ϕ) such that $\phi\in[\phi_B,\phi^B]$ expands; in particular, the value of α such that the radical is nil is lower. Further simulations show that the analysis for general values of θ is qualitatively identical (in particular, this set expands when θ increases because this corresponds to lower dispersion forces). Therefore, I am confident the analysis conveyed is exhaustive.[79]

[79] The Maple worksheet BreakAnalysis.mws is available on request.

Chapter 5. THE HOLD-UP PROBLEM, INDUSTRY-SPECIFIC SKILL ACQUISITION, AND FUNCTIONAL AGGLOMERATION

5.1. Introduction

All chapters thus far fit the so-called 'New Economic Geography' (NEG) paradigm whereby agglomeration, or the clustering of economic activities, is the result of the interaction between increasing returns at the firm level, transportation costs, and factor mobility across sectors or space. The emphasis is put on the technology (increasing returns) and the market conduct (imperfectly competitive segmented markets). By contrast, factors are assumed to be perfectly homogenous and the factor markets to be perfectly competitive. That is, factor market motives for agglomeration are disregarded.

The aim of the present chapter is to remedy to this and focus our attention on labour market considerations. In order to draw as clear conclusions as possible, this chapter abstracts from final goods market considerations.

Many industries are spatially concentrated in the world, and this is by no means a recent phenomena. For instance, Marshall (1920) reported almost a century ago that the British cutlery manufacturing was concentrated in Sheffield. This is still the case nowadays.

Beyond anecdotic evidence, Duranton and Overman (2001) show that 51% of the four-digit UK industries and, confirming earlier empirical work, show that location externalities must be highly localised. Interestingly, their results report that industries that belong to the same branch have similar localisation patterns. This last fact can of course be explained by the NEG if similar industries have similar input-output matrices (as in the work of Puga and Venables 1996). Or it might be that similar industries use workers with similar skills and, in turn, that firms co-locate along the skills they use. The theory of this chapter provides an explanation for the co-location of firms that employ workers with similar skills, regardless of the industry in which they work. In other words, the functional specialisation of a location, to use the terminology introduced by Duranton and Puga (2001), permits the labour market pooling of workers with similar skills. By contrast, when market access and supplier access are crucial for an industry then sectoral agglomeration may arise as the outcome of that tension. This was the topic of all earlier chapters.

As an example, take the accordion industry. The world production is dominated by a bunch of medium and small firms located in an Italian village, Castelfidardo, near Ancona in the region named Le Marche. Clearly, the accordion industry is spatially agglomerated by any measure. According to the NEG, we should suspect the presence of immense scale economies, low transportation costs, and strong input-output linkages. I have not run a proper econometric study on this specific industry, but this line of reasoning looks doubtful as a candidate explanation for the spatial concentration of this industry. Firms in Castelfidardo serve the world market and the stuff they produce is easily transportable. This suggests that market access considerations are not the central concern in the location decision in this industry. In other words, firms in this industry do not seem to have clustered in this medieval town because they belong to the same industry *per se* but, as we shall argue, because they employ workers with similar skills. This distinction is subtle but important.[80]

The reason why this industry is actually located in this particular town is historical.[81] The reason this industry is agglomerated in a single location at all (or almost) is that, locals would tell you, they have the *savoir-faire*. Which puts the question mark one step further: why is this savoir-faire agglomerated? The theory this chapter puts forward relies on solving (or attenuating) a 'hold-up' problem in skill acquisition.

In industries like the accordion industry, where skills are highly specific, workers' investment risk being ‘held-up’ by firms that employ them. An investment is held up if one party must pay the cost while others share in the payoff (Acemoglu and Shimer 1999).

The argument developed in this chapter runs as follows. An entrepreneur has to choose where to locate her firm. To produce, she needs some workers. The entrepreneur

---

[80] Actually, some firms in Castelfidardo specialize in the production of the instruments, others mostly repair them, and yet others produce some components only.
[81] See the website http://www.comune.castelfidardo.an.it/Visitatori/Fisarmonica/storia_fisa.htm

and the workers own complementary production factors. The worker can increase her productivity by incurring some costly investment (education, say) into acquiring specific skills. Once she has done so, the worker decides to join an entrepreneur or another. Since skill acquisition precedes the productive relationship, the firm will pay the worker her reservation wage. In a location in which nobody else has any use of her skill, a worker's reservation wage is independent of her skills. By contrast, if the worker joins a firm that locates near competitors for that skill then her outside options depend positively upon her skills.

The conclusion follows. In case a worker joins a firm that stands alone (in the sense that nobody else value her skills), she has no incentive to acquire these skills in the first place. By contrast, if she joins a firm that stands among many, competition for her skills ensures that her wage will depend on her skills. This provides her with the incentive to acquire the skills in the first place.

The entrepreneur is perfectly aware of this fact. If she locates in isolation she knows that she cannot credibly promise the worker to pay her according to her qualifications in the future when she joins her. Co-location and the resulting competition for the workers' skills, in other words, provide a credible way to pay high wages once the worker has made the investment.

To put it another way, a hold-up problem arises if the investment is relation-specific –which is the case if the firm locates in isolation. This hold-up problem disappears if the investment is industry-specific.

This story works under the following implicit assumptions. First, moving from one location to another is costly for the worker. This can be justified on several grounds. Leaving a city for another entails monetary costs, obviously, but it also entails leaving a network of friends, looking for a new school for the children and other intangibles of the like. Second, the amount and quality of skills acquired might be observable by all parties but is not verifiable by a court, say. As a result, it is not possible to write a contract contingent of the amount of skilled acquired.

As an aside, this story is one about providing investment incentives. It is not a story of committing to pay higher wages per se. Indeed, we assume that location *is* contractible, so that the firm made of the entrepreneur and workers will chose to locate so as to maximize the joint surplus. If this joint surplus is maximised at some positive investment level, then firms will co-locate at equilibrium. In sum, this mechanism shows how a thick labour market for some specialized skills provides a rational for agglomeration.

Our setting is closely related to Rotemberg and Saloner (2000). They assume a simple setting where two *very different* parties are necessary to each other to produce a given good. One has all market power; the other is to make a non-verifiable investment. Since this is essentially the setting described above, they get the same result: agglomeration takes place as it solves the hold-up problem.

