On the Location of Economic and Academic Activity

by

Karen Helene Midelfart Knarvik

A dissertation submitted for the degree of dr. oecon.

Acknowledgements

This thesis would never have seen the light of day without the encouragement of a very special dad, three patient supervisors, and the support and love of a caring husband.

I would have liked to thank my dad who passed away two months after the work on this thesis began, for motivating me to start working on a Ph.D. thesis, and for the incessant interest he always showed for my work. My sincere thanks also to my three supervisors, Jan I. Haaland, Victor D. Norman and Tony Venables for inspiring and valuable comments and discussions. I have learned a lot from each one of them, and am very grateful for their advice, encouragement, and help in practical matters.

Thanks also to my colleagues at the Centre for International Economics and Shipping and the Institute of Economics at the Norwegian School of Economics and Business Administration, and in particular to my two friendly "office mates" Ottar Mæstad and Jostein Tvedt, for providing a stimulating atmosphere for research. Jostein has co-authored one of the essays in this thesis, and I wish to express my gratitude to him for enjoyable and fruitful co-operation.

Financial support from the Research Council of Norway is gratefully acknowledged, and so are the efforts of Jan I. Haaland and Victor D. Norman in arranging the financing.

A special thanks also to Anne Liv Scrase for endless proof-reading.

Last, but not least, I am indebted to my mother and Truls for their support and thoughtfulness.

Bergen, October 1995

Karen Helene Midelfart Knarvik

Contents

Introduction

In times of falling and changing political and economic barriers, it appears as a challenge to attempt analysing how these altered barriers may influence the location of economic activity and people. "Will a completely new economic geography appear?", and "Will we observe increased international specialisation?" are questions that are constantly subject to discussion. So is the question of to what extent falling barriers imply that location actually ceases to matter. Motivated by the renaissance experienced during the last few years by the location theory and Alfred Marshall's concept of external economies of scale, this thesis therefore seeks to explore how the inclusion of a geographic dimension and agglomerative forces in traditional economic models can help us understand how changing trade arrangements and different levels of integration may affect the location of firms and people. Furthermore, we elucidate how new geographic structures influence trade volumes, trade patterns and welfare, and whether we may expect a divergent or convergent development internationally, as time passes by.

An overview of new and older location theory is found in the first essay, while the three succeeding essays propose theoretical models aiming to study further aspects of the location of economic activity within a microeconomic framework. In the fifth essay an empirical model is introduced. We try to test for the importance of an academician's location to his research productivity. Hence, we aim to elucidate whether the kind of forces determining the location and productivity of firms, are also at work as regards the location and productivity of academicians.

Essay 1, *"Economic geography and trade — A survey of the literature"*, is divided into three chapters. The first chapter gives a short introduction to traditional location theory, while the second chapter touches on what has been done on industrial localisation within the business strategy literature. The main part of the essay concentrates on what is often referred to as the "new" location theory or the "new" economic geography. We review the literature as well as the microeconomic tools most commonly applied. As the determinants of geographical concentration are central in the work that has been done on the "new" location theory, a discussion of different kinds of external economies encouraging agglomeration is provided. We elaborate on how policy may affect the location of economic activity when external economies enforcing agglomeration are present, and close the essay with a series of examples of topics that have been and can be analysed, applying the tools of the "new" location theory.

Essay 2, *"Uncertain trade costs and industrial localisation — Concentration versus diversification"*, was motivated by the situation currently faced by firms based in a couple of smaller European countries — countries that decided not to join the EU. The essay studies how uncertainty about future trade arrangements affecting firms' competitiveness and market access, may influence firms' investment behaviour and choice of location. Our analysis is based on a model of imperfect competition, and we show that risk aversion is not a necessary condition for firms to respond negatively to trade cost uncertainty. Despite being risk neutral, firms may be encouraged by uncertain trade arrangements to diversify their investments, and reallocate part of their production to abroad in order to secure market access to a large foreign market. We find that the number of national competitors committing to the same investment strategy, the degree of export dependence, as well as the perceived probabilities of "good" and "bad" future states respectively, are decisive for manufacturers' investment

behaviour and locational choice. What distinguishes this essay from the other essays, is the absence of external economies of scale and the presence of uncertainty. An interesting extension of the article would be to include in the model some kind of externality encouraging industrial agglomeration.

Essay 3, *"Technological spillovers, industrial clusters and economic integration"*, considers the effect of economic integration on industrial structure and trade patterns. We introduce a general equilibrium model of intra-industry trade with two sectors, one perfectly competitive and one imperfectly competitive. In the imperfectly competitive industry, external economies of scale create a tendency towards geographical agglomeration. Within the "new" location theory, different mechanisms leading to agglomeration have been explored. It has been demonstrated that inter-regional labour mobility as well as input-output structures between firms in imperfectly competitive industries, can serve to create linkages and hence agglomeration (see essay 1). But an alternative possibility, that so far has received less attention within the "new" location theory, is that of technological or knowledge spillovers — and it is on this mechanism we focus in essay 3. Localised knowledge spillovers are assumed to be present in the imperfectly competitive sector, and the vehicle for spreading information among firms is the intra-market mobility of skilled personnel. Economic integration implying a reduction in trade costs, is found to encourage agglomeration. The industry subject to external economies concentrates in one region, i.e., divergent industrial structures across countries are observed. Which country that ends up with the industry where spillovers are generated, will depend on the initial distribution of firms between countries, as well as on the level to which trade costs are reduced.

Knowledge spillovers caused by intra-industry labour mobility does also play a major part in Essay 4, *"International trade, technological development and agglomeration"* (co-authored by Jostein Tvedt). However, unlike in essay 3, perfect competition prevails in all markets. We focus on how localised external economies of scale created by knowledge spillovers, encourage industrial agglomeration, and cause uneven technological development internationally. Technological advantage and national industrial structures evolve over time, and we elucidate how the creation of clusters may contribute to explaining regional specialisation. Two cases are considered, one in which we allow for trade in goods only, and one in which there is trade in capital as well. Consistent with results in the "new" location theory, it appears that the critical size of a region regarding its ability to sustain an industrial cluster, is substantially reduced as trade is extended to include capital in addition to goods.

Finally, in essay 5, *"Does it matter where you are? The importance of being localised in the academic world"*, we aim at applying the theory of localised external economies of scale in an empirical model. While the preceding essays analyse the location of firms, in this paper, however, we focus on the location of academicians. In order to shed some light on possible determinants of individual and departmental performance, a cross-sectional study of 175 economists is conducted. In particular, we seek to test the hypothesis that an economist's environment, i.e., colleagues, matters to his research productivity. — And we find that there is reason to believe that localised knowledge spillovers may be present, not just in industries, but in academic milieus as well. Hence, in the same way as industrial clusters may increase firms' productivity, being in an efficient academic cluster may enhance an economist's performance.

Essay 1

Economic geography and trade — A survey of the literature

by

Karen Helene Midelfart Knarvik*

* I would like to thank Jan I. Haaland, Victor D. Norman and Anthony J. Venables for valuable comments.

Introduction

Most of the traditional trade theory rests on the simplifying assumption of constant returns to scale and perfect competition. Although it has been known for a long time, that the presence of increasing returns to scale and imperfect competition may have a significant impact on the results in trade theory, trade theories based on constant returns technologies have been predominant. The reason for this, is mainly the lack of analytical tools to model imperfect competition. As soon as the assumption of constant returns is relaxed, the *market structure* can no longer be ignored (except under special circumstances[1]), and it is necessary to know the precise nature of competition, and how it works.

The predominance of the competitive models lasted until about 1980. Due to the progressing theory of industrial organisation, trade economists were no longer limited to the models of constant returns — the assumption of perfect competition could be relaxed. During the 1980s a new trade theory based on imperfect competition emerged.[2]

It is important to see the development within the trade theory in connection with the general trends in economics. The importance of economies of scale has become gradually more accepted by economists during the last years.[3]. There is probably a close connection between this acceptance and the improvement of the analytical tools for the modelling of economies of scale. Beside the theory of trade, topics such as

---

[1] See Helpman and Krugman (1985), page 31.
[2] For contributions to this new trade theory see Helpman and Krugman (1985); Brander and Krugman (1983); and Krugman (1980) among others.
[3] See Helpman and Krugman (1985).

growth theory[4] (with keywords such as "increasing returns to scale" and "endogenous growth") and business cycles[5] have also experienced a ransformation due to the consideration of increasing returns to scale.

But although several consequences of the existence of technologies with decreasing costs have been analysed lately, there is one aspect of increasing returns that has been fairly neglected by economists: the impact of increasing returns on the location of economic activity in space.

A common feature of general microeconomics and trade theory is that, with few exceptions, the geography of economic activity is ignored. As long as the models rest on the simplifying assumption of constant returns to scale, such an ignorance is, in fact, not that serious. According to Krugman, the study of economic geography is actually condemned to be more or less excluded from economic analyses, as long as constant returns are assumed.[6] — Almost all competitive microeconomic models are consistent with the assumption of a "wonderland of no dimensions"[7]. But as soon as the assumption of constant returns to scale is relaxed, the study of industrial location becomes a natural consequence.

In other words: it was not possible for an economist to say very much about location until the theory of economics had been developed so far that the modelling of increasing returns to scale and imperfect competition were possible.[8] But having

---

4 See especially Romer (1986).
5 See Murphy, Shleifer, and Vishny (1989).
6 See Krugman (1991a), page 4.
7 See Isard (1956).
8 See Krugman (1991a), page 4.

reached so far, the study of location appears as an opportunity of making the models even more realistic, elucidating further interesting aspects of economics.

Nevertheless, in spite of the fact that location theory could give valuable contributions to economic models, little work has been done in this field until recently.[9] Currently, literature on economics and geography is emerging. Ideas concerning the role of location of economic activity, increasing returns and imperfect competition, multiple equilibria, the influence of chance and history, and endogenous growth are becoming popular. But, although different economists have been occupied with this subject lately, only a few path-breaking articles have been published on the specific topic of international economics and geography.[10] An important reason for economists to do geography is the fact that the lines between international economics and regional economics are becoming vague in important cases. The completing of the internal market is one example of this evolution.

The aim of this paper is to study what has been published on the topic of linking economics and geography, and especially in the field of geography and international economics. We shall first take a brief look at what has been done on industrial location within the theory of economic geography. Lately, also business strategy literature has been occupied with the importance of location. Especially the contributions of Porter[11] have received much attention. In the last part of chapter one we shall try to elucidate some of the thoughts that have appeared within the business strategy literature. The

---

[9] The study of location did not really catch the attention of economists until the end of the 1980s. Before this date, the only group of economists to be concerned with location, was the urban economists. See for instance Henderson (1974) and (1988).

[10] See especially Krugman and Venables (1990).

[11] See Porter (1990).

main part of this paper will concentrate on the issues of geography and microeconomics, and geography and trade theory.

1. Economic geography: The theory of industrial location

The study of localisation is an old one that has existed since von Thünen published his book "Der isolierte Staat" ("The isolated state") in 1826.[12] But the first systematic treatment of industrial location theory was due to Alfred Weber, and did not appear until 1909.[13] It focuses on the describing of factors that may have an impact on the allocation of economic activity, and develops principles for determining where various economic activities take place. Important are also the methods that have been developed for the calculation of optimal location for single facilities or systems of facilities. The applied tools are mainly those of operations research, and not those that are used in the common economic models.

Chapter one can be divided into two main parts: the first section summarise important features of traditional location theory, while concentrating on the determinants of industrial location. This section is primarily based on the chapters in Ihde on location theory.[14] Readers may find the references somewhat old, but the fact is that, although the books were written a long time ago, they still contain some of the most important elements of location theory.

---

[12] See von Thünen (1826).
[13] See Weber (1922).
[14] See Ihde (1984).

As for the second part of the chapter, we look at location from a single firm's point of view, i.e., locating as a strategic variable. This is a subject where economic geography and business strategy overlap each other to a great extent. The literature on geography and strategy can, in fact, be divided into an older and a newer one, of which the newer literature has been getting the most attention by far. The older literature is characterised by keywords such as: economic geography, logistics, and corporate strategy. Authors who have made important contributions here are Ihde and Timmermann among others.[15] In section 1.2 we try to convey some of their thoughts.

In paragraph 1.3 we turn to the "new" literature on economic geography and strategy. As Porter published "The competitive advantage of nations" in spring 1990, the theory of location experienced a kind of renaissance. For some reason (if we may take the liberty to be a bit impolite) the book became extremely popular among economists and business managers. Following what one might call the "Porter tradition", a number of economists have made contributions to the analysis of economic geography and strategy during the last few years.

What is the purpose of this chapter? We hope to be able to show that economic geography is actually concerned with many of the same topics as business administration and economics. The difference between them being the way they approach the problems, and which factors that are emphasised.

---

[15] See Ihde (?); Ihde (1984); and Timmermann (1983).

1.1 Determining the location of economic activity in space

When considering where to allocate its production, a firm has to take both advantages and disadvantages into account. It is quite common to divide the positive and negative factors into two groups: primary and secondary determinants. By primary we mean natural, given or exogenous factors, as for example natural resources. As a result of the economic activity itself, we get what is called the secondary determinants of location:[16]

a) Internal economies of scale
b) External economies of scale
c) Transport costs
d) Price of land

These determinants can be either agglomerative or deglomerative. Internal and external economies of scale are thought to have a positive effect on concentration. But as for transportation costs and the price of land, the opposite effect is presumed, i.e., they lead to delocalisation.

The last years have been characterised by decreasing an importance of internal economies of scale and transportation costs, and an increasing importance of external economies and the price of land. This tendency reflects the current development of the product markets; more specialisation than earlier is needed to satisfy the very different preferences of the purchasers, i.e., a larger product selection is necessary. At the same time, the product cycles are getting shorter. Transportation has become cheaper, and

[16] This classifying of the determinants is based on Lösch (1962). There are actually numerous ways of doing this classifying.

an increasing part of the transported goods are superior goods, which means they are not that sensible to additional costs (e.g. transport costs or more general transaction costs). Internal economies and transport costs certainly still matter, but compared with the two other secondary determinants of location, they do presently not receive that much attention.

Figure 1.1 provides a short and simple summary of the influence that the internal and external economies of scale, and transport costs have on industrial localisation.

Transport costs / Internal and external econ. of scale	= 0	> 0
= 0	Indifference	Moving towards delocalization (dispersion)
> 0	The production processes tend to concentrate	A kind of compromise between dispersion and concentration

Figure 1.1

The table displays the possible spatial arrangements of the economic activities in a region. A second interpretation of the table is how, depending on the magnitude of the three variables (internal and external economies of scale and transport costs), the decision of a single firm about how to allocate, changes.

Different authors have chosen to focus on different aspects (determinants) of localisation. von Thünen[17], and later also Alonso[18], have both developed location theories with the price of land as the central variable (see "land use theory"). They assume that the price discrepancies will lead to such a spatial distribution of firms, that the production processes with the minimum need for space will be occupying the most expensive sites.

According to the literature on location theory, the existence of external effects is especially interesting in relation to industrial localisation. Positive (external economies) as well as negative (external diseconomies) effects have received much attention. The kind of externalities that are mainly in focus, are those that arise from the distribution of economic activities in space, known as "localisation economies" and "urbanisation economies".[19] These may either be agglomerative (positive externalities) or deglomerative (negative externalities).[20] The two types of agglomeration economies can be defined as follows:[21]

Localisation economies are gained by firms in a single industry (or a set of closely related industries) at a single location and accrue to the individual production units through the overall enlarged output of the industry as a whole at that location. One

---

17 Related to his theory, von Thünen developed a geometrical model which is known as "die Thünensche Kreise (circles)".

18 See Alonso (1964).

19 These names and ways of grouping the spatial externalities were first used by Hoover (1937). He actually classified the agglomeration factors as:
- large scale economies (i.e. internal economies of scale),
- localisation economies, and
- urbanisation economies.

20 Positive effects are easily found, and some examples will be given later. But lately, also negative effects have become more discussed in the literature. Pollution due to economic activity, and lacking capacity of the infrastructure may, for example, be deglomerative.

21 See Dicken and Lloyd (1990).

example of such an agglomeration effect is the availability of qualified labour, i.e., a pooled labour market.

Urbanisation economies apply to all firms in all industries at a single location, and reflect external economies passed on to enterprises as a result of savings from the large-scale operation of the agglomeration as a whole. High local demand is, for example, a direct result of the concentration of economic activity in space.

1.2 The allocation problem of a single firm

With regard to localisation, each firm has to make two major decisions:

- Where to allocate, i.e. in which region or country.
- Concentration versus dispersion, i.e., to distribute the production or not, and to locate together with other firms or not.

It is often emphasised in the literature, that the nature of these decisions is clearly a strategic one, and that the decisions are part of a firm's strategic planning. What a company requires from its location, depends on what and how it produces. Choosing its location, a firm is concerned about close access to primary factors as well as to secondary factors.[22] Due to the economic evolution, the primary factors are, however, of decreasing importance.

As for the first decision, the political and economic environment in the regions (countries) in question are important. Considered are for instance: tax system,

[22] See section 1.1 for an explanation of primary and secondary factors.

infrastructure, supply of labour, access to capital, tariffs and trade policy. The political and economic environment and the structure of the economic activity, are but constantly changing. Hence, the ability of different regions and countries to meet the requirements of a firm, will be altered accordingly. As a result, the firms' allocation may need to be reconsidered. "The dynamics of allocation" can be illustrated through a simple example: the flagging out of ships.