We shall stress below that a very specific conjunction of assumptions is needed for this result: agglomeration is the efficient spatial organization of production if the party that has to make important relation-specific investments faces a potential hold-up problem. In the example above, if the firm has to make unverifiable investments as well, then it might be better for it to locate somewhere in the 'periphery' and form a one-firm town so as to grab a larger share of the surplus generated at the production process so as to, which in turn gives it higher incentives to invest. This suggests an explanation for the 49% of the four-digit UK manufactures that fail to be agglomerated according to Duranton and Overman (2001).

When firms locate apart, this increases the degree of competition among workers who sell their labour on local labour markets. Hence, in general, it might well be that agglomeration *worsens* the hold-up problem, turning Rotemberg and Saloner's (2000) result upon its head. In sum, the respective identity of the party that exerts market power and of the one that makes an industry-specific investment is crucial. This suggests a potentially rich theory in which the relative importance of the investments of the parties involved, the specificity to their relationship, and their outside options all mater.

This also implies that agglomeration and dispersion forces are of the same nature, by contrast to other papers. (Typically, other studies assume that agglomeration economies arise in thick markets and that congestion costs are the result of land scarcity.) A final technical point deserves some mention here. Technically, agglomeration economies arise when there are increasing returns to scale so that size matters. In this chapter, I assume away all non-convexity in the production function as a simplifying assumption. Hence, there might be benefits associated to the co-location of some agents in the same region but these are not agglomeration proper. I leave this issue in the background so as not to blur the analysis.

The remainder of the chapter is organised as follows. Section 5.2 introduces a simple model in which one party of a production relationship has all the market power (the firm) whilst the other one (the worker) has to make an investment that improves the productivity of the relationship. It is shown that in this simple and very specific example, Rotemberg and Saloner's (2000) result obtains. In Sections 5.3 and 5.4, we establish more symmetry between the parties and concentrate on the question of how market power, investment in relation-specific investments, and location all interact. Finally, section 5.5 concludes.

5.2. A simple model

This section considers a disturbingly simple model that will prove sufficient to illustrate the main point of this chapter: agglomeration might exacerbate or soften the hold-up problem. There are two types of agents in our economy, entrepreneurs and workers. There is a single industry in which production requires agents of both types to establish a productive relationship, which we call 'a firm'. We assume that one party (potentially) exerts market power and that one party makes an industry-specific investment; in this section we assume that these two parties are distinct.

The environment in which entrepreneurs and workers evolve is as follows. There are two 'firms', each made of one entrepreneur (K) and a continuum of workers (L) of mass one. For simplicity, we say that they belong to the same 'industry'. But industry here must be understood from a functional perspective: both firms make use of workers with identical skills to convey the same function (say secretaries), irrespective of what good or service the firm produces.

In this chapter we leave final good market considerations aside. Hence we assume that firms produce a homogenous good under constant returns to scale that they sell on the spot market in a perfectly competitive environment. None of the choice variable of the model has any effect on the price firms might get on the good market.

We also make the following assumptions:

Assumption 1. On the labour market, workers are atomistic but entrepreneurs are discrete units.

This is another way of saying that workers have no market power, whereas entrepreneurs might have some. Next, workers make an industry-specific investment best thought of as acquiring some human capital:

> **Assumption 2.** Workers make a non-verifiable industry-specific investment, which is labour augmenting.

This investment is observable by both parties (no asymmetric information) but not verifiable by a third party (e.g. a court).

As for the location decision, our economy is made of two regions, region 1 and region 2. We assume that location is contractible. Therefore, whoever has to make the location decision chooses to locate in the region so as to maximize the joint surplus of the entrepreneur and the workers. It is natural to assume that firms make that decision.

The exact timing of events, some additional terminology, and the notation are all introduced in the game we define below.

The game

There are two symmetric, ex ante identical regions, region 1 and region 2. If both firms establish themselves in the same location, we dub this region as the 'core'. The empty region is referred to as the 'periphery'. By contrast, we refer to the configuration in which there is one firm in each region as the 'dispersed' outcome. The set of players is made of two entrepreneurs $j \in \{1,2\}$ as well as of a continuum of workers of mass 2: $L \in [0,2]$.

The timing of the game is as follows:

1. Firms chose where to locate. In particular, whether to co-locate or to locate separately.
2. Workers make their investment decisions and join the firms. Firms offer wages that workers may refuse. If they accept, production takes place and the gross surplus (or revenue) R is realised.

In period 1, firms decide where to locate so as to maximize the expected joint surplus, or revenue R. The location decision is contractible, so the identity of who actually makes this decision does not matter for the equilibria of the game. R itself is non-verifiable ex-post and hence non-contractible (see Grossman and Hart 1985 for a full treatment of incomplete contracts).

In period 2 workers make a labour-augmenting, industry-specific investment $\iota \geq 0$ that costs them ι and then join a firm. The entrepreneur makes them an offer that they may refuse, in which case they either get a reservation wage normalised to 0 (crucially, this reservation wage does not depend upon ι) or try to be hired by the other firm. If the two firms locate are dispersed, we assume the cost of moving from one region to the other is prohibitive for the worker. This simplifying assumption is for convenience only; it implies that workers' outside option is zero when there is one firm in each region.

If located apart, it follows from Assumption 1 that entrepreneurs detain full market power. Anticipating the results, no firm has any incentive to pay the workers more than their reservation wage.

If both firms co-locate, we assume that they compete in prices for the workers. This competition a-la Bertrand ensures that the workers are paid their marginal product. This implies that the workers' outside option is their marginal product in this configuration.

This combination of extreme assumptions is purposely designed so as to get bold results to illustrate the mechanism of interest.

Turn to technology. R is defined as $R(\iota L,K)$, where L and K are the mass/number of workers and entrepreneurs of a typical firm, and ι is the labour-augmenting skill acquired at stage 1 by a typical worker. Therefore, ιL is interpreted as the 'effective' labour. There are constant returns to scale in ιL and K. Denote the first derivative to R with respect to its Z^{th} argument as $R_Z(.)$ and its second derivative with respect to the same argument as $R_{ZZ}(.)$.

We impose the following regularity conditions on R:

Assumption 3. (Positive and decreasing marginal products.) $\forall$ Z=1,2, $R_Z>0$ and $R_{ZZ}<0$, all $Z>0$. Moreover, $R_{12}>0$.

Assumption 4. (Inada conditions.) $\lim_{K\to 0}R_2(.)=+\infty$, $\lim_{\iota L\to 0}R_1(.)=+\infty$. Moreover, $\forall Z'=\iota L,K$, $\lim_{Z'\to 0}R_{12}(.)=+\infty$.