Between 1982 and 1987, the flagging out of ships originally sailing under Norwegian flag, increased dramatically. Due to the advantages of the so called "flags of convenience", several shipowners chose to change the legal location of their ships. In this way they were able to lower their costs significantly. After 1987 the opposite trend was observed— the ships were "flagging in" again — not to the traditional Norwegian register (NOR), but to the new open ship register NIS (the Norwegian International Ship register). Altered conditions in Norway due to the establishment of NIS was the reason for this "flagging-in".

Characteristic of the present situation is, in general, that production processes continuously change location within one country and to other countries. As a consequence, the structure of the spatial arrangements is altered as well. The old "core locations" where a lot of industry was concentrated, lose their importance as new cores appear. The change of location tends to take place at an increased speed, i.e., the "location cycles", referring to how frequently a firm changes its location, are getting shorter.

As for the second decision, concentration versus dispersion, the firm is actually faced with two different decisions:

- Which one of the strategic allocation alternatives illustrated in figure 1.2, should the firm choose; concentrated or distributed production?
- What kind of location(s) should the firm choose for its production? Somewhere with a high degree of concentration of economic activity, the opposite, or a kind of compromise between the two extreme alternatives?

a) / b)	NO	YES
NO	Concentrating the production ("Standorteinheit") (1)	Separating the production of different goods (" Standortteilung") (3)
YES	Producing the same products at several locations ("Standortspaltung") (2)	A combination of 2 and 3 ("Standortdiversifikation") (4)

Figure 1.2

The letters in the figure refer to:

a): Spatial specialisation, i.e., similar goods are produced in one single location, and not together with other kinds of goods.

b): The same kinds of production processes are carried out in several places.

Crucial to the choice of whether to concentrate or distribute the firm's production are the kinds of goods produced, level of transport costs, the quality of the infrastructure, and the degree of economies of scale (see also figure 1.1).

In alternative (2) the transport costs are dominating, and the internal economies of scale are of less importance. Products are typically of less value and represent "old" goods, and the surroundings are often characterised by a very poor infrastructure. The opposite of (2) is true for alternative (3). The spatial specialisation observed here, is a result of the importance of decreasing average costs. Firms choosing this alternative usually produce high-value and complex goods in regions with a superior infrastructure. Number (4) is thought to be a result of risk aversion. The spatial distribution of a firm's production, is considered as a way of spreading (and restricting) risks, i.e., risks associated with the different locations.[23]

Concentrating on the western part of the world, the table also describes the historical evolution over time; beginning with (1), moving on to (2), (3), and (4). Hence, typical of the present situation is a kind of spatial diversification, where transport costs, as well as the advantages of internal economies of scale, are of decreasing importance.

Turning to the question of whether to cluster together with other firms or not, concentration may, obviously, have many advantages. A number of firms and/or industries located together at one place, lead to positive externalities, i.e., so-called agglomerative urbanisation and localisation economies. But, if the degree of concentration becomes "too" high, we are likely to observe negative externalities as well. The restricted capacity of the infrastructure system (e.g. roads and communication), could serve as an example of such negative externalities. Advantages and disadvantages from localisation should therefore be elaborated simultaneously. It is desirable to find the optimal degree of concentration and allocate production in a

[23] Such risks can be related to quality, delivery, exchange rates and transport.

region where this degree exists. Figure 1.3 illustrates what is meant by "optimal degree of concentration".

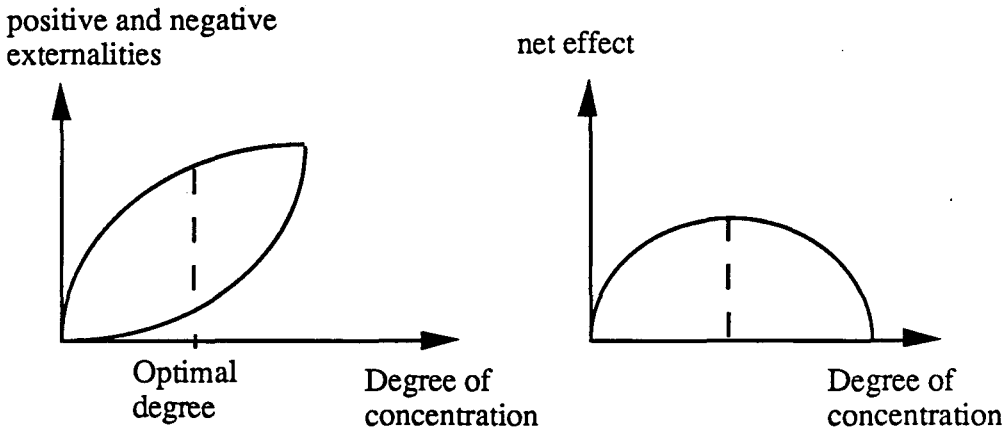


Figure 1.3

The general conclusion in literature on location as a strategic variable, is that access to locations that satisfy the requirements of a firm, is a necessary condition for a firm to succeed, as well as for the inter-regional and intertemporal optimal allocation of resources.

1.3 "Thinking Clusters" — the Porter approach

The idea that Porter was the first to "discover" the link between economic geography and business strategy, seems to be fairly widespread. However, Porter was not the first to consider the importance of geography for the strategic planning. One may guess that the lack of public awareness of other economists' earlier thoughts on the

subject, is either due to the new, simple, and appealing framework, as well as the way of approach used by Porter, or due to his superior marketing knowledge.

Nevertheless, writing a paper on geography and trade, it is not possible to ignore Porter. In this section, we concentrate on the thoughts of Porter published in "The competitive advantage of nations", and the contributions of Enright, Reve, Lensberg and Grønhaug to the subject.[24]

A question receiving frequent attention, is "why do some nations succeed and others fail in international competition"? Governments, business managers, and researchers have always been occupied with the reason for competitiveness. There have been numerous attempts to explain competitiveness. According to Porter, what the explanations are often conflicting, and neither has become *the* generally accepted theory. National competitiveness has been interpreted as:

- a macroeconomic phenomenon, due to exchange rates, interest rates and government deficits,
- a function of cheap and abundant labour,
- a result of the possessing of valuable natural resources,
- a result of government policy, including a protectional trade policy,
- a result of management practices, including labour-management relations.

None of these interpretations of competitiveness are sufficient to give a rationale for the competitive position of a nation's industries. In search of a convincing explanation, we probably need to revise the original question. As a first step we need to abandon the whole notion of a "competitive nation". — No nation can be

[24] See Enright (1991) and Reve, Lensberg, and Grønhaug (1992).

competitive in the production of, and net exporter of, all types of goods and services. Hence, seeking to explain competitiveness at the national level would be the wrong way to approach the subject. Instead, we have to focus on specific industries or industry segments. The study of a national economy reveals significant differences in the competitive success across industries. An international advantage is often concentrated in a single industry or industry segment.

Worth noting is also the fact that within many industries or industry segments, firms possessing international advantages are usually all based in just a few countries. The influence of a nation seems to apply to industries or segments, rather than to the actual firms. Instead of asking "why do some nations succeed and others fail in international competition?", we should probably ask "why do firms based in particular countries achieve international success in distinct segments and industries?".

The classical way to explain the international success of particular industries in a country, in terms of international trade, is the theory of comparative advantage. Comparative advantages are based on differences in access to factors of production, i.e., national differences in factor costs. But there is a general agreement on the fact that, the theory of comparative advantage does not explain the patterns of trade sufficiently, and also, at least for some industries, relies on quite unrealistic assumptions.

Porter concludes, that a new theory is needed in order to explain "which decisive characteristic of a nation that allows its firms to create and sustain competitive advantage in particular fields, i.e., the competitive advantage of nations".

A new theory of national competitive advantage

The answer to the question of the reason for a nation's prosperity, lies in four broad attributes of a nation: (1) factor conditions, (2) demand conditions, (3) related and supporting industries, and (4) firm strategy, structure and rivalry. To illustrate the determinants of national competitive advantage, the so called "Diamond" is often used (see figure 1.4).

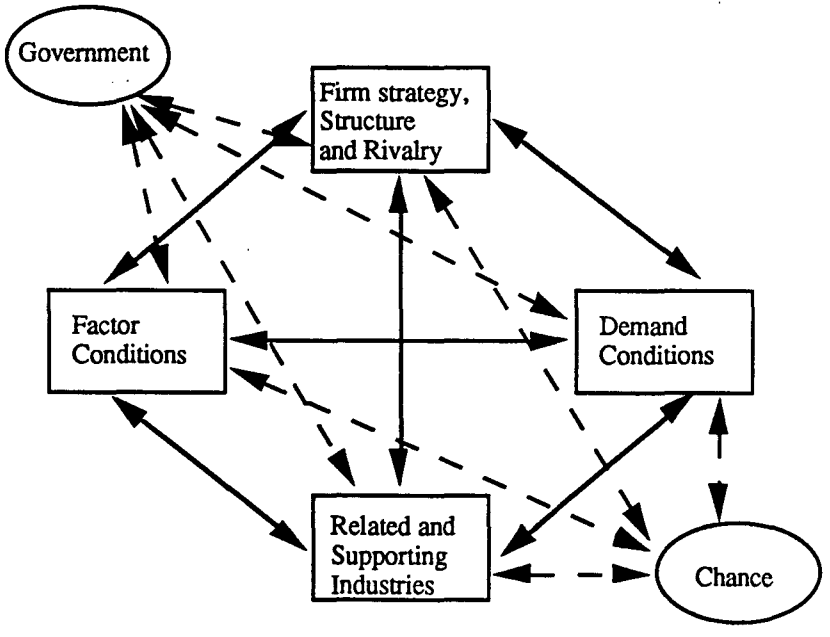


Figure 1.4

According to Porter, nations succeed in particular industries because their home environment is the most forward-looking, dynamic and challenging. The importance of the home nation is, in fact, more significant than ever. Companies benefit from having strong domestic rivals, aggressive home-based suppliers, and demanding local

customers. Emphasised is especially domestic rivalry, because it is thought to have the most powerful and stimulating effect on all the other points of the diamond.

The points of the diamond actually constitute a system. The main forces behind this system are domestic rivalry and geographic concentration. Geographic concentration magnifies the interaction between the four separate influences on a nation's competitive advantage. Crucial is, in other words, the observation that the development of competitive advantages is a highly localised process. The more localised the rivalry (for instance), the more intense, and therefore the better it is for firms' competitiveness.[25]

Industry clusters

Typical of the nature of the diamond, is the way it creates an environment that promotes industry clusters. In general, we do not only find one competitive industry in a country, on the contrary, we usually find several. These industries are somehow linked together. They influence each other, and receive mutual gains from their interdependence. The existence of a cluster leads to a continuously competitive upgrading and new entries. The clusters are characterised by geographic concentration, which can be observed in numerous countries, and is vital to how the cluster works.[26]

Geographic concentration reviewed

The phenomenon "geographic concentration" may be critical for how firms' competitiveness evolves. A geographic concentration of domestic rivals is quite often

[25] Examples of such geographic concentration are: Cutlery companies in Solingen, Germany, pharmaceutical companies in Basel, Switzerland, and Italian jewellery companies in Arezzo and Valenza Po.

[26] Examples mentioned by Porter are the Italian footwear industry and the German paper industry.

surrounded by suppliers and located in places with a concentration of demanding customers. This leads to a reinforcement of efficiency and specialisation, and further innovation and improvement. Moreover, the concentration of an industry or several related industries (clusters) attracts highly skilled labour. Today, communication and transportation costs are decreasing, and trade barriers are reduced. As a result, international competition is increasing and locational advantages with respect to industry innovations are becoming even more significant than earlier.

1.4 The need for a microeconomic foundation

In this chapter we have tried to convey some insight in to what economic geography and industrial location are all about. Locating, as an important part of a firm's strategic planning, has been elucidated, and ultimately, we have elaborated the relationship between geographic concentration, a firm's competitiveness, and the competitive advantage of nations.

Several appealing ideas and causal relations appear in the literature on which the previous sections are based. But nevertheless, we are left with the impression that the subject lacks a more formal foundation and structure. Through an incorporation of geography in traditional microeconomic models one may hope, that this main drawback of the subject of economic geography could be eliminated. Hence, in the next chapter we shall look at what has actually been done on geography within a microeconomic framework and by whom.

2. Microeconomics and geography — Towards a "new" location theory

In the preceding chapter we reviewed some of the main ideas of traditional location theory, and looked at how the inclusion of geographic aspects has influenced the subject of business strategy. Bearing this in mind, we turn to the issue of microeconomics and geography. A significant part of the literature covering this topic is often referred to as the "new location theory". We shall concentrate on this part of the literature, where the reasons for, and the dynamics related to, agglomeration of firms and consumers play a crucial role.[27] Especially the contributions of Krugman have received much attention, but there have also been articles by Abdel-Rahman, Fujita, Rivera-Batiz, Venables, a.o.[28] Chapter two introduces the microeconomic ideas and methods of the "new location theory". In chapter three the implications of the principles of the "new location theory" for trade theory are elaborated. We provide an overview of issues that may be (and have been) analysed by means of the tools of the "new" location theory.

---

[27] Models in this literature are characterized through endogenous agglomeration economies. There is a larger literature on exogenous agglomeration economies, but for reasons mentioned in 2.1.1, this will not be elaborated in this paper. Much work has also been done on issues such as spatial competition and price discrimination, where implications of space for price policies and competition are discussed. But since this work concentrates on aspects of space and economics that are not necessarily related to those of industrial concentration and spatial equilibria, it will not be elucidated in this paper.

[28] See Krugman (1991a); (1991b); and (1991c), a.o.

2.1 Increasing returns and geographic concentration

2.1.1 A rationale for imperfectly competitive models

Summarising what we have written about geographic concentration, there are two dimensions of concentration associated with firms: (1) A single firm can either choose to distribute its production, or to concentrate it in one location. (2) The firm can choose to locate together with several other firms, or to locate more or less alone. The other firms can either be similar and related firms, or associated with quite different industries. Hence, the firm can either try to exploit localisation economies and/or economies of urbanisation. The existence — and degree — of internal economies of scale are decisive for a firm's choice of number of production plants. Regarding where to locate, external economies of scale are crucial.

The central question is, how to include geography in microeconomic models (or how to link geography and economics), and to find the sufficient conditions for this. If we simply want to study the clustering of firms (e.g. in urban models), this could be compatible with the use of perfect competitive models. But such a "competitive approach" to agglomeration economies is contingent upon direct assumptions about the source and size of the external effects that cause the concentration. Because the explicit modelling of externalities is analytically awkward, and based on a more or less vague concept, it would be desirable to find another way of approaching the issue of agglomeration within the theory of economics.[29]

[29] See Krugman (1991a), page 5.

As to the strategic problem of whether or not a firm should divide its production, it should be obvious, that as long as constant returns to scale are assumed, there is, ceteris paribus, no reason why a firm should not distribute its production, in order to gain as close access as possible to all consumers. Accordingly, it may choose to divide its production between places with great industry concentration and more peripheral locations. The determining of an equilibrium may thus become quite difficult. Furthermore, it should not be ignored that in the real world the importance of scale economies internal to the firm is significant in several industries, and consequently of great importance to a firm's allocation decision.

Through the relaxation of the assumption of internal constant returns to scale, it is possible to provide a better reproduction of the real world; at the same time as agglomeration economies may be modelled without any direct assumption about external effects. The externalities leading to concentration emerge as a consequence of market interactions alone. In other words, a suitable model for economic geography is one of imperfect competition, in which the firms possess technologies with increasing returns to scale.

2.1.2 Modelling geographic concentration

We would now like to draw attention to a microeconomic model incorporating imperfect competition and geography - the "Core-periphery" model. The model is one of economic geography, developed by Krugman. It is published in his article "Increasing returns and economic geography" and in his book "Geography and trade",

in 1991.[30] First we present a sketch of the model, trying to convey some intuition of how it works, and then turn to the more formal version of the model.

However, it should be noted that Krugman was not the first to present a model including imperfect competition and spatial agglomeration. To our knowledge, Fujita's and Rivera-Batiz's articles in Regional Science and Urban Economics in 1988 constitute the path-breaking work in this field.[31] Until these papers were published, the models of economic geography developed by economists, were characterised through external-economy approaches like those of Henderson. Henderson let the centripetal force arise from assumed localised external economies in production, and let the centrifugal force be urban land rent.

The precise way in which Krugman let agglomerations emerge from the interaction between internal increasing returns (at firm level), transport costs and factor mobility, without making any assumptions about localised externalities presents an appealing and promising approach. There are two further advantages of Krugman's model: it provides a good overview of factors decisive for how and where firms locate, and

---

30 Krugman has later provided an extended version of the Core-periphery model (See Krugman (1992).), where the dynamics are modelled explicitly. Different from the original model are also the number of agglomerations that are modelled. In the former model, just one agglomeration is possible, but as for the latter model, the possibility of multiple agglomerations is considered. We shall concentrate on the first version of the Core-periphery model, because it is less complex than the later version, at the same time as it includes most the main points and results.