The first of these assumptions means that production (and hence revenue) increases with the use of one factor –keeping the other one fixed– at a decreasing rate. Also, the factors are complementary in that marginal productivity of either factor increases with the amount of the other factor used. The Inada conditions state that the marginal productivity of any factor grows unbounded when this factor is used in ever-smaller quantities. By contrast, the reservation wage in finite (and normalised to zero). This ensures that the first best amount of investment ι is positive.

Note that because ι is labour-augmenting, we have $R_L(\iota)\equiv\partial R/\partial L=\iota R_1(.)$, $\partial^2 R/\partial L^2=\iota^2 R_{11}(.)$, $R_\iota(\iota)\equiv\partial R/\partial\iota=LR_1(.)$ and $\partial^2 R/\partial\iota^2=L^2R_{11}(.)$, so the Inada conditions above ensure that the worker will make a positive investment in any equilibrium in which her pay is an increasing function of her marginal product. Finally, by the assumption of non-increasing returns workers will join firms in equal proportions at any symmetric equilibrium, viz. L=K=1 for any firm. Hence, with some abuse of notation we write $R(\iota 1,1)=R(\iota)$.

Accordingly, the joint surplus achieved when workers make no investment is defined as R_0, where $R_0\equiv R(0)=R(0,1)$. R_0 can be smaller or larger than the reservation wage. None of the results below actually depend on the size of R_0.

Finally, we must characterize the payoffs. The entrepreneur is the residual claimant and gets the profit $\pi=R(\iota)-w$, namely the revenue minus the wage he pays to the workers. By the non-contractibility assumption, w cannot be contingent to ι. By constant returns, we have $R_L(\iota)+R_K(\iota)=R(\iota)$. Everybody is risk-neutral.

We now turn to the solutions of the game.

Solution to the game

As usual, we solve the game backward. In stage 2, workers observe the firms location decisions and decide how much to invest in industry-specific skills and then join firms, foreseeing the outcome on the labour market in each case (co-location or dispersion). In stage 1, firms perfectly anticipate the investment decisions and decide where to locate accordingly.

Consider the dispersed equilibrium first. In stage 2, the firm is a monopsonist on the labour market, made of a unit mass of workers (since the two entrepreneurs' problem is identical, we drop subscripts). Its problem is then to maximize its profit:[82]

(5-1) $$\max\{\pi^{Disp} = R(\iota) - w : w \geq 0\}$$

The firm decides how many of the workers who followed it in its location to hire and makes a take-or-leave-it offer in wages. Clearly, the firm offers the reservation wage w=0 to all workers, an offer they accept.

In stage 1, since the two locations are segmented the investment the worker might make becomes relation-specific. Foreseeing her wage to be zero in stage two whatever the amount of ι she chooses (namely, facing a 'hold-up problem') the worker does not invest. The total surplus generated by each firm is then R_0.

Next, consider the concentrated equilibrium. In stage 2 firms compete on the labour market. Competition for workers takes a 'Bertrand' form. The Nash equilibrium (w_N*,w_S*) to this sub-game is defined as:

(5-2) $$w_N^* = \arg\max\{\pi_N^{Conc} = R(\iota L_N) - w_N L_N : w_N \geq 0\}$$

with

(5-3) $$L_N = \begin{cases} 0, w_N < w_S^* \\ 1, w_N = w_S^* \\ 2, w_N > w_S^* \end{cases}$$

and w_S is defined symmetrically. As is well known, competition in prices for a homogenous commodity, as here, is equivalent to perfect competition: workers are paid at their marginal productivity, so w*=R_L(ι). Therefore, expected wages are an increasing function of ι, even though ι is not contractible. This happens because competition in the labour market creates a link between the workers' productivity and their expected wage.

At stage 1, workers maximize their expected wage, anticipating w*:

(5-4) $$\max_{\iota \geq 0} R_L(\iota) - \iota$$

---

[82] Formally, the firm also chooses the number of workers it wishes to hire, L∈[0,1]. Since it is price taking on the good market and labour is in fixed supply, it clearly chooses L=1 whenever the resulting profit is non-negative.

The solution to this problem is given by

(5-5) $$1 = R_{L\iota}(\iota^*) \equiv R_1(\iota^*,1) + \iota^* R_{11}(\iota^*,1)$$

Define ι^{FB} as the first best solution to the investment problem, that is, the investment in human capital that solves $R_\iota(\iota)=1$. The following result from Assumption 3 and the homogeneity of degree one in R(.).

> Result 5-1. Workers under-invest under agglomeration as well as under dispersion. However, they under-invest less under agglomeration than they do under dispersion, viz. $0<\iota^*<\iota^{FB}$.

To see this, note that ι^{FB} is implicitly defined as $R_1(\iota^{FB},1)=1$. Subtracting from (5-5), we find:

(5-6) $$R_1(\iota^{FB},1) - R_1(\iota^*,1) = \iota^* R_{11}(\iota^*,1) < 0$$

The inequality follows from Assumption 3 (decreasing marginal returns). Assumption 3 also implies $\iota^{FB}>\iota^*$. Finally Assumption 4 (Inada conditions) ensures $\iota^*>0$. This demonstrates Result 5-1.

The intuition for this result is simple. The hold-up problem is reduced in the concentrated configuration, so workers internalise part of their investment. Since they are paid at their marginal productivity, and hence they invest so as to increase their wage. Another way of interpreting this result is as follows: when firms conglomerate in a unique region, each worker's outside option improves. Because her investment is industry-specific in this case (and not only relation-specific), any investment she makes also improves her outside option.[83]

The bottom line is that, absent feasible contracts, agglomeration acts as a credible commitment for the entrepreneurs to pay more productive workers a higher wage where a contract could not do this by assumption. Hence, the unique sub-game perfect Nash-equilibrium of the game described above is one in which firms cluster in one region only. Remember that we have made the reasonable assumption that location is contractible so that firms locate where the revenue is highest. In particular, the

---

[83] But note that they do not internalise the positive effect their investment has on the entrepreneur's marginal productivity, so that workers invest less than the first best.

distribution of R between π and w is of no interest, even though no side-payment are needed here (both wages and profits are higher under agglomeration).