31 See Fujita (1988) and Rivera-Batiz (1988). The models povided by these authors may be described as follows: Fujita provides a spatial version of the Dixit-Stiglitz model: there is a continuum of households and firms; firms prefer to locate near households, and vice versa. Unlike Rivera-Batiz, Fujita introduces land consumption by firms, not just by households. The paper demonstrates that pure market processes based on price interactions alone can generate spatial agglomeration of economic activities. Additionally, welfare implications of the equilibrium configurations of cities are elaborated.
Rivera-Batiz applies the Dixit-stiglitz approach to both consumers and producers. An increased city size enlarges the variety of services available locally, and leads to a higher consumer utility. Increased city size is associated with an agglomeration of firms, which in turn raises the derived demand for intermediate goods, allowing the suppliers of these goods or services to specialize more. A basic source for agglomeration is therefore also the desire of the service industry to employ a variety of intermediates to achieve a higher degree of productivity.

explains how these may be modelled within the context of a microeconomic model. Second, it is simple. Yet, one should note that, there is one important aspect of economic geography, which is neglected in the model we present, namely geographic distance.

2.1.2.1 An informal approach to the issue of geographic structure

The main idea of the "Core-periphery" model is: "what determines the firms' choice of location, and what are the possible spatial equilibria?" In this model the importance of a "market potential" is emphasised, so that the best locations are those with good access to markets.

In the "Core-periphery" model increasing returns, transport costs, labour mobility, and demand are decisive variables. They represent the centripetal and centrifugal forces in the economy, and their interaction determines the degree of geographic concentration. Given strong economies of scale, each firm wishes to locate in one single place, instead of distributing its production geographically. Bearing this in mind, a firm will choose such a location that total transport costs are minimised, i.e., it will locate where the demand is high. Demand will obviously be high in places with a substantial concentration of manufacturers employing much labour. In other words, there is a clear interdependence between manufacturing concentration and demand concentration. The consequence of this interaction is that, once a manufacturing belt has been created, it tends to be very sustainable. It should be noted that, the reason for a "belt" to be established in a certain place does not necessarily have anything to do

with economics. The "belt's" location may just as well be determined by chance, history, or self-fulfilling prophecy.[32]

Our model is the following: in a country there are only two possible locations, East and West. Two kinds of goods are produced: agricultural and manufactured goods. There are two factors of production, each of which is specific to a particular sector. "Farmers" produce agricultural goods, and "workers" produce manufactured goods. We let π be the share of the total population engaged in manufacturing, and $1-\pi$ the share of the population engaged in agriculture. There is, in fact, a third factor of production, namely land — a location specific factor. This factor is, however, not included in our formal production function.

Farmers produce agricultural goods using the location-specific factor land, and as a result, the agricultural population is divided exogenously between East and West. The production is homogenous and characterised by constant returns to scale and perfect competition. "Farmers", as the immobile factor in the economy, provide the centrifugal force, which works against a concentration of manufacturers.

Manufactured goods can be produced in any location. The goods represent product varieties, and the number of varieties is large. The firms possess a technology with increasing returns to scale, and the market is characterised by imperfect competition. A firm can choose to locate in East *or* West, but will then be faced with transport costs due to sales in the other location. Its third alternative is to divide the production between East and West, but this will incur increased fixed costs related to the set-up of a second plant.

---

[32] We shall return to the importance of history, chance, and self-fulfilling prophecy later in this chapter, see 2.1.3.

The manufacturing labour force in a location is proportional to the amount of manufacturing that takes place in the location, and the demand in each location is proportional to the population living there.

A numerical example

We assume, that 60 percent of the labour force are farmers; that the farmers are divided equally between East and West; that total demand for manufactured goods is 10 units; that fixed costs of a manufacturing firm are 4; and that the transport cost per unit is 1.

A typical firm is faced with three different locational strategies: East only, fifty-fifty split between East and West, and West only. Depending on the locational strategies of all other firms, i.e., on the distribution of manufacturing labour (which implicates "demand"), the firm will solve its allocation problem.

It turns out that there are multiple possible equilibria:

1) If all other firms are located in the East, it will be most profitable for a single firm to locate its total production in the East as well.
2) If the firms are equally split between East and West, each firm will also want to split its production.
3) Ultimately, if manufacturing is concentrated in the West, each firm will concentrate its production in the West.

The three possible equilibria, contingent on the different locational strategies of all other manufacturers, represent the respective cost minimising location alternative for a

single firm. Below we illustrate graphically the possibility of more than one equilibrium.

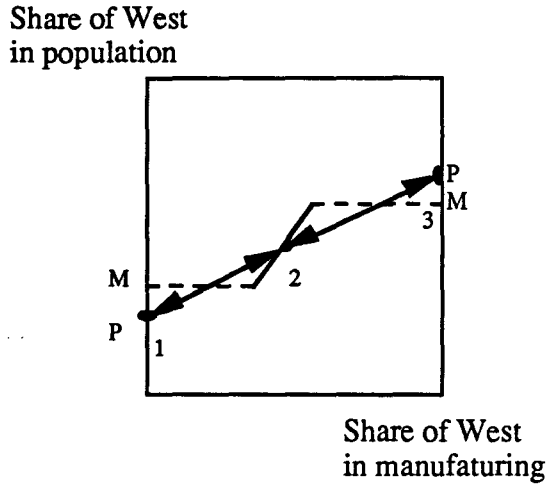


Figure 2.1

The horizontal axis measures the share of the manufacturing labour force employed in the West. As for the vertical axis, it measures the share of the total population living in the West. The two lines MM and PP can be described like this: both illustrate the interaction between the distribution of population and the distribution of manufacturing. MM shows how manufacturing depends on population, while PP explains how population is dependent on manufacturing.

Bearing in mind that π is interpreted as the share of the population engaged in manufacturing, the share of the population living in the West, s_N, can be written as a function of the share of workers employed in the West, s_M:

$$s_N = \frac{1-\pi}{2} + \pi s_M$$

This function is drawn as the line PP in the figure, and it is clear that it is increasing in s_M. Note that the part of the population that is situated in one of the locations can never be less than $(1-\pi)/2$.

The MM curve can be explained as follows: if one of the locations only has a very small part of the total population, it will not be profit-maximising to produce in this part of the country, due to the high fix cost incurred by the opening of a plant. In other words, production will only take place in the West, given that the population here has reached a certain size. Assuming that the fix cost is not too large relative to transport costs, an equal division of the inhabitants between West and East will lead manufacturers to produce in both locations.

The MM line illustrates the function $s_M(s_N)$. This function can be described mathematically in the context of s_N, fix cost (F), transport costs (t) and the sales of a typical manufacturing firm (x):

$$s_M = \begin{cases} 0 \text{ if } s_N < \dfrac{F}{tx} \\ s_N \text{ if } \dfrac{F}{tx} < s_N < 1 - \dfrac{F}{tx} \\ 1 \text{ if } 1 - \dfrac{F}{tx} < s_N \end{cases}$$

If $s_N xt < F$, it is cheaper to service the West from the East, and if $(1-s_N)xt < F$, it is cost-minimising to service the East from the West. The MM curve does, however, only represent the function $s_M(s_N)$ correctly, provided that the fix cost is not "too" high

relative to the transport costs. For this condition to be satisfied, $F < tx/2$ must be true. If not, it will never be profit-maximising behaviour to produce in both locations.

The numbers in figure 2.1 depict the different equilibria described above. Depending on where we start, each of these equilibria may be stable. The dynamics of locating are illustrated by the arrows, showing how we adjust towards an equilibrium. The original situation, in other words history, will determine which of the alternatives that will appear as *the* equilibrium.

A general condition for concentration

In the numerical example we got several equilibria, but it does not have to be like this. Depending on the magnitude of centrifugal and centripetal forces, there might be situations in which there is just a single possible equilibrium. If the agglomeration economies are not strong enough, we might get just one unique stable equilibrium where the production is split between East and West (see figure 2.2).

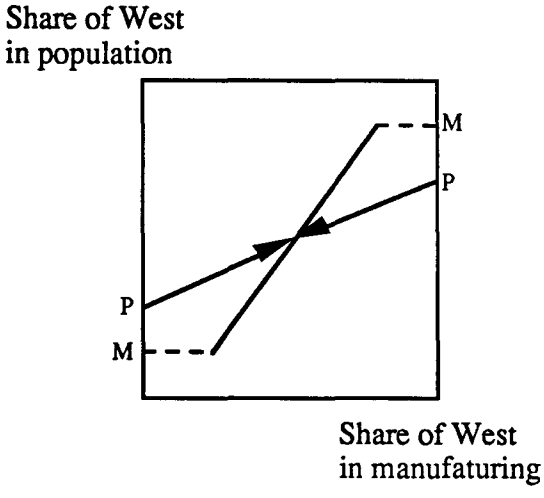


Figure 2.2

This is a situation where the fix cost of setting up a plant is relatively small compared to transport costs, and/or the share of the population that is immobile, is relatively large.

It is possible to derive a necessary condition for concentration of manufacturing in one location. Such a condition tells us how strong agglomeration economies have to be, to ensure the possibility of an equilibrium with concentration. Once a manufacturing cluster is established, the production will remain concentrated in the original location as long as the condition below is satisfied:

$$F > \frac{1-\pi}{2} tx.$$

If this condition is not met, history does not matter. Where the production was concentrated originally is of no importance. The distribution of agricultural activity alone, will determine the localisation of manufacturing. But it is obvious that the higher the fix cost (or the stronger the economies of scale), the lower the transport costs, and the larger the share of the population engaged in manufacturing, the more persistent is an established geographical concentration of manufacturers.

The endogeneity of transport costs

The quality of the transport network is another factor of importance when determining the geographical concentration (or dispersion) of an industry. One characteristic of transport networks is great regional divergence in quality and capacity. As a result, manufacturers located in one part of the country will have a much better market access than those located in another part, and their basis for choice of location is a totally different one.

The process of economic change

Another interesting aspect of the subject of geographical concentration, is the influence of economic change. Krugman elucidates two ideas related to the consequences of change in the environment or underlying conditions:

a) A geographical structure may be stable for a long time, but when it changes, the change often happens rapidly. The case of explosive change due to gradual changing of conditions, does not seem unrealistic.

b) The reasons for change are not necessarily of an economic nature. Change may just as well be caused by expectations — possibly self-fulfilling expectations. We will not elaborate the issue of expectations here, but return to this later in section 2.1.4.

In order to illustrate a change in geographical structure graphically, we make the assumption that farmers are originally unevenly divided between the two locations, East and West. Due to a migration of farmers from East to West, a reallocation process takes place over time, so that the earlier dominating location, say East, loses its dominance. As a result, the PP line shifts upwards as illustrated in figure 2.3, and we get the new line P*P*.

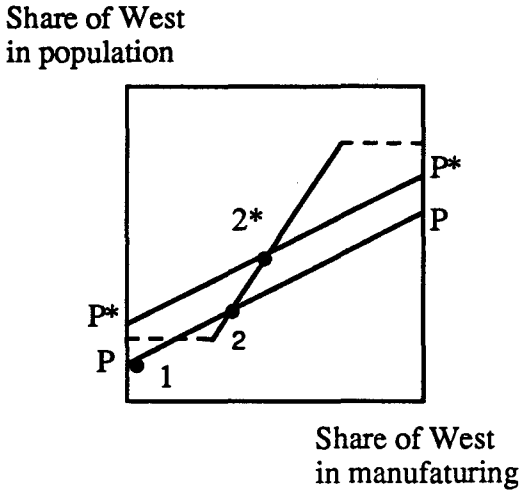


Figure 2.3

Although point 2 in the figure could have been a possible equilibrium, we assume that East, for some reason, has originally a dominating position and, as a result, all manufacturing is concentrated in the East. Point 1 depicts, in other words, our first equilibrium. As the dominance of the East is reduced, the population in the West increases. This leads to increased manufacturing in the West, which in turn leads to a further increase in western population. What we get is a cumulative process, and ultimately we get a new equilibrium like 2*.

An example of such a process, is the one that started with the discovery of oil in California in the late nineteenth century. Prior to this, the region had hardly had any industry at all, due to the lack of a local market, but the oil started an explosive process and turned California into a manufacturing centre.

2.1.2.2 A formal model

This model is based on the assumptions which we described in the preceding section. To close the model, we need some further assumptions: we let all the individuals in the economy share the same tastes, i.e., the same utility function. The function is a Cobb-Douglas function of the form

$$U = C_M^{\pi} C_A^{1-\pi} \tag{1}$$

with the consumption of agricultural goods (C_A) and a manufactures aggregate (C_M), determining the level of utility. U can also be thought of as a welfare function since the individuals possess the same preferences. We remember that π — the share of total expenditure received by manufacturers — also depicts the share of the population engaged in manufacturing. The function embodies a desirability for variety, and the manufactures aggregate, C_M, is defined as a constant-elasticity-of-substitution function of all the different potential product varieties,

$$C_M = \left[\sum_{i=1}^{n} c_i^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}. \tag{2}$$

There exists a large number of differentiated products, and due to the special kind of utility function, the elasticity of demand for any individual manufactured good is $\sigma > 1$, equal to the elasticity of substitution among the manufactured products.[33]

[33] Since the utility function is a Cobb-Douglas function, the elasticity of substitution between manufactured and agricultural goods is 1, and, in other words, less than the elasticity of substitution among the differentiated products. The intuition behind this, is that the differentiated products should be closer substitutes among themselves than a differentiated product and agricultural goods.

Consistent with the assumptions we made in the preceding section, we assume that farmers cannot move, and think of the geographical distribution of farmers as given. Similar to our numerical example, the agricultural labour force is split equally between East and West, with $(1-\pi)/2$ farmers in each location. The workers always move to the location that offers the highest wages.

The production of manufactured goods is subject to increasing returns to scale, which is reflected through a linear cost function in terms of the manufacturing labour (L_M):

$$L_{Mi} = \alpha + \beta x_{Mi} \tag{3}$$

$$TC \equiv wL_{Mi} = w\alpha + w\beta x_{Mi} \tag{4}$$

w= wage

x_i= produced amount of one product variety

The transportation of agricultural goods is assumed to be costless, a simplifying assumption that ensures that the farmer's wage rate and the price of agricultural goods are the same in East and West. But the transport of manufactured products from one location to the other, incurs costs. These costs take the form of Samuelson's "iceberg" costs, which means that only a fraction $\tau < 1$ of the transported goods arrives.

Market structure

Because there are economies of scale and the firms in the industry produce close substitutes, the manufacturers do not have any incitement to produce more than one

variety. Accordingly, the market structure is characterised by monopolistic competition.[34]

Solving the utility-maximising problem of a typical consumer and the profit-maximising problem of a representative firm, we get the demand for manufactured goods and the profit-maximising price. This price is as it would have been in a pure monopoly case, a constant mark-up over marginal costs.

Demand: $$p_i = \mathrm{U}'(q) \cdot \Phi'(c_i) \tag{5}$$

with $\Phi(c_i) = c_i^{\frac{\sigma-1}{\sigma}}$ and $q = \sum \Phi(c_i)$, where $\Phi(c_i)$ is a concave function.

Market price: $$p_i = \frac{\sigma}{\sigma - 1}\beta w \tag{6}$$

Free entry is assumed, implying zero-profit and price equal average cost in the industry in the long run:

Zero-profit condition: $$(p - \beta w)x = \alpha w. \tag{7}$$

The equations (5), (6), and (7) characterise the partial equilibrium of the market for manufactured products. Using these three equations, we find the equilibrium degree of economies of scale in the industry, the equilibrium output of a typical manufacturing firm, and the equilibrium number of manufactured goods (firms) in a region:

---

[34] The model we shall use is the Spence-Dixit-Stiglitz model of monopolistic competition, see Dixit and Stiglitz (1977).

Economies of scale can be measured by the ratio of average costs to marginal costs, which gives us $\sigma/(\sigma-1)$, implying that σ can be interpreted as an inverse index of the importance of economies of scale.

Output of a representative firm:

$$x_i = \frac{\alpha(\sigma-1)}{\beta}. \tag{8}$$

Number of varieties produced (firms) in one region with a manufacturing labour force of:

$$n = \frac{L_M}{\alpha+\beta x} = \frac{L_M}{\alpha\sigma}. \tag{9}$$

The persistence of a core-periphery pattern

Using the developed model it is desirable to derive under which conditions an established core will remain *the* place of concentration. The demand externalities from the larger market, and the fact that each manufacturer wishes to concentrate his production in one location due to increasing returns to scale, encourage agglomeration. The convex preferences of the demanders (leading to the concavity of the Φ function) embody the desire for variety, which is another centripetal force that should not be ignored: workers want to have access to all manufactured goods, not just to those produced by themselves. Ultimately, there is a deglomerative force at work, namely the fact that the manufacturers also want to service the more peripheral agricultural location.

In order to derive a criterion for the sustainability of a core, we derive what the total income and sales are in the two locations.

Income

It was mentioned earlier that the model ensures equal wages for farmers independent of their location. Due to the assumption of zero-profits and the assumption that the share of expenditure used on manufactured goods is proportional to the part of the population employed in manufacturing, workers and farmers will also have equal wage rates.

Assuming that all manufacturing is concentrated in one location, for instance in the East, the income distribution looks like this:

Income of East $$Y^E = \frac{1+\pi}{2} \tag{10}$$

Income of West $$Y^W = \frac{1-\pi}{2} \tag{11}$$

Sales

Given the situation where all manufacturing is located in the East, the sales of a representative Eastern firm will be:

$$s^E = \frac{\pi}{n}. \tag{12}$$

If a manufacturer wants to open a plant in the West, he will have to pay the workers a higher wage than in the East. Such a compensation is necessary because the other products, for which the individuals have preferences, will have to be imported from

Eastern firms and will incur transport costs. Only a fraction τ of the shipped goods will reach the West, i.e., the Eastern manufactured goods will be $1/\tau$ times more expensive in the West. The implication of this is an overall price index (= a geometric average of manufactured and agricultural goods) in the West, that is $\tau^{-\pi}$ times as high as in the East. Accordingly, a manufacturer in the West would have to pay his workers a wage $\tau^{-\pi}$ times higher than in the East, resulting in a higher price for goods manufactured in the West:

$$p^W = p^E \tau^{-\pi}. \tag{13}$$

The prices the consumers have to pay for imported goods may differ from the producer price because of the transport cost. The relative consumer price of a Western good to a consumer in the East, will be $p^w/\tau p^E$, and the relative consumer price of a Western good to a Western consumer, will be $\tau p^w/p^E$.

Remembering that σ depicts the elasticity of demand, a one percent increase in the price of the Western manufactured good will lead to a σ percent decrease of consumption of this good. But due to the valuation effect caused by the rise in the relative price, the expenditure (or sales value)will only be reduced by $(\sigma - 1)$ percent. Using this information, it is possible to derive the value of the sales of a manufacturing firm choosing to allocate in the West instead of in the East:

$$s^W = \frac{\pi}{n}\left[\frac{1+\pi}{2}\left(\frac{p^W}{p^E\tau}\right)^{-(\sigma-1)} + \frac{1-\pi}{2}\left(\frac{p^W\tau}{p^E}\right)^{-(\sigma-1)}\right]. \tag{14}$$

The value of the sales of a single Western firm relative to that of a firm located in the core (East), follows by dividing equation (14) by equation (12) and substituting for p^w by using equation (13):

$$\frac{s^W}{s^E} = \frac{1+\pi}{2}\tau^{(1+\pi)(\sigma-1)} + \frac{1-\pi}{2}\tau^{-(1-\pi)(\sigma-1)}. \tag{15}$$

It may seem profitable for a firm to defect and locate in the West if $s^W/s^E>1$. But this is in fact, not quite true. We also have to consider the fixed cost associated with the opening of a plant. Because of the special kind of cost function, this cost will also be affected by the higher wages in the West, and it will only be profitable to defect if

$$\frac{s^W}{s^E} > \tau^{-\pi}. \tag{16}$$

We define a new variable, K, whose value can tell us if it is profitable to start production in the peripheral location:

$$K \equiv \frac{\tau^{\pi} s^W}{s^E} \tag{17}$$

$$K = \frac{\tau^{\pi\sigma}}{2}\left[(1+\pi)\tau^{(\sigma-1)} + (1-\pi)\tau^{-(\sigma-1)}\right] \tag{18}$$

The criterion for the sustainability of the core-periphery pattern and equilibrium is $K<1$. If $K>1$, it is relatively more profitable for a firm to open a plant in the periphery than in the core location. The variables determining the value of K are: the transport costs, the share of income spent on "footloose" production (manufacturing), and the degree of economies of scale in production.

Determining the spatial equilibrium

The equation for K defines a boundary. We aim to evaluate the properties of K around 1, in order to see in what ways the three parameters *transport costs, degree of economies of scale, and share of income spent on manufactured goods,* influence the persistence of a core-periphery equilibrium.

The share of income spent on manufactured goods can be written as:

$$\frac{\partial K}{\partial \pi} = \sigma K \ln(\tau) + \frac{\tau^{\sigma\pi}}{2}\left[\tau^{(\sigma-1)} - \tau^{-(\sigma-1)}\right] < 0. \tag{19}$$

An increase in the share of expenditure related to manufactured goods has an unambiguously negative effect on K. It follows that an increase in p helps to sustain the core-periphery pattern: the larger the share of income spent on manufactured goods, the lower are the sales of a defecting firm; the smaller is K, and the more persistent is the core. The fact that an increased π leads the workers to demand a higher wage premium in order to be willing to move to the periphery, at the same time as the "home market effect" in the East gets stronger, explains why an increase in π has the described impact on K and on the sustainability of the core-periphery pattern.

Transport costs:

Equation (18) tells us that if transport costs are equal to zero (τ=1), then K=1, and the question of location is irrelevant. But if the transport costs are very high (τ approaches zero), then K approaches

$$\lim K_{\tau \to 0} = \frac{1}{2}(1-\pi)\tau^{1-\sigma(1-\pi)}. \tag{20}$$

Unless the economies of scale are very large (σ very small) or the share of expenditure on manufactured goods is substantial, high transport costs (low τ) will incur a $K>1$. More precisely, we have, that as long as $\sigma(1-\pi)>1$, for low τ values K will become arbitrarily large, and the sustainability of the core will decline. In general, the effect of transport costs on the nature of the spatial equilibrium is, however, ambiguous:

$$\frac{\partial K}{\partial \tau}=\frac{\sigma\pi K}{\tau}+(\sigma-1)\frac{\tau^{\sigma\pi}}{2}\left[(1+\pi)\tau^{\sigma-2}-(1-\pi)\tau^{-\sigma}\right]. \qquad (21)$$

As τ approaches 1 (transport costs lose importance), the second term of the expression approaches $\pi(\sigma-1)>0$, and since the first term is always larger than zero, $\partial K/\partial\tau$ will be positive for values of τ close to 1. In figure 2.4 below, the K equation is illustrated as a function of transport cost. The two other parameters are kept constant, and as values for σ and π we have chosen σ=4 and π=.3. These values satisfy the above assumption, $\sigma(1-\pi)>1$, which is crucial to the influence of transport costs on localisation. Given that this condition is not met, i.e., $\sigma(1-\pi)<1$, this has an important economic interpretation: the economies of scale are so strong and the share of manufacturers relative to farmers so large, that no matter how large the transport costs are, it will never be profitable to open a plant in the peripheral region (West). Explained in another way: workers receive a higher real wage in the region with a manufacturing cluster, even though the costs of transport are infinite. In other words, $\sigma(1-\pi)>1$ is necessary for transport costs to be of relevance to localisation, and we shall therefore keep this assumption for the rest of the section.

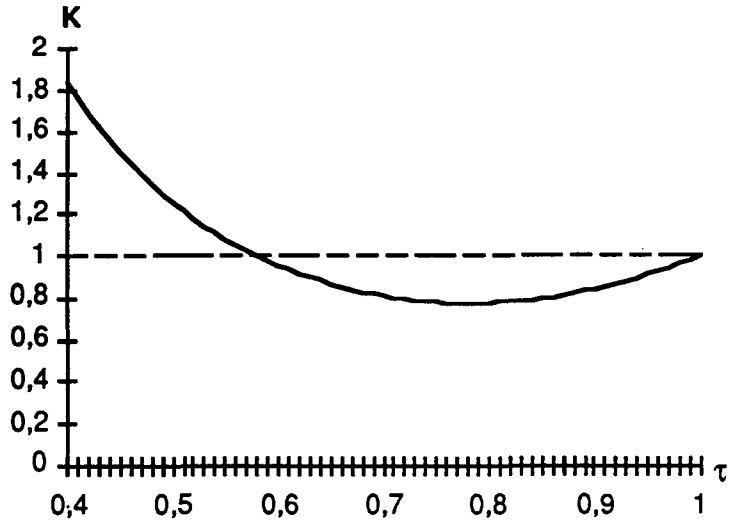


Figure 2.4

Studying the figure above, we can observe that for small values of τ (i.e., high transport costs), $K>1$, and it will be profitable to defect and start production in the periphery. For some critical value of τ, K falls below the boundary (=1), and a core-periphery pattern is sustainable. We note that when τ reaches the critical value, depicting the boundary between dispersion and concentration, $\partial K/\partial\tau$ is negative. An increase in τ (equivalent to a reduction of transport cost) has a positive effect on concentration. Figure 2.4 further illustrates, that the forces for agglomeration are greatest at the intermediate level of transport costs.

Economies of scale:

Remembering that σ is interpreted as an inverse index of the importance of scale economies, it is possible to derive how economies of scale affect K:

$$\frac{\partial K}{\partial \sigma} = \ln(\tau)\left\{\pi K + \frac{\tau^{\pi\sigma}}{2}\left[(1+\pi)\tau^{\sigma-1} - (1-\pi)\tau^{-(\sigma-1)}\right]\right\} \qquad (22)$$

Comparing equation (22) with equation (21), it becomes obvious that they will always have opposite signs. Given that $\partial K/\partial \tau < 0$, which is the case when τ approaches the critical value, then $\partial K/\partial \sigma > 0$, and a higher elasticity of substitution (implying a less significant degree of economies of scale) has a negative effect on the sustainability of a core-periphery pattern.

Based on the expression for K, it is possible to derive a boundary in a π, τ space, which marks the values at which a firm is just indifferent between staying in the location with concentration or opening a plant in the periphery.(see figure 2.5). The equilibrium in an economy inside this boundary will never be characterised through a concentration of industry in just one region. But if an economy lies outside the boundary, a core-periphery pattern will develop.

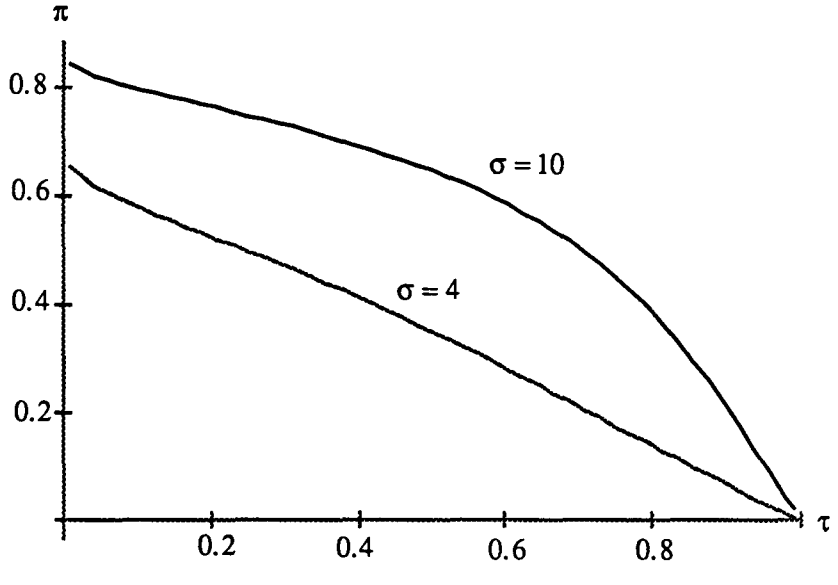


Figure 2.5

Knowing that as τ approaches the critical value (and takes on this value), $\partial K/\partial \tau$ is negative, it is clear that the slope of the boundary in the τ, π space will also be negative:

$$\frac{\partial \pi}{\partial \tau} = -\frac{\partial K/\partial \tau}{\partial K/\partial \pi} < 0 \tag{23}$$

A change in the degree of economies of scale will shift the boundary in or out, depending on whether σ has increased or decreased:

$$\frac{\partial \pi}{\partial \sigma} = -\frac{\partial K/\partial \sigma}{\partial K/\partial \pi} > 0 \tag{24}$$

This interdependence between π and σ implicates that, if the importance of scale economies is reduced (σ increased), the boundary shifts out, and it becomes more difficult to sustain a core-periphery pattern.

Conclusion

In an economy characterised by high transport costs, insignificant economies of scale, or a small share of footloose manufacturing, the distribution of manufacturers will be determined by the distribution of location-specific, immobile labour (farmers). But if the scale economies are strong, transport costs low, or the share of the population engaged in manufacturing comparatively high, then manufacturing will probably take place in the region which, for some reason, gets a head start.

The number, type, and stability of spatial equilibria depend on the parameter values. Parameter changes can lead to marginal as well as "catastrophic" changes in a spatial

equilibrium. Depending on the original situation, an alteration of parameters can induce only marginal changes of the equilibrium, but it may also incur a change from concentration to diversification (or vice versa).

2.1.3 The importance of history and expectations in determining an equilibrium

A typical characteristic of a model with increasing returns and positive externalities is the possibility of multiple equilibria. Crucial is the nature of the ultimately established equilibrium, and how it is determined. The traditional view has been that, in a situation with multiple possible equilibria, history determines the equilibrium. But during the last few years, the belief that expectations or self-fulfilling prophecy have an important impact on the establishing of an equilibrium, has also been pronounced by economists.[35] The rationale for considering the role of expectations can be explained as follows: the idea that history determines the equilibrium goes back as far as Marshall. He assumed that resources move in response to differences in current earnings. But due to costs of adjustment, resources move gradually, and the owners of resources will therefore not only be interested in current, but also in future earnings. In the presence of some kind of externality, future returns depend on the factor allocation of other people, which in turn depends on their expectations of future returns. There should, in other words, be a potential for self-fulfilling prophecy: if everybody believes that the economy will end up in equilibrium 1, then it will; and vice versa, if everybody believes that the final equilibrium will be equilibrium 2, then it will.

---

[35] See for instance Baldwin and Lyons (1991); and Krugman (1991c).

A short illustration: given that all the workers expect the other workers to move from East to West, they will assume the real wage to be higher in the West than in the East, and will therefore decide to reallocate to the West. Due to some head start in industrialisation in the East, the real wage is originally higher in the East than in the West. Despite this fact, the expectations about the decisions of the other workers may lead to a migration from East to West, and as a result, to a higher real wage in the West.

Most economic models have tended to either consider history *or* expectations, but there will clearly be cases where both determinants are relevant. The model presented below was developed by Krugman (see Krugman (1991c)), and is an attempt to explain what the respective roles of history and expectations are.

A simple formal model

The assumptions of the model: two regions; one mobile factor of production, namely labour (L); the difference between the real wages in regions 1 and 2 is increasing in the share of the total labour force located in region 1:

$$w_1 - w_2 = \alpha(L_1 - \bar{L}) \tag{25}$$

with $\bar{L} = L/2 = (L_1 + L_2)/2$.

Labour is assumed to migrate towards the region offering the highest real wages, but moving from one location to another is not costless. This cost of migration can be thought of as some kind of "costs of adjustment", and is convex in the aggregate rate of moving.

Consistent with the assumptions of the model, the total income of all workers will be equal to their wages minus the costs related to migration:

$$Y = w_1 L_1 + w_2 L_2 - \frac{1}{2\gamma}\left(\dot{L}_1\right)^2. \tag{26}$$

Let r be the exogenous discount rate. Given that the workers are forward looking with perfect information about the future real wages in both regions at any time s, the present value of being in location 1 instead of in location 2 is

$$q(s) = \int_s^{\infty} [w_1(t) - w_2(t)] e^{-r(t-s)} dt. \tag{27}$$

Differentiating q(s) we get

$$\begin{aligned} \dot{q} &= rq - (w_1 - w_2) \\ &= rq - \alpha(L_1 - \bar{L}). \end{aligned} \tag{28}$$

The rate of workers moving from one region to another will be such that the gain from moving equals the marginal costs of moving

$$\dot{L}_1 = \gamma q. \tag{29}$$

Equation (28) and (29) define a dynamic system, which illustrates the dynamics of migration. Figure 2.6 shows how the system works, and that an equilibrium where the labour is divided evenly between the two locations is unstable. Over time, the whole labour force will become concentrated either in region 1 or in region 2.

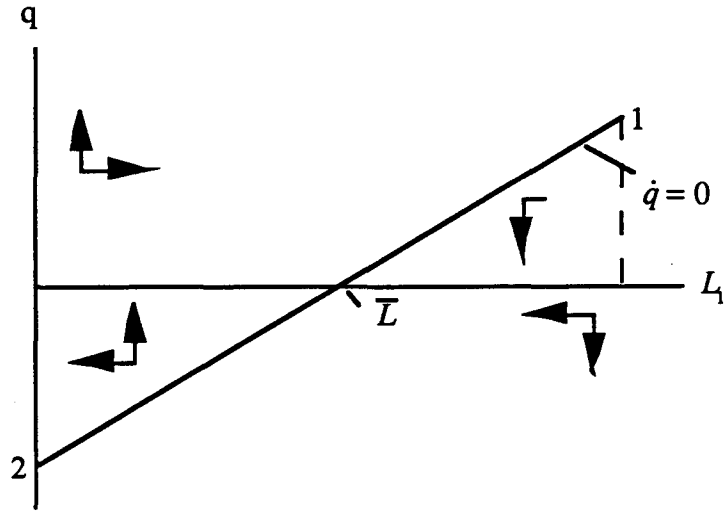


Figure 2.6

As for the nature of the local dynamics around the unstable equilibrium, this can be established by solving the differential equations (28) and (29) with standard techniques. The eigenvalues of the dynamic system are then given by:

$$\lambda_1 = \frac{r+\sqrt{r^2-4\alpha\gamma}}{2}$$
$$\lambda_2 = \frac{r-\sqrt{r^2-4\alpha\gamma}}{2}. \qquad (30)$$

There are two alternative solutions: either both roots are positive, or both roots are complex. If the roots are positive, the paths leading to the long-run equilibrium are illustrated in figure 2.7. The labour force will, in the long run, be concentrated in the region that for historical reasons gets a head start, i.e., the region which initially has the largest share of workers. Expectations cannot change this evolution by

outweighing the role of history. The only way expectations might influence the long-run equilibrium, is through reinforcing the importance of historical events.