5.3. Entrepreneur investment

In this section we turn the previous result over its head by removing one but crucial assumption and imposing its perfect symmetric instead. Assume now that the entrepreneur has both market power on the labour market and has to chose whether to invest and how much. In particular, we maintain Assumption 1 but replace Assumption 2 with the following:

> **Assumption 5.** Entrepreneurs make a non-verifiable, industry-specific investment, which is K-augmenting.

That is, we now write the gross revenue as R(I), defined as R(L,IK) for L=K=1, where I is the entrepreneur's investment. Also, the symmetric of Assumptions 3 and 4 hold for I replacing ι. Finally, because I is capital augmenting, we have $R_K(I) \equiv \partial R/\partial K = IR_2(.)$, $\partial^2 R/\partial K^2 = I^2 R_{22}(.)$, $R_I(I) \equiv \partial R/\partial I = KR_2(.)$ and $\partial^2 R/\partial I^2 = L^2 R_{22}(.)$.

Under dispersion, the entrepreneur maximizes R(I)-I and chooses the first best I^{FB}, where I^{FB} solves

(5-7) $$R_I(I^{FB}) = 1$$

The entrepreneur chooses the first best under dispersion because she can fully appropriate the returns on her investment. Under agglomeration, Bertrand (i.e. perfect) competition on the labour market forces each entrepreneur to pay the workers their marginal product, so (by the assumption of constant returns to scale in production) entrepreneurs are left with their marginal product in value. As a consequence they chose I so as to maximize $R_I(I)$. The solution to that problem is $\hat{I}$, where $\hat{I}$ is implicitly defined as:

(5-8) $$R_{KI}(\hat{I}) \equiv R_2(1,\hat{I}) + \hat{I}R_{22}(1,\hat{I}) = 1$$

Using the same methodology as for proving Proposition 1 above, we have:

(5-9) $$R_2(1,I^{FB}) - R_2(1,\hat{I}) = \hat{I}R_{22}(1,\hat{I})$$

By the assumption of decreasing marginal products (Assumption 3), (5-9) implies the following result:

Result 5-2. Entrepreneurs under-invest under agglomeration and choose the optimal level of investment under dispersion, viz. $0<\hat{I}< I^{FB}$.

Clearly, the entrepreneurs under-invest under agglomeration, because they internalise only the effect of their investment on their marginal productivity. Assuming that location is contractible at date 1, though, each firm would then locate in different locations. (The entrepreneur can compensate the worker for the smaller wage she is going to pay her.) Dispersion provides the right incentives to the right person.

5.4. Core-periphery or dispersion?

We now have everything at hand to state our key proposition. This makes the comparison between Result 5-1 and Result 5-2:

Result 5-3. Under Assumption 2 (workers invest), agglomeration is preferable. Under Assumption 5 (entrepreneurs invest), dispersion is preferable (the first best is even achieved).

In both cases, the nature of the agglomeration and dispersion forces is the same: any firms should locate where the hold-up problem is minimized. Under agglomeration, the workers' market power is higher than under dispersion (her outside options are more valuable); the converse is true for the entrepreneurs. So, firms should locate so as to give maximum market power to the side that invests.

5.5. Discussion

In this chapter we have seen how the interaction between market power, location and non-verifiable investments gives rise sometimes to agglomeration, sometimes to dispersion. The simple model above yields powerful results. When market power and firm-specific investment are complementary this acts as an agglomeration force. When market power discourages investment this acts as a dispersion force.

The most novel insight is that if agglomeration of firms in a given industry increases the market power of the side whose investment is the most important to the (joint) surplus, then we should expect to observe all firms of this industry to locate in a few 'cores'. Conversely, if agglomeration reduces the incentives to invest, then we should expect this industry to be dispersed in more numerous locations. Clearly, the proximity of other firms (or the absence hereof) engenders externalities on the incentives to invest within other firms.

The empirical predictions of the model are that spatial units should be specialised along functional characteristics. This does not conflict with, but rather complement, the NEG prediction that market access and supplier access considerations trigger sectoral agglomeration. Duranton and Puga (2001) propose a NEG model in an urban setting in which there is a tension between sectoral and functional specialisation of cities. The model suggests that falling managerial costs enable firms to fragment the production process so that various plants can be set in various locations so as to benefit from agglomeration economies that are specific to the production stage of the various plants.

This chapter has assumed that investments are industry-specific, but the result can be generalized: if a worker can choose to invest in general skills or to become more specialised, then the presence of other industries in the same location might induce her to invest too much in general skills so as to improve her outside options (in other words, her prospects on the local labour market); this in turn generates negative externalities between industries. To be more specific, and leaving entrepreneurs' investment in the background, assume that ι is bi-dimensional: along one dimension, the worker chooses what amount of skills she wants to acquire; along the other dimension, she decides how specific it is to a given industry. Both are costly. The proximity of another industry B might induce the workers of industry A to invest more (a positive externality), but to do so in less specific skills (a negative externality). Grossman and Helpman (2002) have a general equilibrium setting in which service providers decide how much to invest in relation-specific skills, and how much to invest in more standard ones so as to increase their outside options and hence their bargaining power.

Consequently, a theory of location based on this insight can give rise to a rich set of results. If an agent is to choose between a relation-specific investment and a more general one, for instance, dispersion might then increase the joint surplus, as then the investing party does not undertake the less valuable investment that would increase her outside option instead. We wish to pursue in this direction along the lines of Hart (1995) and Grossman and Hart (1986) for future work.

CONCLUSION

This thesis has contributed to the 'New Economic Geography' in several ways. On positive issues, it has proposed a relatively tractable model in which agglomeration forces stem from vertical linkages between firms. This model encompasses a well-known trade-cum-geography model, the so-called 'Footloose Capital' model of Flam and Helpman (1987) and Martin and Rogers (1995). This latter model does not display self-enforcing agglomeration forces and, as a direct consequence of this, is analytically very convenient to work with. As a result, it can be applied to various extensions. Conceptually then, its results can be thought of as giving a first approximation to a more complete model that allows for agglomeration forces. This more complete model is the topic of Chapter 1 of this thesis.

The core of the models of economic geography builds on Dixit and Stiglitz's (1977) monopolistic competition framework with trade/transportation costs a-la Samuelson (1952). The contribution of Chapter 2 was the completion of the characterisation of the equilibria of the simplest models among these. In this way, it filled a gap that even the monograph of Fujita, Krugman, and Venables (1999) let open. This chapter also showed how seemingly different models are deeply isomorphic.

New Economic Geography models are well suited to analyse situations in which comparative advantages of nations change endogenously as core parameters vary. Indeed, an extension of the vertical-linkages model of Chapter 1 that allows for decreasing returns to scale in the background sector shows how decreasing trade/transportation costs generate production specialisation and income divergence between countries with seemingly similar factor endowments and technology when trade/transportation costs are low; and how further trade integration generates convergence in the same variables. This result is reminiscent of Krugman and Venables (1995) and results from the tension that arises between agglomeration economies and cheap labour in less industrialised areas. The model of Chapter 4 then goes on showing how falling communication costs help integrate spatially segmented labour markets and, doing so, relieve one of the congestion cost associated with an industrial base. As a result, the existence of the industrial base is sustainable over a wider range of trade/transport costs. Firms co-locate so as to benefit from specialised intermediate inputs and outsource the most labour-intensive tasks in the less industrialised countries where wages are lower –or set up a subsidiary to that purpose.

Marshall (1890) and others have pointed to other forces driving firms to locate near to each other's. Labour market imperfections are thought to provide one prominent such force. It is often argued that thick labour markets provide firms with on average better skilled workers and that, in turn, these workers can expect a better match and higher wages in denser labour market. The argument often goes as follows (see e.g. Hesley and Strange 1990). The firms' skill requirements and the workers' abilities are heterogeneous. There are increasing returns at the firm level, so there is finite number of firms operating at equilibrium. As a result, there is a tension between the quality of the average match (which increases in the number of firms) and the scale economies associated with a small number of firms. Crucially, the set of possible skills is supposed to be finite and constant. Therefore, large cities relieve that tension and hence generate higher per-capita returns. The World is not like a single big city because congestion costs (e.g. commuting costs) eventually outweigh the advantages associated with a large labour market.

The final chapter of this thesis contributed to this literature by suggesting that the pooling of the labour market provided co-location economies of another kind. In effect, firms that use workers with similar skills compete for those workers when they locate near one another. This way, they can credibly commit to pay workers in accordance to their skills. This, in turn, provides the workers the incentives to acquire those skills in the first place. Under mild conditions, this benefits both the firms and the workers. Interestingly, the lack of market power on the side of the firm may reduce its own incentive to make specific investment. The argument is the flip side of the previous motive for co-location. Without market power, the firm cannot capture the full rents its investment generates. In other words, co-location of firms avoids a hold-up problem on the side of the workers but creates such a problem on the firm's side. This implies that agglomeration and dispersion forces are of the same nature. This might provide an explanation for why some industries are agglomerated whereas others are not, as reported in Duranton and Overman (2001), for instance.

This thesis has also dealt with normative issues, a rare exercise in the New Economic Geography. For a long time the New Economic Geography has been devoid of policy analysis, for its results are 'too stark to be true' (Neary 2001). The welfare implications of these models in the simplest form are stark indeed: it is always good for you to live in the region or the country in which the whole of industry clusters (see Chapter 1). Does this mean that no policy implication can be drawn out of these models? As it turns out, one can be optimistic about this issue. On the one hand, it appears that simple alterations of the simple models make them 'ambiguous enough to be true' (Baldwin et al. 2001).

On the other hand, some policy analysis can be carried out even in the simplest setting. If the market spatial allocation of industry cannot be Pareto improved (i.e. the size of the pie is optimal), then one may still ask, how does a given policy affect the spatial equilibrium? How is this policy chosen at equilibrium (i.e. how is the pie shared)? How does the political economy environment interact with the usual economic geography determinants of location? This was the approach taken in Chapter 3. There I showed how socio-economic differences across various populations within the same constituency affect the political equilibrium in a model of electoral competition. The result was that more homogenous populations are more attractive to politicians and, as a consequence, obtain regional policies that may help maintain an industrial base where the market would have decided otherwise.

The way ahead

A lot of potentially interesting topics related to the subject of this thesis have been put aside. On the descriptive side, it should be easier to work out extensions now that the functioning of traditional New Economic Geography models is better understood, and thanks to the recent apparition of analytically simpler and potentially richer models (Ottaviano, Tabuchi and Thisse 2000). Various studies have already taken a fresh look at such topics as trade policy, regional policy, and tax competition within a New Economic Geography framework (see Baldwin et al. 2001 for a comprehensive treatment).

From a political economy perspective it would be especially interesting to see how institutions affect the spatial landscape. Arguably, political systems in which small constituencies have a disproportionate representation –as is often the case in bicameral systems– should deliver a more even political landscape. Also, it is often heard that industries are more agglomerated in the United States than in Europe (a claim which is difficult to confirm or reject on sound empirical basis). Moreover, it is often argued in the same breath that the lack of labour mobility in Europe is responsible to this fact. In sum, it seems that people vote with their feet in the US and with their pencil in Europe. What is the effect of this on the spatial concentration of industries?

Cities tend to grow and decline with the sectors in which they are specialised. Also, a wide range of empirical studies suggests that sectoral protection is biased towards shrinking industries.[84] How does this affect both the urban landscape and the aggregate growth rate of the economy at equilibrium? These issues deserve at least as much interest as those tackled in Chapter 3. As a consequence, they will be among the objects of further research.

Firms in chapter 4 break up a vertically integrated process in search for cheap inputs according to the factor intensity of various stages of production process. But the model is silent about the internal organisation of the firm. In particular, does the firm outsource the tasks that used to be processed within the same location? Or does it merely set up a subsidiary in a foreign country? In other words, does the firm become a multi-location firm or does it buy these inputs on the spot market?

Likewise, the current view among economists about what is a firm is that it firms are a set of incomplete contracts that provide incentives to contracting agents to undertake the best possible actions under unforeseeable contingencies. The location choice of the firms of Chapter 5 plays a similar (albeit slightly different) role.

An obvious question follows: how do the location and the 'make vs. buy' decisions are related? How do they interact? Is there circular causality of sorts between these two decisional dimensions? Questions of this sort are the object of a budding literature body. They are also left for future work.

---

[84] Many explanations are consistent with the fact that protection and subsidies are skewed towards 'sunset industries' at the expense of booming sectors. Possibly, shrinking industries are better able to capture the lobbying rents since they do not fear these rents to be dissipated by further entry.

REFERENCES

Acemoglu, D. and R. Shimer (1999). 'Holdups and efficiency with search frictions.' International Economic Review 40: 827-49.