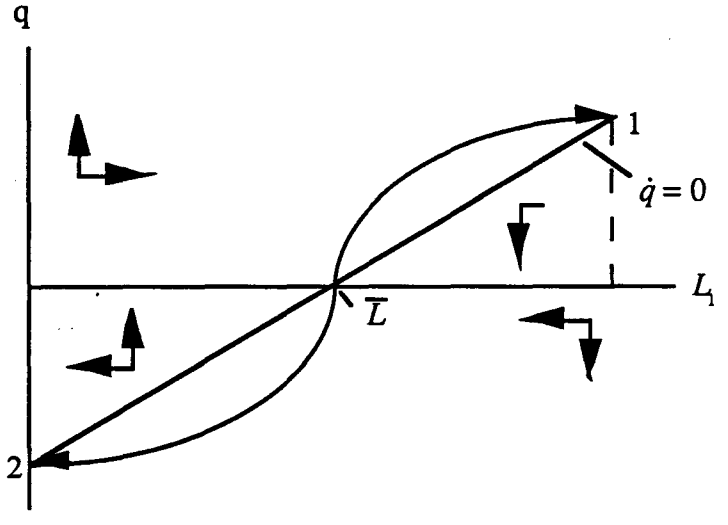


Figure 2.7

But given that the roots are complex, the importance of the determinant "expectations" should not be neglected. In this case, i.e. when $r^2 < 4\alpha\gamma$, the paths leading to a long-run equilibrium will spiral outward from the middle, overlapping each other (see figure 2.8). If L_1 initially takes on a value in the region where the spirals overlap, $\left(L_1^A, L_1^B\right)$, then there exists at least one path of self-fulfilling expectations leading to either long-run equilibrium (i.e., concentration in 1 or 2).[36] — The nature of the equilibrium is undetermined: although region 1 starts with more than fifty percent of

[36] In a comment to Krugman (1991c), Fukao and Benabou (1993) argue that Krugman's presumption about the terminal condition is not correct, and that the correct formula for $q(s)$ is not the equation (27), but $q(s) = \int_s^T [w_1(t) - w_2(t)] e^{-r(t-s)} dt$. From this follows that the true equilibrium paths are different from Krugman's, and that his analysis of the width of the "overlap" also needs to be revised. According to Fukao and Benabou the "overlap" is much narrower than Krugman argues.

the total labour force, it is possible that region 2 ends up as the core, if all workers expect this to happen.

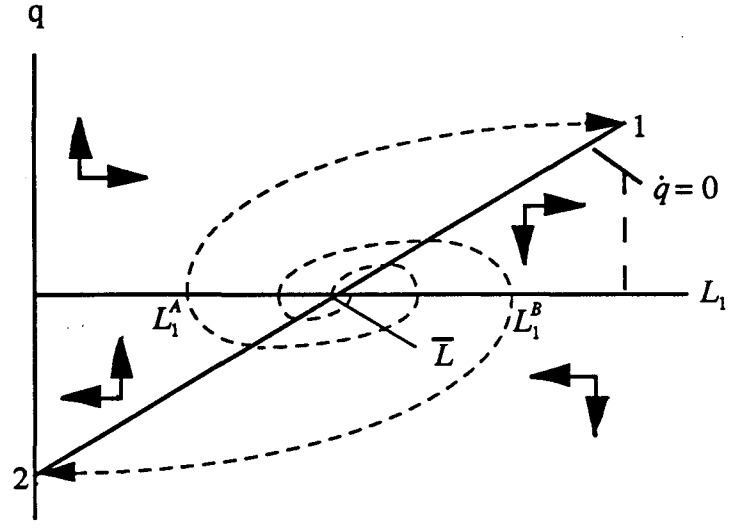


Figure 2.8

The factors determining the existence and the extension of an "overlap" tell us whether self-fulfilling expectations matter or not to the evolution towards an equilibrium. By investigating the criterion for an overlap, which is $r^2 < 4\alpha\gamma$, it appears that if

- the discount rate is low;
- increasing returns are strong, so that an expected future shift in population distribution moves the real wage differential quickly;
- the external economies, defined as the phenomenon that makes people want to be allocated in the same location that everyone else, are high;
- the rate at which resources can be reallocated between locations is high,

then an overlap is likely to occur, and self-fulfilling expectations may play a decisive role. Yet, the starting position of the two regions must not be too unequal if self-fulfilling prophecy shall be able to outweigh history.

The subject of history versus expectations has been further discussed by Krugman in his paper "First nature, second nature, and metropolitan location" (Krugman (1993a)), where the role of historical accident and self-fulfilling expectations with regard to metropolitan location is explained. Another approach to the issue of expectations has been made by Baldwin and Lyons (1991), who try to formalise the assertion that highly optimistic expectations about European economic and monetary integration may prove to be self-fulfilling.

A fairly common attitude to the decisive role of expectations is that "smaller" events may be determined by expectations, but as for phenomena of a "larger scale", as for instance Silicon Valley, history plays the most important role.

2.2 Sources of industry clustering

In the previous sections we elucidated the forces determining the establishment and sustainability of a cluster. The models included centripetal as well as centrifugal forces. Significant determinants of agglomeration were internal increasing returns to scale, transport costs, labour mobility, and the share of expenditure spent on "footloose" production (in our model "manufacturing"). Models like this mainly emphasise the importance of *a "market potential" for final goods*: the optimality of a production location depends on its access to markets. But there are also other kinds of

externalities leading to spatial concentration. In this section we try to provide an overview of these external economies.

The most cited concept of external economies was presented by Alfred Marshall in his book "Principles of Economies" in 1920. Despite its age, it still gives a very good review of some of the most important externalities leading firms to concentrate. Marshall mentioned three specific reasons for localisation:

1) *Labour market pooling,*
2) *Intermediate inputs,* and
3) *Technological spillovers.*

Labour market pooling

It is widely believed, that a spatial concentration of firms in an industry (a so-called industrial cluster) which allows for a pooled market for labour with specialised skills, will be beneficial to both firms and workers. A concentrated labour market is, in other words, believed to be a significant reason for localisation. To explain why, it is convenient to use an example.

Suppose that there are two possible locations and two firms. Both firms use the same kind of skilled labour. The firms can produce in either region, but because of increasing returns to scale in production, they will choose to locate in only one of the regions. For some reason, the firms' labour demand is uncertain and imperfectly correlated: times may be "good" or "bad". In "good" times a firm needs 125 workers,

but in "bad" times it only demands 75 workers. The total supply of workers is 200, so that on average demand equals supply.[37]

Why would both firms and workers benefit from a concentration of both firms in one region?

- If the firms are located in two different regions, we assume that this leads to a fifty-fifty split of the labour force between the two locations. As a consequence, a firm will not be able to exploit good times when it needs 125 workers to satisfy the demand. However, if both firms and all workers were located in one region, then, at least occasionally, one firm's good times would coincide with the other firm's bad times, so that both firms' demands for labour would be satisfied.
- Turning to the labour force, the effect of pooling in one location is similar to the effect that is experienced by firms: if each region is dominated by one company, then bad times for the company mean bad times for the workers too. But given that both firms have their plants in one region, then there will be a chance that one firm's bad times will be offset by the good times of the other. Thus, a clustered industry implies a lower average rate of unemployment.

We may conclude that both firms and workers gain from a pooled labour market, and that a pooling of input factors will therefore encourage concentration. But one should be aware that this conclusion relies heavily on the assumption of increasing returns and some kind of uncertainty.

---

[37] The wage rate is, in other words, taken as given, and we assume that it is set at an expected market-clearing level. But it can be shown, that a flexible wage rate would not change the importance of a pooled labour market for the establishing of an industry concentration (See Krugman (1991a), pp 44-45.): an equal division of the labour force and firms between two locations, leads to a higher wage in good times, and a lower wage in bad times. A pooled labour market will imply a constant wage rate equal to the average of the high and low wage rates. Each firm will be employing more people in good times, and fewer in bad times than they would have done if they were isolated from the other firm. Each firm's gains from a pooled labour market will therefore be larger than the losses.

The process that leads to a concentration of firms and labour may be illustrated graphically:

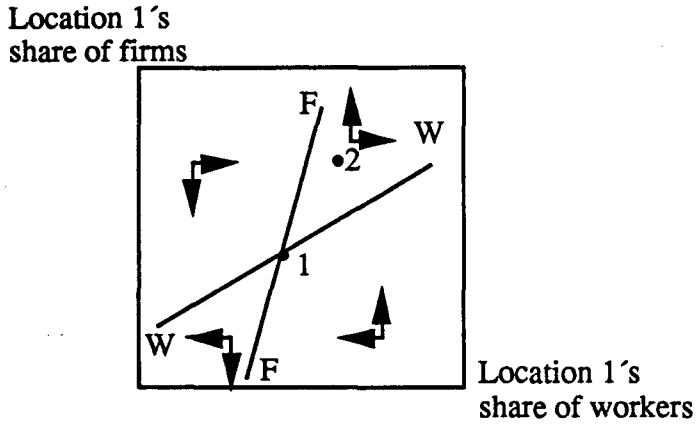


Figure 2.9

There is a fixed number of firms possessing a technology with increasing returns to scale, and a certain number of workers with specialised skills. The FF curve and the WW curve show the distributions of firms and workers that will leave the typical firm and worker respectively, indifferent between the two locations.

Point 1 depicts where the two curves cross each other, i.e., the point in the centre of the diagram where both firms and workers are split evenly between the two locations. Because of the dynamics in the model, which will be explained soon, there are three possible equilibria: all firms and workers in location 1, all firms and workers in location 2, or a fifty-fifty split of firms and workers between the locations.

Due to increased competition for the available workers, the profit of a firm located in one specific region will decrease if the share of firms located in this region increases. The attractiveness of a location to a firm is reduced, if an increase in the number of firms is not offset by an increase in the number of workers in the region. As a result FF must be upward sloping.

For the workers the effect is the opposite. The workers in one location would prefer to locate together with as many firms as possible, and as few other workers as possible, in order to decrease the risk of becoming unemployed.[38] WW also has to be upward sloping.

To see why FF must be steeper than WW, consider point 2 in the figure: at this point the ratio of firms to workers is the same in both locations, but location 1 possesses more of each, thereby offering a more attractive location for both firms and workers than location 2. The "graphical" consequences are, that point 2 has to lie below FF (implicating that location 1 is preferred by the firms); and above WW (implicating that location 1 is also preferred by the workers). This implicates an FF curve steeper than a WW curve at point 1.

Reviewing the dynamics of the model, we can see that the equilibrium in the centre of the diagram is unstable. Firms and workers will choose to concentrate in only one of the locations, i.e., we shall see a convergence towards location 1 or 2.

[38] Replacing the assumption of a fixed wage rate with an assumption of flexible wages, would still leave the workers with a motivation for allocating close to a cluster of firms: the more firms in one location, the higher the wages ceteris paribus.

The results above are robust to other assumptions about the wage rate (see footnotes 41 and 42), as well as to the possibility that a firm would try to exploit its monopsony power. Such a firm-behaviour would, in fact, only increase the effect that a pooled factor market has on localisation.[39]

Intermediate inputs and vertically linked industries

The availability of specialised inputs and services is another externality, assumed to cause agglomeration of firms belonging to the same industry (and therefore demanding the same kind of intermediate inputs). A geographically concentrated industry allows for a greater variety of, and more specialised, intermediate inputs, at a lower cost, at the same time as a substantial concentration of downstream firms provides upstream firms with a beneficial market access. In other words, cost, variety, and demand linkages all encourage agglomeration of vertically linked industries. That vertical links trigger agglomeration, does, however, depend critically on an assumption of increasing returns to scale in production and imperfect competition.

Comparing the kinds of agglomeration economies elucidated in the "Core-periphery" model with the one just described, we observe that in both cases the phenomenon *"market potential"* is central. They differ in the way the market-size is made endogenous. In the "Core-periphery" model, the size of the market at different locations was endogenous due to labour mobility. But if industries are vertically linked, it is the movement of the downstream industry that affects the market-size for the upstream firms.

---

[39] See Krugman (1991a), pp 46-49.

Anthony Venables has developed a formal model considering the allocation problem of firms in an upstream and a downstream industry that are vertically linked (see Venables (1993)). Since this model is basically quite similar to Krugman's "Core-periphery" model, we will not elaborate it in detail here, but just provide a brief description of framework and results.

Both industries are imperfectly competitive, and the interaction between the downstream and the upstream industry creates a force for agglomeration. Centrifugal forces are labour supply and final demand, which are spread across locations, and assumed immobile. Although the specific forces working for and against concentration are different from those in the "Core-periphery" model, the economies of agglomeration (positive externalities) are derived from the same, single source in both models, namely the market interactions. Technological externalities are neither considered by Krugman, nor by Venables.

A key parameter is the costs of market access.[40] Similarly to the results regarding the sustainability of a core in the "Core-periphery" model, it turns out that the forces of agglomeration are greatest at the intermediate level of these costs. A reduction in transport costs from a high to an intermediate level, leads to agglomeration and divergence among regions. But a further reduction in transport costs to a very low level will, according to Venables' results, cause diversification and increased convergence of economic structure and income among regions.

It is interesting to note that, a reduction in costs affecting the incentives to cluster, and thereby the spatial equilibrium, may lead to reallocation in different directions,

[40] Costs of market access are equal to what we have earlier referred to as "transport costs". For the sake of simplicity, we will keep referring to these kinds of costs, as "transport costs".

depending on the characteristics of the industry in question and other vertically related industries. Furthermore, depending on initial conditions and industry characteristics, the effect of parameter changes may be small or large, incurring a slightly different equilibrium or a dramatic change from concentration to dispersion (or vice versa).

Venables also discusses the implications of a regional "industrial base": because a firm's choice of location depends on the allocation of other firms, like in the "Core-periphery" model, the possibility of multiple equilibria arises. Another consequence of this interdependence between firms is that, if one firm's (or several firms') locational choice is altered, the whole vertically linked chain of firms may be affected, and as a result we get a totally changed equilibrium.

Ultimately, we return to the crucial assumption of market imperfections. This assumption is critical to the results achieved by Venables. If, for instance, the upstream industry were perfectly competitive with all manufacturers producing the same homogenous goods, the downstream industry would probably always use its local supplier. But product differentiation (as Venables assumes in his model) ensures that all downstream firms use all upstream firms' products. Due to demand linkages, the firms in the upstream industry would, in this case, have an incentive to concentrate their production in one location, and the price for intermediate goods will be lower the more upstream firms there are in one location. This implies lower costs for the downstream industry and encourages agglomeration.

Technological spillovers

This is the very "common" externality that is (almost) always mentioned among the reasons for agglomeration in space. Silicon Valley and Boston's Route 128 are

famous examples of industry clusters arising from knowledge spillovers between nearby firms. The difference between these kinds of externalities and the external economies rising from labour market pooling and intermediate goods supply, is that knowledge spillovers are, as Krugman describes them, invisible. In other words, technological spillovers are, difficult to measure and an economist needs to make several assumptions about their form and size before he can include them in a model.

Comparing the modelling of spillover effects with the models in which market interactions are responsible for agglomeration, the advantage of the latter models is clear: explicit assumptions about the externalities are not necessary. The external economies rise from interaction between increasing returns, factor mobility, and transport costs.[41] Currently, the modelling of knowledge spillovers relies totally on the vague concept of external effects. As a consequence of the difficulties related to the models with these kinds of agglomeration economies, the issue of technological spillover effects has not received much attention within the new field of geography and economics.

[41] As, for instance, from the interaction between internal increasing returns, factor mobility, and transport costs in the Core-periphery model.

3. International trade and geography

3.1 The difference between a region and a nation

What is the difference between a region and a nation? The answer to this question may seem obvious. But a proper definition of these two notions is necessary in order to understand the relation between the subjects microeconomics and geography, and international trade and geography.

A nation is not necessarily equal to a region. A nation may consist of several regions, but it is also possible that a number of nations constitutes a region. The region depicts the relevant territory where external economies leading to agglomeration and development of core-periphery patterns, apply. Within the theory of imperfect competition and international trade, it has been widely believed that, external economies apply at the level of a nation, implicating that political borders have the same function as regional boundaries.[42] This presumption is, in fact, not quite correct. We may find examples proving that, indeed, it is the borders that determine the unit, within which the external effects work. But then, this is because of national policies, not because of economic factors. In other words, it is the regional boundaries, not the national ones, that are decisive for where the external effects apply.

But in spite of all we have said about a nation being unequal to a region, due to the great importance of national governments' policies influencing the movements of factors and goods between countries, in many cases it will be both possible and

[42] Based on such an assumption, Krugman (1980), for instance, finds that allowing for increasing returns, transport costs, and differences in country size, a large country will tend to export goods characterised by increasing returns in production, while a small country will tend to export goods produced subject to constant returns to scale.

correct to compare a nation with one region. What we wish to emphasise is that, although there are both cases where a country is equal to just one region and cases where it contains several regions, the framework and model structure used in chapter two are also applicable within the theory of international trade and geography. Actually, it is not possible to separate the subjects microeconomics and geography, and trade and geography. Models can include two or more regions, allowing for the study of one or multiple agglomerations (cores), independent of which of the two subjects we are working with.