Anderson, S., A. De Palma, and J. Thisse (1992). Discrete Choice Theory of Product Differentiation (MIT Press, Cambridge MA).

Andersson, F. and R. Forslid (1999). 'Tax competition and economic geography.' CEPR WP 2220.

Baldwin, R. (1999). 'Agglomeration and endogenous capital.' European Economic Review 43(2): 253-80.

Baldwin, R. (2001). 'The Core-Periphery model with forward looking expectations.' Regional Science and Urban Economics 31: 21-49.

Baldwin, R. and R. Forslid (2000). 'Trade liberalisation and endogenous growth: A q-theory approach.' Journal of International Economics 50(2): 497-517.

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud (2001). Public Policy and Spatial Economics. Manuscript (Graduate Institute of International Studies, Geneva).

Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud (2002). 'The Core-Periphery model: Key features and effects.' In S. Brakman and B. Hejdra (eds.) The Monopolistic Competition Revolution in Retrospect, forthcoming (Cambridge University Press, Cambridge).

Baldwin, R. and P. Krugman (2001). 'Agglomeration, integration and tax harmonization.' HEI Working Papers in Economics HEIWP 01/2001.

Baldwin, R. and P. Martin (1999). 'Two waves of globalisation: Superficial similarities, fundamental differences.' In H. Siebert (ed.) Globalisation and Labour (Tubingen: J.C.B. Mohr for Kiel Institute of World Economics).

Baldwin, R., P. Martin, and G. Ottaviano (2001). 'Global income divergence, trade, and industrialization: The geography of growth take-offs.' Journal of Economic Growth 6: 5-37.

Baldwin, R. and F. Robert-Nicoud (2000). 'Free trade agreements without delocation.' Canadian Journal of Economics 33(3): 766-786.

Becker, G. (1983). 'A theory of competition among pressure groups for political influence.' Quarterly Journal of Economics 98: 371-400.

Berliant, M., R. Reed and P. Wang (2000). 'Knowledge exchange, matching and agglomeration.' Institute for Empirical Macroeconomics discussion paper 135 (Federal Reserve Bank of Minneapolis).

Besley, T. and S. Coate (2000). 'Issue unbundling via citizens' initiatives.' Mimeo (LSE and Cornell University).

Bismans, F. (1998). 'Croissances et convergences régionales en Belgique. Une approche macroéconomique et historique.' In: Treizième congrès des économistes belges de langue française: Portrait socio-économique de la Belgique (Université du Travail, Centre Interuniversitaire, Charleroi).

Bolton, P. and M. Whinston (1993). 'Incomplete contracts, vertical integration, and supply insurance.' Review of Economic Studies 60: 121-48.

Cadot, O., L. Röller, and A. Stephan (2001). 'Contribution to productivity or pork barrel? The two faces of infrastructure investment.' Mimeo (HEC Lausanne and WZB Berlin).

Campa, J. and L. Goldberg (1997). 'The evolving external organisation of manufacturing industries: Evidence from four countries.' Economic Policy Review (New York Federal Reserve Bank).

Combes, P. and G. Duranton (2001). 'Labour pooling, labour poaching, and spatial clustering.' CERAS working paper 01-03.

Coughlin, P. (1992). Probabilistic Voting Theory (Cambridge University Press, Cambridge).

Davis, D. (1998). 'The home market effect, trade and industrial structure.' American Economic Review 88: 1264-76.

Davis, D. and D. Weinstein (1996). 'Does economic geography matter for international specialisation?' NBER working paper 5706.

Davis, D. and D. Weinstein (1996). 'Economic geography and regional production structure: An empirical investigation.' European Economic Review 43(2): 379-408.

Dixit, A. and J. Stiglitz (1977). 'Monopolistic competition and the optimum product diversity.' American Economic Review 67: 297-308.

Dumais, G., G. Ellison and E. Glaeser (2002). 'Geographic concentration as a dynamic process.' Review of Economics and Statistics 84(2): 193-204.

Duranton, G. and H. Overman (2001). 'Localisation in UK manufacturing industries: Assessing non-randomness using micro-geographic data.' Mimeo (LSE).

Duranton, G. and D. Puga (2000). 'Diversity and specialization in cities: Why, where and when does it matter?' Urban Studies 37(3), 533-555.

Duranton, G. and D. Puga (2001). 'From sectoral to functional urban specialisation.' Mimeo (LSE and Toronto niversity).

Englmann, F. and U. Walz (1995). 'Industrial centres and regional growth in the presence of local inputs.' Journal of Regional Science 35:3-27.

Ewers, H. (1995). 'The locational impact of declining costs of transportation and communication.' In H. Giersch (ed.) Urban Agglomeration and Economic Growth (Springer-Verlag, Berlin).

Faini, R. (1984). 'Increasing returns, non-traded inputs and regional development.' Economic Journal 94: 308-23.

Feenstra, R. (1998). 'Integration of trade and disintegration of production in the global economy.' Journal of Economic Perspective 12(4): 31-50.

Felli, L. and K. Roberts (2001). 'Does competition solve the hold-up problem?' STICERD discussion paper TE/01/414.

Flam, H. and E. Helpman (1987). 'Industrial policy under monopolistic competition.' Journal of International Economics 22: 79-102.

Forslid, R. (1999). 'Agglomeration with human and physical capital: An analytically solvable case.' CEPR Working Paper 2102.

Forslid, R. and G. Ottaviano (2001). 'Trade and location: Two analytically solvable cases.' Mimeo (Stockholm University).

Freeman, R. (1995). 'Are you wages set in Beijing?' Journal of Economic Perspectives 9(3): 15-32.

Fudenberg, D. and J. Tirole (1991). Game theory (MIT Press, Cambridge MA).

Fujita M., P. Krugman and A. Venables (1999). The Spatial Economy (MIT Press, Cambridge MA).

Fujita, M. and J. Thisse (1996). 'Economics of agglomeration.' Journal of the Japanese and International Economies 10: 339-78.

Fujita, M. and J. Thisse (2002). Economics of Agglomeration. (Cambridge University Press, Cambridge).

Gao, Ting (1999). 'Economic geography and the development of vertical multinational production.' Journal of International Economics 48: 301-20.

Grossman, G. and E. Helpman (1994). 'Protection for sale.' American Economic Review 84, 833-850.

Grossman, G. and E. Helpman (2002). 'Integration vs. outsourcing in industry equilibrium.' Quarterly Journal of Economics 117(1): 85-120.

Grossman, S. and O. Hart (1986). 'The cost and benefits of ownership: A theory of vertical and horizontal integration.' Journal of Political Economy 94(4): 691-719.

Harris, R. (1995). 'Trade and communication costs.' Canadian Journal of Economics 28: 46-75.

Harris, R. (2001). 'A communication based model of global production fragmentation.' In W. Arndt and H. Kierzkowski (eds.) Fragmentation (Oxford University Press, Oxford).

Hart, O. (1995). Firms, Contracts, and Financial Structure (Clarendon Press, Oxford).

Hart, O. and J. Moore (1990). 'Property rights and the theory of the firm.' Journal of Political Economy 98(6): 1119-58.

Helleiner, G. (1981). Intra-firm trade and the developing countries (Macmillan Press, London).

Helpman (1998). 'The size of regions.' In D. Pines, E. Sadka and I. Zilcha (eds.) Topics in Public Economics (Cambridge University Press, Cambridge).

Henderson, V. (1985). Economic Theories and the City (Academic Press, Orlando FL).

Herrendorf, B., A. Valentinyi and R. Waldmann (2000). 'Ruling out multiplicity and indeterminacy: The role of heterogeneity.' Review of Economic Studies 67: 295-307.

Hesley, R. and W. Strange (1990). 'Matching and agglomeration economies in a system of cities.' Regional Science and Urban Economics 20: 189-212.

Hinich, M. (1977). 'Equilibrium in spatial theory: The median voter result is an artefact.' Journal of Economic Theory 16, 208-219.

Holmstrom, B. and J. Tirole (1989). 'The theory of the firm.' In R. Schmalensee and R. Willig (eds.) Handbook of Industrial Organisation, Vol. 1 (North-Holland, Amsterdam).

Jaffe, A., M. Trajtenberg and R. Henderson (1993). 'Geographic localisation of knowledge spillovers as evidenced by patent citations.' Quarterly Journal of Economics 58: 577-98.

Jones, R. (1965). 'The structure of simple general equilibrium models.' Journal of Political Economy 73: 557-72.

Jones, R. and H. Kierzkowski (1990). 'The role of services in production and international trade: A theoretical framework.' In R. Jones and A. Krueger (eds.) The Political Economy of International Trade (Basil Blackwell, Oxford).

Jones, R. and H. Kierzkowski (2001). 'A framework for fragmentation.' In W. Arndt and H. Kierzkowski (eds) Fragmentation (Oxford University Press, Oxford).

Karp, L. (2000). 'Fundamentals vs. beliefs under almost common knowledge.' Mimeo (University of California at Berkeley).

Kim, S. (1990). 'Labour heterogeneity, wage bargaining, and agglomeration economies.' Journal of Urban Economics 28: 160-77.

Kim, S. (1999). 'The rise of multiunit firms in US manufacturing.' Explorations in Economic History 36(4): 360-86.

Kind, H., K. Midlefart Knarvik and G. Schelderup (2000). 'Competition for capital in a lumpy world.' Journal of Public Economics 78(3): 253-74.

Krugman, P. (1980). 'Scale economies, product differentiation and the pattern of trade.' American Economic Review 70: 950-9.

Krugman, P. (1991a). Geography and Trade (MIT Press, Cambridge MA).

Krugman, P. (1991b). 'Increasing returns and economic geography.' Journal of Political Economy 99: 483-99.

Krugman, P. (1991c). 'History versus expectations.' Quarterly Journal of Economics 106 (2): 651-67.

Krugman, P. and A. Venables (1995). 'Globalization and the inequality of nations.' Quarterly Journal of Economics 110 (4): 857-80.

Krugman, P. and A. Venables (1997). 'Integration, specialisation and adjustment.' European Economic Review 40: 959-68.

Ledyard, J. (1984). 'The pure theory of large two-candidate elections.' Public Choice 44: 7-41.

Lindbeck, A. and J. Weibull (1987). 'Balanced-budget redistribution as the outcome of political competition.' Public Choice 52: 273-297.

Lösch, A. (1940). Die Räumliche Ordnung der Wirtschaft (Gustav Fischer Verlag, Jena).

Ludema, R. and I. Wooton (2000). 'Economic geography and the fiscal effects of regional integration.' Journal of International Economics 52(2): 331-357.

Markusen, J. (1995). 'The boundaries of multinational enterprises and the theory of international trade.' Journal of Economic Perspectives 9(2): 169-89.

Marshall, A. (1890). Principles of Economics (Macmillan, London, 8^{th} ed. published in 1920).

Martin, P. (1999). 'Public policies, regional inequalities and growth.' Journal of Public Economics 73: 85-105.

Martin, P. and C. Rogers (1995). 'Industrial location and public infrastructure.' Journal of International Economics 39: 335-351.

Martin, P. and G. Ottaviano (1999). 'Growing locations: Industry location in a model of endogenous growth.' European Economic Review 43(2): 281-302.

Martin, P. and G. Ottaviano (2001). 'Growth and agglomeration.' International Economic Review 42(4): 947-68.

Mas-Colell, A., M. Whinston and J. Green (1995). Microeconomic Theory (Oxford University Press, Oxford).

Matsuyama, K. (1991). 'Increasing Returns, industrialization, and indeterminacy of equilibrium.' Quarterly Journal of Economics 106 (2): 617-50.

Matsuyama, K. (1995). 'Complementarities and cumulative processes in models of monopolistic competition.' Journal of Economic Literature 33: 701-29.

Matsuyama, K. and T. Takahashi (1998). 'Self-defeating regional concentration.' Review of Economic Studies 65: 211-34.

Myerson, R. (1993). 'Incentives to cultivate favoured minorities under alternative electoral systems.' American Political Science Review 87(4): 856-869.

Neary, P. (2001). 'On hypes and hyperbolas: Introducing the new economic geography.' Journal of Economic Literature 39:536-61.

Olson, M. (1965). The logic of collective action (Harvard University Press, Cambridge MA).

Osborne, M. (1995). 'Spatial models of political competition under plurality rule: A survey of some explanations of the number of candidates and the positions they take.' Canadian Journal of Economics 27: 261-301.

Ottaviano (2002). 'Home market effects and the (in)efficiency of international specialisation.' Mimeo (Bocconi University).

Ottaviano, G. (2001). 'Monopolistic competition, trade, and endogenous spatial fluctuations.' Regional Science and Urban Economics 31: 51-77.