Some will probably argue that the models described in the previous chapter did not include tariffs or any other kinds of trade barriers. However, "transport costs", which were included, need not be interpreted as costs related to transport only. Reviewing Venables' model of vertically integrated firms (see paragraph 2.2), we mentioned that "costs of market access" should be considered equal to what we had earlier referred to as "transport costs". There is no reason why we cannot think of these costs as a mixture of transport costs, tariffs, difficulties related to communication, cultural differences, and other kinds of trade costs: what they all have in common is that they tend to reduce geographical concentration. The interpretation of the notion "transport costs", is determined by whether we are studying trade between national regions or between countries.

3.2 The implications of the "new" location theory for trade theory

The development of what we have called a "new" location theory allows us to study new, interesting fields within the theory of international economics. Additionally, it may force modifications and alterations of earlier results attained in international trade.

Where production takes place is, without doubt, of great importance to the individual nation. With respect to the welfare of its inhabitants, it is not trivial whether a large share of the world's production is located within the country or in a neighbouring country. Hence, it is in the own interest of a national government to conduct an economic policy which reinforces localisation of industries within its national borders.

Typically appealing international trade issues that may be analysed employing the tools of the "new" location are, for instance:

- The way changes in tariffs and other trade policies influence localisation and national welfare. How, for example, the results from the GATT negotiations might influence a country's production and welfare.
- Economic integration, and the implications of for instance the European Union (EU).
- How changes in transport costs influence localisation and trade. Transport costs are obviously also affected by the technological evolution, leading to for example cheaper transportation. But the costs related to transportation may also be influenced by a firm's or an industry's "production philosophy": currently "just in time" is a well known notion, implicating the importance of time for the producers. The necessity of precise (and fast) delivery, may implicate increased transport costs.

- The technological evolution, affecting the importance of economies of scale in production, and thereby a firm's allocation decision.
- Endogenous growth within the context of economic integration.
- A country's regional policy. We learned in the previous chapter that immobile factors of production, immobile suppliers, and immobile demand have a centrifugal effect, i.e., a negative effect, on agglomeration. A government's influence on the settling of households and firms, may have a significant effect on localisation and on the conditions of production and trade.
- How the price of land influences agglomeration and a firm's incentives to locate within a specific region or nation. Price of land is usually presumed to have a negative effect on agglomeration, because localisation and urbanisation are developments that enforce an increase in the price of land.
- How the availability and quality of public goods can have either negative or positive effects on agglomeration (external diseconomies or external economies of scale). A country with a good infrastructure will provide its firms with better conditions for concentration, and better possibilities of exploiting internal and external economies of scale.

To provide a first insight into how issues like these may be analysed, we offer an example: considering the first two issues, it is obvious that changes in trade costs are in focus. These changes may be analysed in the context of the "Core-periphery" model described in the previous chapter. We only make one modification: in order to be able to study the case of multiple agglomerations or cores, we allow for more than two regions. We assume that there are three possible geographical structures: depending on the parameter values (increasing returns, transport costs etc.), there may be one core, two cores or no cores at all.

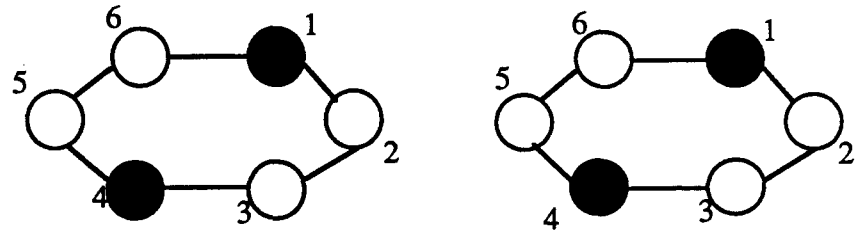


Figure 3.1

Figures 3.1 illustrates the possibilities of one and two cores respectively. (The cores are marked in black.) Our economy is assumed to contain 6 regions. Transport between the regions is only possible along the lines connecting the regions. In the case with two cores, each core will have a hinterland consisting of two neighbouring regions. (Hence, hinterland of region 1 is regions 2 and 6, and regions 3 and 5 form the hinterland of region 4.)

Given strong economies of scale in production, low transport costs, and a large share of the economy's industries being "footloose", the economy will probably support one core only. If the parameters take on the reverse values, then there will probably be no cores at all. As for intermediate levels of all the parameters, there will be a chance of multiple (in this case: two) cores.

We now divide our economy into two countries: one country consisting of two regions, and one country consisting of 4 regions. The broader line in figure 3.2 illustrates how the economy is divided. Between the countries there are trade barriers causing higher transport (trading) costs internationally than nationally.

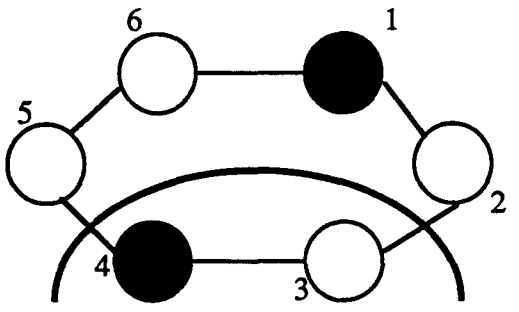


Figure 3.2

What will happen if the transport costs between the countries are reduced, or the two economies become fully integrated? There are more than one possible outcome.

Depending on how large the reduction in transport costs is, there will be an equilibrium with one or two cores. Imagining that we end up with one core, this will probably be located in the larger country (with four regions), because of this nation's head start caused by its larger home market. The larger nation will gain, and the smaller will lose, due to the integration.

But if the equilibrium is characterised by two cores, this may be beneficial for the small country, but not for the large country. The reason why being that the core in the small country will presumably be able to gain extra hinterland, namely region 5, which was earlier hinterland of the core in the large country, and is actually more closely located to the core in the small country. Manufacturers in the small country will get the opportunity to expand at the expense of the manufacturers in the larger country.

As for welfare effects of economic integration, we first need to define whose welfare we are concerned with. If the welfare of a nation is defined as the welfare enjoyed by

its immobile factors, the relationship between transport costs, localisation, and welfare may be illustrated as in figure 3.3. Suppose that there are two regions, and transport costs incurred by trade between the regions are high. The firms will not have an incentive to cluster in just one of the regions. The core-periphery pattern will not appear. The degree of welfare is the same in both regions.

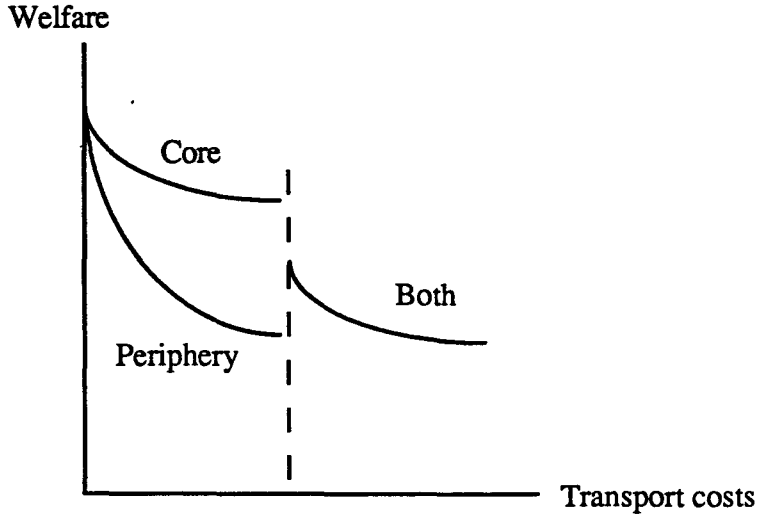


Figure 3.3

A gradual reduction in transport costs raises welfare in both regions, until a critical level of transport cost is reached. At this level the transport costs have declined to such an extent that the geographical structure changes dramatically. A core emerges, and we can observe great differences in welfare between the core-region and the periphery-region. But as the transport costs decline even more, both regions will experience a growing, converging welfare. As transport costs reach zero, location ceases to matter, and the degree of welfare will be the same in both regions.

In other words, a reduction in transport costs to an intermediate level leads to a concentration in one region, i.e., the country possessing the core gains, whereas the more peripheral country loses. If both nations are to gain from an economic integration, we need a close integration with dramatically reduced transport costs, not only a limited one. Figure 3.4 shows how welfare in the nation condemned to become the periphery, depends on the degree of economic integration.

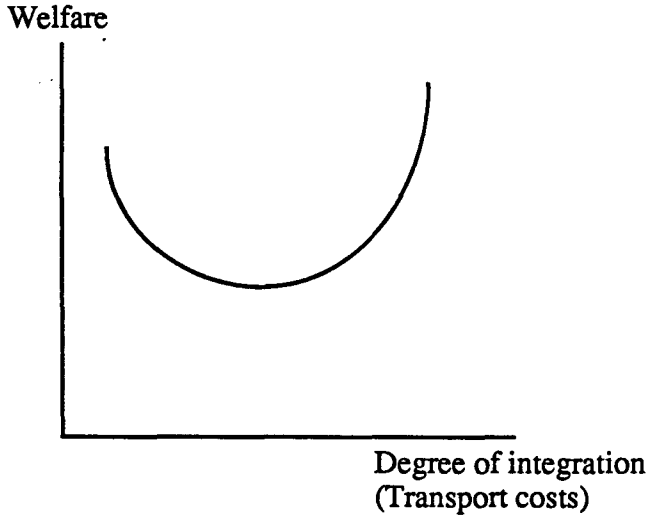


Figure 3.4

A few words about the implications for trade policy: being close to the critical level of transport costs (see figure 3.4), only a small change in these costs will be enough to achieve a significant raise in welfare in one of the nations. Hence, even modest policy alterations may, under certain circumstances, be enough to incur a dramatic change in one country's welfare, at the expense of the welfare of another country.

3.3 Applications of economic geography within trade theory

Most of the authors who have been preoccupied with geography, international economics and trade, have chosen to concentrate on the European integration. The economic integration within Europe has, in fact, probably been the prime motivation for several economists to devote themselves to the subject of geography. The completing of the internal market raises many important questions from an economist's point of view. Some of these cannot be elaborated or answered without the inclusion of geography and space in traditional economic models.

The article "'Integration and the competitiveness of peripheral industry", written by Krugman and Venables (1990) has been pathbreaking on the topic of trade and geography. They approach a problem that has received much attention from economists as well as politicians: how the completing of the internal market, involving a greater economic integration of the EU countries, will affect the role played by the Southern peripheral regions of the community.

It is quite common to consider two possible outcomes of the integration process:

- The economic integration may permit the Southern countries in Europe to exploit their comparative advantage, and lead to an expansion of the relatively labour-intensive industries in these economies. In other words, due to labour abundance and lower relative wages, we might observe a process of reallocating of industries from the North towards the Southern periphery.

In the context of trade theory, we explain this result as follows: in a situation with restricted trade, a small country will not be able to exploit the possibility of economies

of scale in production at the same time as a high level of competition is achieved. This "cost of smallness" is removed by free trade. The outcome is somehow consistent with what we would have expected if the Heckscher-Ohlin theory had been applied to the problem: the nations being relatively labour abundant will be net-exporters of labour-intensive products.

- The fact that the Northern countries in the EU have a more economically central location within the community, leads to a closer and better market access for the firms situated in these countries. If the importance of "market potential" is evaluated higher than cheap labour force by the firms, the Southern countries might experience a relocation of its industries towards the North, as a consequence of the economic integration.

In their article Krugman and Venables elaborate how the role of the Southern periphery is determined, i.e., under what circumstances the different outcomes seem likely. Their model consists of two economies, one central with a large local market, and one small (labour abundant) economy. Each economy has an imperfectly competitive (and labour-intensive) manufacturing sector, with firms possessing technologies with increasing returns to scale. Intra-industry trade will occur, since firms in both countries supply consumers in both countries. Trade between countries incurs transport costs.

The authors conduct the following experiment: they reduce the barriers to trade between the countries, look at how the relative competitiveness of the manufacturing sector in each country is affected, and elucidate how production and trade may change. Consistent with what we have written earlier, the following results appear:

- If trade barriers are very high, markets in both economies will be served by local firms, i.e. self-sufficiency determines the location of production. Due to the production technologies with high fixed costs, the price must be relatively higher in the small economy.
- As trade barriers are reduced, production tends to move towards the centre, and the periphery becomes a net importer of manufactured goods. At intermediate levels of barriers, the advantages of greater access to the larger market are most important, i.e., the centripetal forces are the strongest.
- If trade barriers are very low, and there are no cost differences across the economies, location ceases to matter: firms have equally good access to all markets, independent of location.

The importance of market access is countervailed by factor price differences between the centre and the periphery. Assuming that the small country is relatively labour abundant with relatively low wages, and that the manufacturing sectors are labour intensive, this pulls production towards the peripheral country.

Integration will lead to an increased divergence in wages until an intermediate level of transport costs is reached. At this level, where the advantages of a good market access are most important, the wages in the periphery will also be the lowest, due to their initial level, and due to the country's decline in production during the first part of the integration process. As the process of integration continues, the wages in the two countries will converge. It appears that it is possible to draw a U-shaped curve to explain how the wages in the periphery change as the process of economic integration proceeds.

In general, the relationship between output (welfare) in the peripheral country and integration can also be described through a U-shaped curve: as integration approaches an intermediate level, production in the periphery declines. But as we reach a very low level of transport costs, relative factor endowment determines the direction of net trade, and the relatively labour abundant small economy becomes net exporter of manufactures.

The authors emphasise, that forces tending to equalise factor prices, such as international factor mobility, will reinforce the importance of market access in determining the location of manufacturing.

Krugman and Venables conclude that the effects of the completing of the internal market are ambiguous. Because of lower trade costs, there could be an expansion of the production in the Southern peripheral countries. But we might just as well observe a concentration of production in the North, leading to a reduction in peripheral wages. Anything impeding the relative wages in the South from changing, reinforces the tendency to concentration in the Northern centre. But since we do not know the initial levels of trade barriers within the EU, i.e., on which side of the U-curves we are, it is not possible to tell whether the completing of the economic integration will improve or worsen the competitiveness of the peripheral industry.

The two authors have later followed up the work just referred to, with Krugman and Venables (1993c), where they demonstrate the possibility that due to the growing integration of the European market, Europe may develop an American-style economic geography. "American-style" in the sense that each industry only has one major centre of production. Their main concern is that although such a development is found to

raise real incomes in the end, there may be serious adjustment problems along the way. Possible transition costs associated with such a shift in the economic geography are elaborated.

Other authors have also expressed their concern about the possibility that the European integration may lead to a divergent economic development within the Community:

Blanchard and Katz stress the problems related to a monetary union in Europe.[43] The two authors have studied how the states of the U.S. respectively, adjust after having been affected by an adverse shock to employment. They find that the dominant adjustment mechanism is labour mobility. Based on their results from the U.S., Blanchard and Katz criticise one of the common arguments for a common currency area in Europe: once a common currency is introduced and exchange rates are fixed, "firms and workers will no longer expect to be bailed out by monetary expansion and depreciation". But what will now happen if the firms face a decline in competitiveness? Due to a missing flexibility, wages will not decline enough to prevent an increase in unemployment. According to evidence from the U.S., the regions affected by a negative shock, will experience a labour emigration as the economy adjusts towards equilibrium.

What Blanchard and Katz suggest, is that due to a lower labour mobility in Europe than in the U.S., shocks may have larger and longer lasting effects on relative unemployment in Europe.

[43] See Blanchard and Katz (1992).

Probably motivated by Baldwin (1989), who argues that economic integration may be expected to increase the level of European production as well as the European growth rate, Bertola emphasises that not all European regions will necessarily experience an increased economic growth as an impact of the integration. To provide a better understanding of his argument, he offers a model of localised endogenous growth.[44]

Bertola's model is based on the assumptions of two regions; increasing returns in production, constant returns to capital in each region, which enjoys potentially different levels and growth rates of production before integration, and free factor mobility.

The exploitation of scale economies made possible by the economic integration, can only be realised if the factors of production are concentrated in the region with the highest productivity. It is stressed that as long as geographic concentration leads to higher aggregate production, policy should not prevent factor mobility. Instead, national governments and the EU politicians should concentrate on a distributional policy in favour of the owners of the immobile factors (such as land) in the underprivileged regions. Bertola actually states that a divergent economic development does not necessarily need to be inconsistent with better efficiency and faster aggregate growth. So far, he seems to agree with Baldwin.

But it is then argued, that due to empirical evidence about Europe, it would perhaps be more correct to assume that returns to capital accumulation are constant in the aggregate of two distinct localities, but decreasing if each locality is considered

[44] See Bertola (1992).

Norges Handelshøyskole
Biblioteket

separately.[45] Based on this altered model where the role of externalities is emphasised, Bertola finds that the market interactions will yield a privately optimal, but probably not a socially optimal outcome. When a geographical dimension is added to the externalities in a model of endogenous growth, free factor mobility may not optimise the overall performance of an integrated economy.

To influence private investment decision, and prevent undesirable side-effects of economic integration, which lead to a slower rather than a faster growth in the integrated economy, distortionary taxation and lump-sum redistribution are needed. Bertola concludes that a united European tax structure ensuring the internalisation of geographical linkages between productive activity in different localities, might be needed to promote aggregate European economic developments.

Martin and Rogers (1994) examine the impact of public infrastructure on industrial location, when increasing returns are present. They analyse the role of infrastructure in the context of trade integration, and find that, not surprisingly, trade integration implies that the sector with increasing returns to scale will locate in the countries with the best infrastructures. As for policy implications, it turns out that if the aim of an infrastructure policy is to foster industrial convergence between a rich and a poor country, the infrastructure policy must be biased in favour of infrastructure that facilitates intra-regional trade rather than international trade.

Ultimately, we would like to mention a few papers on geography and international trade that are concerned with problems associated with the Third World, rather than

[45] Bertola argues that the capital inflow experienced by the capital-poor countries Spain and Portugal since they joined the EU, does not support the thought of constant returns to capital in all European countries. If this were the fact, we should expect an inflow to the richer members of the EU.

with the European integration. Krugman and Venables (1994), Puga (1994), as well as Elizondo and Krugman (1992) have all been concerned with such questions. Krugman and Venables consider how globalisation affects the location of manufacturing and the gains from trade enjoyed by the Western and Third World countries. In the two latter articles, the impact of trade policy on the regional structure of a country is discussed. We shall give a short summary of the article by Elizondo and Krugman:

Today, many of the largest cities of the world are located in developing countries. But in the development literature little, if any, attention has been paid to the issue of urbanisation and the growth of giant cities. When trade policies in developing countries are discussed, the effects such policies could have on internal economic geography in these nations, are neglected.

Elizondo and Krugman argue that this neglect is serious, if not a mistake, because the trade policies of developing countries and their tendency to develop large metropolitan centres are closely linked: the rise (and gradual fall) of giant cities in developing countries can be explained by the strong backward and forward linkages that exist when the manufacturers try to serve a small, closed domestic market.[46] It is argued that closed markets promote a centralisation of production and huge central metropolises, whereas open markets discourage agglomeration and the emergence of such cities.

[46] Backward linkages depict access to consumer markets; and forward linkages depict access to other firms' goods. The goods of other firms can either belong in the consumption basket of the firm's workers, or they can be intermediates.

The giant metropolises in the Third World may be seen as an unintended by-product of import-substitution policies, because protection is assumed to imply stronger backward and forward linkages enforcing concentration. As developing countries liberalise, one might expect that the large cities in developing countries would tend to shrink due to weaker (national) linkages. Employing a model of economic geography, Elizondo and Krugman actually show that starting with a concentrated population, a process of liberalisation affecting the degree of reliance on the domestic market will reduce concentration. Increased export of final products and increased import of intermediates decrease the importance of being close to a country's metropolitan centre.

***

The subject of trade and geography is a new one with several unanswered questions. Much research is still to be done: existing models may be extended and improved by the inclusion of additional parameters and variables, and new models should be developed in the attempt to study further aspects of economic geography and localisation within a microeconomic framework.

References

Abdel-Rahman, Hesham and Masahisa Fujita (1990): "Product variety, marshallian externalities and city sizes", *Journal of Regional Science* 30 (1990): 165-183

Alonso, W. (1964): *Location and land use. Towards a general theory of landrent.*, Cambridge, Massachusetts, 1964

Baldwin, Richard E. (1989): "The growth effects of 1992", *Economic Policy* 9 (1989): 248-281

Baldwin, Richard E. and Richard Lyons (1991): "External economies and Integration: the potential for self-fulfilling expectations", CEPR Discussion paper No. 471, London, 1991

Bertola, Giuseppe (1992): "Models of economic integration and localized growth", CEPR Discussion Paper No. 651, London, 1992

Blanchard, Olivier J. and Lawrence F. Katz (1992): "Regional evolutions", *Brookings paper on economic activity*, Vol. 1, Washington D. C., 1992

Böventer, E. von (1962): *Theorie des räumlichen Gleichgewichts*, Tübingen, 1962

Brander, James A. and Paul R. Krugman (1983): "A "reciprocal dumping" model of international trade", *Journal of International Economics* 15 (1983): 313-323

Christaller, Walter (1933): *Die zentralen Orte in Süddeutschland. Eine ökonomische-geographische Untersuchung über die Gesetzmässig-keit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen*, 1. Auflage, Jena, 1933

Dicken, Peter and Peter E. Lloyd (1990): *Location in Space. Theoretical perspectives in economic geography.*, 3rd ed., New York, 1990

Dixit, Avinash K. and Joseph E. Stiglitz (1977): "Monopolistic competition and optimum product diversity", *American Economic Review* 67 (1977): 297-308

Elizondo, Raul Livas and Paul R. Krugman (1992): "Trade policy and the third world metropolis", NBER Working Paper No. 4238, Cambridge, Massachusetts, 1992

Enright, Michael J. (1991): *Geographic concentration and industrial organization*, unpublished PhD dissertation, Harvard University, 1991

Fujita, Masahisa (1988): "A monopolistic competition model of spatial agglomeration", *Regional Science and Urban Economics* 18 (1988): 87-124

Fukao, Kyoji and Roland Benabou (1993): "History versus expectations: a comment", *Quarterly Journal of Economics* 108 (1993): 535-542

Helpman, Elhanan and Paul R. Krugman (1985): *Market structure and foreign trade*, Cambridge, Massachusetts, 1985

Henderson, J. Vernon (1974): "The sizes and types of cities", *American Economic Review* 64 (1974): 640-656

Henderson, J. Vernon (1988): *Urban development: theory, fact, and illusion*, New York, Oxford, 1988

Hoover, Edgar M. (1937): *Location theory and the shoe and leather industries*, Cambridge, Massachusetts, 1937

Ihde, Gösta B. (1984): *Transport, Verkehr, Logistik: Gesamtwirtschaftliche Aspekte und einzelwirtschaftliche Handhabung*, München, 1984

Ihde, Gösta B.: "Standortdynamik als strategisches Antwort auf wirtschaftliche Strukturveraenderungen", in Gaugler, Eduard, O. H. Jacobs and Alfred Kieser (eds.): *Strategische Unternehmehmensfuehrung und Rechnungslegung*

Isard, Walter (1956): *Location and Space-economy*, Cambridge, Massachusetts, 1956

Krugman, Paul R. (1980): "Scale economies, product differentiation, and the pattern of trade", *American Economic Review* 70 (1980): 950-959

Krugman, Paul R. (1984): "Import protection as export promotion: International competition in the presence of oligopoly and economies of scale", in Kierzkowski, H. (ed.): *Monopolistic competition and international trade*, Oxford, 1984

Krugman, Paul R. and Anthony J. Venables (1990): "Integration and the competitiveness of peripheral industry", in Bliss, Christopher J. and Jorge Braga de Macedo (eds.): *Unity with diversity in the European economy: the Community's southern frontier*, Cambridge: Cambridge University Press, 1990

Krugman, Paul R. (1991a): *Geography and trade*, Cambridge, Massachusetts: MIT Press, 1991

Krugman, Paul R. (1991b): "Increasing returns and economic geography", *Journal of Political Economy* 94 (1991): 483-499

Krugman, Paul R. (1991c): "History versus expectations", *Quarterly Journal of Economics* 106 (1991): 651-667

Krugman, Paul R. (1991d): "History and industry location: the case of the manufacturing belt", *American Economic Review* 81 (1991): 80-83

Krugman, Paul R. (1991e): "Cities in space: three simple models", NBER Working Paper No. 3607, Cambridge, Massachusetts, 1991

Krugman, Paul R. (1992): "A dynamic spatial model", NBER Working Paper No. 4219, Cambridge, Massachusetts, 1991

Krugman, Paul R. (1993a): "First nature, second nature, and metropolitan location", *Journal of Regional Science* 33 (1993): 129-144

Krugman, Paul R. (1993b): "On the number and locations of Cities", *European Economic Review* 37 (1993): 293-298

Krugman, Paul R. and Anthony J. Venables (1993c): "Integration, specialization, and adjustment", NBER Working Paper No. 4559, Cambridge, Massachusetts, 1993

Krugman, Paul R. and Anthony J. Venables (1994): "Globalization and the inequality of nations", CEPR Discussion Paper No. 1015, London, 1994

Launhardt, W. (1885): *Mathematische Begründung der Volkswirtschaftslehre,* Leipzig 1885 (Neudruck der Ausgabe, Aalen, 1963)

Lösch, August (1962): *Die räumliche Ordnung der Wirtschaft*, 3. Auflage, Stutgart 1962 (orginally published in Jena in 1940)

Marshall, Alfred (1920): *Principles of Economics*, London: Macmillan, 1920

Martin, Philippe J. and Carol Ann Rogers (1994): "Industrial location and public infrastructure", CEPR Discussion Paper No. 909, London, 1994

Miksch, L. (1951): "Zur Theorie des räumlichen Gleichgewichts", *Weltwirtschaftliches Archiv*, 1951

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1989): "Increasing returns, durables, and economic fluctuations", NBER Working Paper No. 3014, Cambridge, Massachusetts, 1989

Porter, Michael E. (1990): *The competitive advantage of nations*, London and Basingstoke: Free Press, 1990

Puga, Diego (1994): "The dynamics of urbanisation", London School of Economics, Centre for Economic Performance Working Paper No. 652, London, 1994

Puga, Diego (1995): "Preferential trade arrangements and industrial location", London School of Economics, Centre for Economic Performance Working Paper No. 724, London, 1995

Reve, Torger, Terje Lensberg, and Kjell Grønhaug (1992): *Et konkurransedyktig Norge*, 1992

Rivera-Batiz, Francisco L. (1988): "Increasing returns, monopolistic competition, and agglomeration economies in consumption and production", *Regional Science and Urban Economics* 18 (1988): 125-153

Romer, Paul M. (1986): "Increasing returns and long-run growth", in *Journal of Political Economy* 94 (1986): 1002-1037

Thünen, Johann H. von (1826): *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg 1826

Timmermann, Manfred (1983): "Standort-Diversifikation als Instrument der Unternehmenspolitik", *Die Unternehmung* 28 (1983): 41-49

Venables, Anthony J. (1993): "Equilibrium locations of vertically linked industries", CEPR Discussion Paper No. 802, London, 1993

Weber, Alfred (1922): *Ueber den Standorten der Industrien*, Tübingen 1922

Essay 2

Uncertain trade costs and industrial localisation — Concentration versus diversification

by

Karen Helene Midelfart Knarvik*

Introduction

Experience has showed us that uncertainty about the political and economic environment and a fear of future policy reversal, may cause reallocation of firms and capital flight, and have serious implications. To stimulate business and growth in a country, commitment and irreversible policy changes are required.

This paper was inspired by the choice of a few small European countries to remain outside the European Union (EU), in a time when the world is dividing itself into rival trading blocks. The future of countries remaining outside these blocks may seem quite uncertain. Three of these small European "outsider" countries chose to sign the European Economic Area (EEA) agreement, which extends the EU's "four freedoms" to themselves. But doubt has been expressed as to what extent this agreement provides

* I would like to thank Jan I. Haaland, Victor D. Norman, Jostein Tvedt and Anthony J. Venables for valuable ideas and comments.

assurance that policy changes are stable and irreversible to the same degree as EU membership would have done.[1]

There are two "types" of uncertainty related to the EEA: (i) its somewhat uncertain future and the possibility of parliaments choosing to terminate the agreement. (ii) The always existing possibility of parts of the agreement being suspended, due to dissension about new EU measures.[2] The impact of such suspensions may be severe for commercial interests. It is often suggested that the EEA agreement will therefore probably *not* provide firms with sufficient assurance that policy changes are permanent, in contrast to an EU-membership, which is widely believed to eliminate any artificial reason for firms to relocate production from EFTA to EU.

What we seek to elucidate, is how a situation of uncertainty regarding future trade policy, international agreements, and thereby trade costs[3], may affect firms' competitiveness, investment behaviour and allocational choice. Of main interest are the implications uncertainty might have for the allocation of economic activity, for the distribution of such between countries, and for international trade.

To be concerned about the impact of uncertainty does indeed not represent anything new within the subject of international economics. There is a literature — though not a substantial one — on international trade and uncertainty with important contributions

---

[1] Wijnbergen (1985), a.o. demonstrate, that uncertainty about future policy reversals coupled with irreversible investment, may lead to capital flight.

[2] In accordance with the political system of the EEA, the EFTAns do not actually have to approve every new EU measure, but they never have the choice between status quo and accepting the new law. In other words, the law must be accepted or the whole relevant part of the EEA agreement is suspended. (See Baldwin (1994).)

[3] Trade costs should be interpreted as a measure of an extensive range of costs, related to the export from one country to another, and thereby including trade taxes, transport costs, costs of frontier formalities and differing product standards. A discussion of some of these costs is found in Baldwin (1994), p. 43; and an overview of non-tariff barriers is given in the Cecchini report (1988).

by Helpman, Razin, Grossman and Cheng, a.o.[4] Characterising most approaches to the subject, are the assumptions of competitive firms and a degree of risk aversion. The prevailing types of uncertainty are price uncertainty, technological uncertainty or uncertainty in preferences. The mainstream of contributions is concerned with the question of to what extent the fundamental theorems of the theory of international trade carry over to uncertain environments, and studies the pattern of trade under uncertainty.

More related to the subject of this paper is probably the work done by Baldwin, Dixit and Krugman on how exchange rate fluctuations influence entry and exit decisions, and may yield hysteresis.[5] The implications that trade cost uncertainty has for the decision of a profit maximising, imperfectly competitive, internationally trading firm about how and where to locate, have, however, been subject to less attention. Nor can the traditional theory of location provide us with much insight regarding such a firm's allocation.[6] And although the issue of industrial location coupled with the theory of trade has recently become subject to increasing attention, uncertainty aspects still remain pretty ignored.[7]

The objective of our approach is to focus on how firms' choice of allocation internationally, and thereby foreign direct investments, are influenced by uncertain trade arrangements. We analyse how an oligopolistic firm that is critically dependent

---

[4] See e.g. Helpman and Razin (1978a) and (1978b), Grossman and Razin (1985), Cheng (1983). Based on a survey by these authors, Dumas (1980) offers a good overview of the subject, by categorising the different approaches within the literature.

[5] See e.g. Baldwin and Krugman (1989), and Dixit (1989a) and (1989b).

[6] There have been approaches to the issue of uncertainty and imperfectly competitive firms. But the location of the firm is then confined to a set of points along a line between output and input markets, and one solves for the optimum location on this line. (See Mai (1984), and Park and Marthur (1990).) — In other words, an approach consistent with traditional location theory, but not addressing the aspects of uncertainty, that we wish to emphasise.

[7] See e.g. Krugman (1991), Venables (1993), and Krugman and Venables (1993).

upon access to a large foreign market, may respond to uncertain trade costs. Will it diversify investments and allocate some of its production facilities abroad in order to avoid market foreclosure, and if so, what are the necessary conditions for a firm to react in this way?

This paper is organised as follows: section 1 offers an overview of the basic elements of the model. Section 2 sets out the formal model. In section 3 the conditions for foreign direct investment and geographically dispersed production in the absence of uncertainty, are derived. How to distribute production capacity internationally under trade cost uncertainty is considered in section 4, and the case with uncertainty is compared with that characterised by no uncertainty. Section 5 concludes.

1. Basic elements of the model

We introduce a model of intra-industry trade between two economies with different market size (s). In order to focus on the implications of trade cost uncertainty for the production and welfare of a region or nation, and bring out the importance of relative market size, comparative advantages — differences in technology and factor endowments - are assumed away. Hence, countries only differ in relative size. The two economies will be labelled subscripts i=1,2. The market in economy 1 is assumed smaller than that in economy 2, so that $s_1 < s_2$. The industry in focus is imperfectly competitive with firms producing homogenous products. There are barriers to entry, so that the number of firms in the industry is fixed, and exogenous to the model. One should note that the number of firms in the two-country world does not necessarily equate the number of plants. One firm may have one or more plants. Though it will

never have more than one plant in each country. Thus, we have a model of international oligopoly and possibly foreign direct investment.

There are costs related to international trade that are subject to uncertainty. We imagine that there are two possible future states: one "good" and one "bad". In a "good" state trade costs are low — or possibly infinitely small. But in a "bad" state these costs are so high that they may, in fact, obstruct trade completely, because foreign manufacturers are not able to compete with domestic manufacturers. We let the probability of low trade costs be ρ, $0 < \rho < 1$, so that the probability of high trade cost is $(1-\rho)$.[8] Regarding the firms' attitude towards risk, all firms are assumed risk neutral, and they act in order to maximise expected profits. In order to avoid the uncertain trade costs, a firm can undertake foreign direct investment, i.e. diversify. The transition to becoming a multinational firm enables the firm to conduct foreign manufacturing, and thereby to serve its foreign consumers from a plant in their own country.[9]

Launching a two-step decision model of oligopolistic firms undertaking investment in production facilities, we shall consider the implications of trade cost uncertainty for a firm's distribution of production between its home country and other countries.

---

[8] In accordance with the "common prior assumption" (also known as the Harsanyi "doctrine"), we assume that the probability of different policies, which should be regarded as a subjective probability assessment, is common to all firms, because they are all thought to have access to the same information. However, following Kreps (1990), we argue that this is, in fact, a very strong assumption with substantial implications for the results derived in the proceeding model.