Ottaviano, G., T. Tabuchi and J. Thisse (2000). 'Agglomeration and trade revisited.' International Economic Review (forthcoming).

Ottaviano, G. and J. Thisse (2002). 'Integration, agglomeration, and the political economics of factor mobility.' Journal of Public Economics 83(3): 429-56.

Overman, H., S. Redding and A. Venables (2001). 'The economic geography of trade, production, and income: A survey of empirics.' CEPR working paper 2978.

Persson, T. and G. Tabellini (1992). 'The politics of 1992: Fiscal policy and European integration.' Review of Economic Studies 59: 689-701.

Persson, T. and G. Tabellini (2000). Political economics: Explaining economic policy (MIT Press, Cambridge MA).

Puga, D. (1999). 'The rise and fall of regional inequalities.' European Economic Review 43(2): 303-34.

Puga, D. and A. Venables (1996). 'The spread of industry: Spatial agglomeration in economic development.' Journal of the Japanese and International Economies 10: 440-64.

Puga, D. and A. Venables (1997). 'Preferential trading agreements and industrial location.' Journal of International Economics 43: 347-68.

Puga, D. and A. Venables (1999). 'Agglomeration and economic development: Import substitution vs. trade liberalisation.' Economic Journal 109 (455): 292-311.

Rauch, J. (1993). 'Does history matter only when it matters little? The case of city-industry location.' Quarterly Journal of Economics 108(3): 843-867.

Redding, S. and A. Venables (2000). 'Economic geography and international inequality.' Mimeo (LSE).

Romer, P. (1990). 'Endogenous technological change.' Journal of Political Economy 98: 71-102.

Rosenthal, S. and W. Strange (2001). 'The determinants of agglomeration.' Journal of Urban Economics (forthcoming).

Rotemberg, J. and G. Saloner (2000). 'Competition and human capital accumulation: A theory of interregional specialization and trade.' Regional Science and Urban Economics 30: 373-404.

Rubinstein, A. (1989). 'The e-mail game: Strategic behaviour under almost common knowledge.' American Economic Review 79: 385-91.

Samuelson, P. (1952). 'The transfer problem and transport costs: The terms of trade when impediments are absent.' Economic Journal 62: 278-304.

Saxenian, A. (1994). Regional Comparative Advantage: Culture and Competition in Silicon Valley and Route 128 (Harvard University Press, Cambridge MA).

Stolper, W. and P. Samuelson (1941). 'Protection and real wages.' Review of Economic Studies 21: 1-20.

Tabuchi, T. and J. Thisse (2001). 'Taste heterogeneity, labour mobility and economic geography.' Mimeo (CORE).

Venables, A. (1994). 'Economic integration and industrial agglomeration.' Economic and Social Review 26(1): 1-17.

Venables, Anthony J. (1996a). 'Equilibrium location of vertically linked industries.' International Economic Review 37: 341-59.

Venables, Anthony J. (1996b). 'Trade policy, cumulative causation, and industrial development.' Journal of Economic Development 49(1): 179-97.

Von Thünen, J. (1826). Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökomomie (Perthes, Hamburg).

Wilson, M. (1995). 'The office farther back: Business services, productivity, and the offshore back office.' In P. Harker (ed.) The Service Productivity and Quality Challenge (Kluwer, Boston).

ERRATA CORRIGE

p. 8, line 18: 'the existence of a limited *number* of cities'.

p. 8, line 23: 'What do we learn from those models *that* we did not know from neoclassical *models* of international trade?'

p. 14, last paragraph: 'let me ~~to~~ compare the model'.

p. 15, line 22: 'it exploits in a spatial setting the kind of *circular* causality'.

p. 17, Table 1: the table should be amended as follows:

Table 1. Four classes of Economic geography models

	Fixed firm size	Specific factor
No Agglomeration Forces	Krugman (1980)	**FC** (Footloose Capital) Flam and Helpman (1987)
Vertical Linkages	**CPVL** (Core-Periphery, Vertical Linkages) Krugman and Venables (1995)	**FCVL** (Footloose Capital, Vertical Linkages) This thesis
Factor Migration	**CP** (Core-Periphery) Krugman (1991)	**FE** (Footloose Entrepreneur) Forslid and Ottaviano (2001)
Factor Accumulation		**CC** (Constructed Capital) Baldwin (1999)

p. 20, last paragraph, line 2: 'a large *part* of the intractability'.

p. 25, line 17: 'The neoclassical paradigm then predicts that ~~each~~ locations or countries'.

p. 29, last paragraph, line 4: 'Now assume the contrary case in which the firm ~~that~~ has to make an unverifiable investment. In this situation the firm ~~it~~ does better'.

p. 36, line 20: 'the closed-form solution is $(1-s)>\mu(1-\alpha)(\sigma-1)/(\sigma(1-\alpha)+\alpha-\mu)$.'

p. 40, line 2: 'determine ~~*the*~~ where the system ends up *are* confirmed'.

p. 42, line 20: 'This implies that the e's *are* constant.'

p. 43, line 19: '(*In*)stability of the dispersed equilibrium'.

p. 50, last paragraph, line 1: 'capital ownership was uniform across ~~*each*~~ the populations of each *region*.'

p. 58, line 12: 'for ~~*all*~~ any value'.

p. 66, line 16: '(iv) $\rho = 1/\sigma$'.

p. 77, line 2: '$h(\phi) \equiv \chi + \phi^2 - (1+\chi)\phi^{1-\theta}$.'

p. 86, line 4: '~~*Some proofs are relegated to the appendix.*~~'

p. 87, line 6: 'her disposable income is $wL + \rho K - LT$.'

p. 94, line 16: 'whether living in 1 (as L_1 of them do) or in 2 (as L_2 of them do)'.

p. 105, line 7: 'this chapter analysed the impact *of* regional policy'.

p. 108, last line: 'and *exporting* it'.

p. 127, line 12: 'capital *owners lose*'.

p. 130, line 4: 'like *that* depicted in Figure 4-5.On this diagram, the break and sustain points are *plotted* as a function'.

p. 147, line 23: 'and their outside options all matter.'

p. 160, last paragraph: 'how *are* the location and the 'make vs. buy' decisions ~~*are*~~ related? How do they interact? Is there circular causality of sorts between these two decisional dimensions? Questions of this sort are the object of a budding *body of* literature ~~*body*~~.'