[9] Within the theory of trade, there is a substantial literature on multinational firms. Four different incentives to disperse economic activities geographically are usually considered. The emergence of these firms is explained as 1) a response to tendencies of factor rewards to differ across countries, 2) a way to save transport costs and tariffs, 3) a way to internalise international transactions, and 4) a response to international differences in tax systems. See e.g. Helpman and Krugman (1985) and Ethier (1986). The emergence of multinationals may, nevertheless, also be ascribed to the existence of uncertainty and the wish to avoid this. Although this reason for diversification has not been much discussed within the context of trade theory, it has been widely recognised within the theory of corporate policy under the notion of "diversification of location" — perhaps better described by the German word "Standort-Diversifikation". See Timmermann (1973), a.o.

A two-stage game is considered: at the first stage firms decide about total production capacity, and employ the inputs needed to exploit this capacity. If a firm wants to divide its production between locations in order to secure access to the foreign market, this decision will also have to be made at this stage. Each firm has two options: (1) to establish one plant at home; (2) to establish two plants, one at home and one abroad. This early in the game, the firms do not know what the actual trade cost will be. However, as the second stage is reached, the trade cost has been unveiled. But since the level of production and its localisation were already determined at a previous stage, the only decision the firm is left with at this stage, is how to distribute the output between the markets. In other words, at the second stage firms face a constrained profit maximisation problem. We assume competition at both stages to take the form of the Cournot variety. — A simplifying assumption, but which makes the model considerably more tractable.

A representative firm located in economy 1 (a "firm 1" for short), has to make two decisions before it knows what the trade costs will be: it has to decide about total production capacity, and also what share of the capacity, if any, to locate abroad in the larger economy 2. Regarding a representative firm in economy 2 (a "firm 2" for short), we assume that an international diversification is not considered, since by assumption, the export market is relatively small compared to the home market. Hence, at the first stage, this firm only decides on total production capacity to install in its home location.

It follows that a firm 2 will only be serving market 1 from its plant in location 2, i.e. through export. As for a firm 1, it will either serve market 2 from its factory in location 1, or possibly from a second plant in location 2, or from both plants if

insufficient capacity has been invested in the plant in location 2. To simplify the analysis, we exclude the possibility of a firm 1 serving its home market (market 1) from a plant in location 2.

What we wish to emphasise is, under which circumstances does a firm based in the small country (market s_1) wish to invest in capacity abroad in order to serve the foreign market from a local plant. Of special interest is how trade cost uncertainty affects a firm's decision whether or not to disperse production and invest in production facilities abroad.

2. The formal model

Demand curves are assumed linear. The demand that a firm experiences in country i is identified by the inverse demand curve

$$p_i = a - \frac{1}{s_i}\left\{n_i x_{ii} + n_j x_{ji}\right\} \quad i,j = 1,2, \quad i \neq j, \tag{1}$$

where x_{ii} denotes the sales of a single firm located in country i in his home country, and x_{ji} is the quantity produced by a single firm based in country j, and consumed in country i.

Defining the total production capacity of a representative firm based in economy 1 as k_1, the part of its capacity invested abroad as k_{12}, the total capacity of a representative firm based in economy 2 as k_2, and the firm i's exports as e_i, we may write:

- a firm 1's sales at home in location 1 as $x_{11} = k_1 - k_{12} - e_1$ (2)
- a firm 1's sales abroad in location 2 as $x_{12} = k_{12} + e_1$ (3)
- a firm 2's sales at home in location 2 as $x_{22} = k_2 - e_2$ (4)
- a firm 2's sales abroad in location 2 as $x_{21} = e_2$. (5)

We assume that the firms are always better off exploiting their capacity completely, since according to the model, capacity has to be installed, and inputs employed, before trade costs are known.

The firms' production involves a constant marginal cost, c_i. Since comparative advantages are not being considered, this marginal cost is assumed the same for all firms regardless of in which economy they are based, as long as their production takes place in their home location.

Exporting from one economy to another involves trade cost. The trade cost (T) is assumed to be invariant of the nationality of the exporting firm. The trade cost is subject to an uncertainty: there is a probability of $1-\rho$ that the trade cost will turn out as $t+\varepsilon$, and a probability of ρ that the trade cost is $t-\varepsilon$, and $0<\varepsilon\leq t$. The expected trade cost for the shipping of one unit of output between countries is obviously $t+(1-2\rho)\varepsilon$. If trade costs are high, i.e., equal to $t+\varepsilon$, international trade does not entail profit, but if they are lower than $t+\varepsilon$, trade is possible and profitable. In the former case, we imagine that the firms face a situation where their optimal export is zero, because non-negative profits from exports cannot be realised. In other words, there is a $1-\rho$ chance of international trade being "impeded", and the two separated markets being served from local plants alone, because foreign suppliers are not able to compete with the domestic suppliers at such a high level of trade costs.

Using the equations (1) - (5), and the information about the firms' costs, the profits of the representative firms in countries 1 and 2 respectively, can, in general, be expressed as

$$\pi_1 = (p_1 - c)(k_1 - k_{12} - e_1) + (p_2 - c)(k_{12} + e_1) - Te_1 - bk_{12} \tag{6}$$

$$\pi_2 = (p_2 - c)(k_2 - e_2) + (p_1 - c - T)e_2. \tag{7}$$

b depicts an additional marginal cost related to producing abroad. In other words, marginal cost incurred when producing at a firm 1's second plant would be $(c+b)$. One might object that, it would have been more correct to define b as an additional fixed cost related to the set-up of a new plant. However, in such a case, more precise information about market size of country 2, levels of trade costs, and the size of the additional fixed cost incurred by foreign direct investment, is required in order to compare the different alternatives of a firm with regard to supplying a foreign market.

In order to ensure that the division of capacity between countries actually represents an alternative to export for a firm 1, $b < t + \varepsilon$ must be met. But unless $t - \varepsilon < b$ is true, a firm will always prefer setting up a second plant abroad to the conducting of international trade. Hence, we need to assume that

$$t - \varepsilon < b < t + \varepsilon. \tag{8}$$

At the first stage the firms decide on capacity, and where to locate it. Since we mainly wish to focus on the geographic distribution of a firm 1's investments, i.e., the size of k_{12}, the determination of k_1 and k_2 is left out. The total capacity of each firm is

regarded as fixed — decided on at a stage zero. To derive k_{12}, we use backwards induction, and solve the second stage of the game first: at the second stage firms decide on how much to export, in other words, which fraction of the ex ante determined home production that should be sold in the foreign market. The optimised exports expressed in terms of capacities are subsequently used to substitute in the firms' first stage maximisation problem. Firms' expected profits are now defined solely by k_1, k_{12} and k_2.We maximise the expected profits with respect to k_{12}, and find the optimal international distribution of production capacities.

3. Investing abroad in the absence of trade cost uncertainty

Before we turn to the analysing of investment abroad under trade cost uncertainty, we take a quick look at the case in which trade costs are known and equal to the expected trade cost, $E[T]=t+(1-2\rho)\varepsilon$. — A case that will be referred to as the "certainty case", and depicted by the superscript "c". For technical convenience we define

$$t^E \equiv t+(1-2\rho)\varepsilon. \tag{9}$$

We intend to demonstrate which conditions must be satisfied, if investment abroad is to be undertaken in a situation where trade costs are *not* subject to uncertainty.

In the certainty case the profits of a firm 1 and a firm 2 respectively, take the forms

$$\pi_1^c=\left(p_1^c-c\right)\left(k_1-k_{12}^c-e_1^c\right)+\left(p_2^c-c\right)\left(k_{12}^c+e_1^c\right)-t^E e_1^c-bk_{12}^c, \tag{10}$$

$$\pi_2^c = (p_2^c - c)(k_2 - e_2^c) + (p_1^c - c - t^E)e_2^c. \tag{11}$$

Since trade costs are known for certain, firms possess the same information about costs when making the decision whether or not to invest abroad, as they do when exports are optimised. Hence, as strategic behaviour is absent, the game has a single stage with a simultaneous choice of exports and foreign direct investments.

As for the firms based in the small market, differentiating (10) with respect to e_1 and k_{12} while using equations (1) - (5), we obtain the necessary conditions for a profit maximum:

$$\frac{s_2}{s_1 + s_2} k_1 - \frac{s_1 s_2}{s_1 + s_2}(p_1^c - p_2^c) - \frac{s_1 s_2}{s_1 + s_2} t^E - e_1^c - k_{12}^c \leq 0, \quad e_1^c \geq 0 \tag{12}$$

$$\frac{s_2}{s_1 + s_2} k_1 - \frac{s_1 s_2}{s_1 + s_2}(p_1^c - p_2^c) - \frac{s_1 s_2}{s_1 + s_2} b - e_1^c - k_{12}^c \leq 0, \quad k_{12}^c \geq 0 \ . \tag{13}$$

From the expressions (12) and (13) it can be deduced that if, and only if, $t^E = b$, firms based in the small country may conduct exports as well as undertake investment in plants abroad. In this case the outcome is ambiguous because a manufacturer will ceteris paribus be indifferent between concentration and diversification of production. If $t^E < b$ no investment will be undertaken abroad, and both markets will be served by domestic and foreign suppliers. However, if $t^E > b$, then the firms based in the small country will choose to serve their foreign market by a local plant rather than by export. It follows that $t^E = b$ is a necessary — though not a sufficient — condition for investment abroad to be undertaken, while $t^E > b$ is a sufficient condition for foreign direct investment to be pursued.

Differentiating equation (11) with respect to e_2, and using equations (1) - (5) to substitute for prices, a firm 2's best response function in terms of capacities, is

$$e_2^c = \frac{s_1}{s_1+s_2}k_2 - \frac{n_1 s_2}{(1+n_2)(s_1+s_2)}k_1 - \frac{s_1 s_2}{(1+n_2)(s_1+s_2)}t^E + \frac{n_1}{1+n_2}e_1^c + \frac{n_1}{1+n_2}k_{12}^c > 0. \quad (14)$$

Since the firms in country 2 never consider to invest abroad, their exports, e_2, are always positive in the case of certain trade costs. In order to concentrate on the firms based in the small country, we shall not elaborate on the profit maximisation problem of the large country firms any further.

From the first order conditions, the equations characterising the demand functions, and the firm 2's best response function, i.e., the equations (1) - (5) and (12) - (14), profit maximising k_{12} and e_1 can be derived. Depending on the value of b relative to t^E, a representative firm in country 1 will export and/or invest in production facilities abroad in accordance with the expressions below:

$$e_1^c = \frac{s_2}{s_1+s_2}k_1 - \frac{(1+2n_2)s_1 s_2}{(1+n_1+n_2)(s_1+s_2)}t^E \quad \text{and} \quad k_{12}^c=0 \quad \text{if } t^E < b \quad (15)$$

$$e_1^c + k_{12}^c = \frac{s_2}{s_1+s_2}k_1 - \frac{(1+2n_2)s_1 s_2}{(1+n_1+n_2)(s_1+s_2)}t^E \quad \text{if } t^E = b \quad (16)$$

$$k_{12}^c = \frac{s_2}{s_1+s_2}k_1 - \frac{(1+2n_2)s_1 s_2}{(1+n_1+n_2)(s_1+s_2)}b \quad \text{and} \quad e_1^c=0 \quad \text{if } t^E > b. \quad (17)$$

The results are illustrated graphically in figure 1:

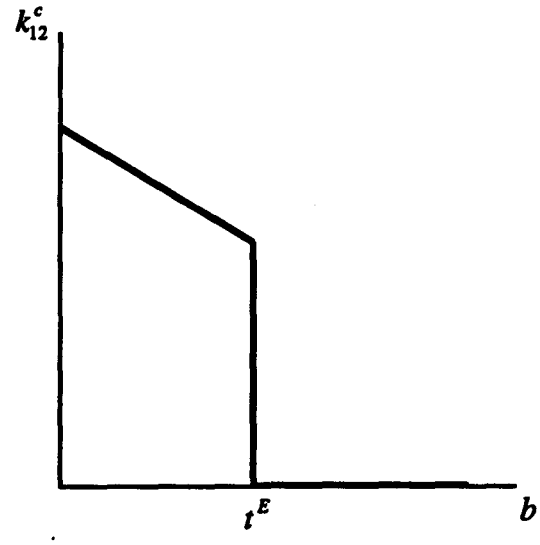


Figure 1

Decisive for the results obtained above, is the assumption about certainty. When deciding how to distribute production capacity internationally, every firm knows what the future trade cost will be. The decision how investment should be divided geographically, is based on cost considerations. Thus, a firm invests in production capacity abroad if, and only if, this results in costs lower than, or equal to, those that would have been incurred if all production had taken place in its home country plant. If the profits from serving the foreign market are the same invariantly of how this market is served — from a separate plant in the foreign market or by exports — other variables than those considered in this model, will determine the outcome.

4. Investing abroad under trade cost uncertainty

We now turn to the case where trade costs are subject to uncertainty, and may appear as high or low. The superscript "u" depicts the presence of uncertainty, while "l" and "h" depict the cases with low and high trade costs respectively. If trade costs are low the firms have access to their respective foreign markets through trade, and trade is presumed to be profitable. Optimally chosen exports e_1 and e_2 in terms of capacities, are found by differentiating the firms' profits. Profits under low trade costs for a firm 1 and a firm 2 respectively, are

$$\pi_1^{ul} = \left(p_1^{ul} - c\right)\left(k_1 - k_{12}^u - e_1^{ul}\right) + \left(p_2^{ul} - c\right)\left(k_{12}^u + e_1^{ul}\right) - (t-\varepsilon)e_1^{ul} - bk_{12}^u \qquad (18)$$

$$\pi_2^{ul} = \left(p_2^{ul} - c\right)\left(k_2 - e_2^{ul}\right) + \left(p_1^{ul} - c - t + \varepsilon\right)e_2^{ul}, \qquad (19)$$

and the first order conditions can be written as

$$\frac{s_2}{s_1+s_2}k_1 - k_{12}^u - \frac{n_2 s_1}{(1+n_1)(s_1+s_2)}k_2 - \frac{s_1 s_2}{(1+n_1)(s_1+s_2)}(t-\varepsilon) + \frac{n_2}{1+n_1}e_2^{ul} - e_1^{ul} \le 0,$$

$$e_1^{ul} \ge 0 \qquad (20)$$

and

$$\frac{s_1}{s_1+s_2}k_2 - \frac{n_1 s_2}{(1+n_2)(s_1+s_2)}k_1 - \frac{s_1 s_2}{(1+n_2)(s_1+s_2)}(t-\varepsilon) + \frac{n_1}{1+n_2}\left(e_1^{ul} + k_{12}^u\right) - e_2^{ul} \le 0,$$

$$e_2^{ul} \ge 0. \qquad (21)$$

By assumption, optimised $e_1, e_2 > 0$ if trade costs are low. From the first order conditions we derive the firms' best response functions. Substituting for e_2^{ul} and e_1^{ul}

respectively, in the best response functions profit maximising exports for a firm 1 and a firm 2 are obtained:

$$e_1^{ul} = \frac{s_2}{s_1+s_2}k_1 - k_{12}^u - \frac{(1+2n_2)s_1s_2}{(1+n_1+n_2)(s_1+s_2)}(\mathrm{t}-\varepsilon) \tag{22}$$

$$e_2^{ul} = \frac{s_1}{s_1+s_2}k_2 - \frac{(1+2n_1)s_1s_2}{(1+n_1+n_2)(s_1+s_2)}(\mathrm{t}-\varepsilon). \tag{23}$$

Using equations (1) - (5) and the expressions for the optimal exports under low trade costs, the equilibrium prices under low trade costs can be written as functions of capacities

$$p_1^{ul} = a - \frac{1}{s_1+s_2}\left\{n_1k_1 + n_2k_2 + \frac{(n_1-n_2)s_2}{(1+n_1+n_2)}(t-\varepsilon)\right\} \tag{24}$$

$$p_2^{ul} = a - \frac{1}{s_1+s_2}\left\{n_1k_1 + n_2k_2 + \frac{(n_2-n_1)s_1}{(1+n_1+n_2)}(t-\varepsilon)\right\}. \tag{25}$$

However, trade costs are just as likely to turn out high. In this case the profits of a firm 1 and a firm 2 take the forms

$$\pi_1^{uh} = \left(p_1^{uh}-c\right)\left(k_1 - k_{12}^u - e_1^{uh}\right) + \left(p_2^{uh}-c\right)\left(k_{12}^u + e_1^{uh}\right) - (t+\varepsilon)e_1^{uh} - bk_{12}^u \tag{26}$$

$$\pi_2^{uh} = \left(p_2^{uh}-c\right)\left(k_2 - e_2^{uh}\right) + \left(p_1^{uh}-c-t-\varepsilon\right)e_2^{uh}, \tag{27}$$

and international trade is presumed never to be profitable. This assumption implies that optimally chosen exports always equal zero, $e_1 = e_2 = 0$. This will be true if, and only

if, the first order conditions for a firm 1 and a firm 2 respectively, hold with inequality, which requires that the conditions (28) and (29) are satisfied:

$$\frac{s_2}{s_1+s_2}k_1 - k_{12}^u - \frac{n_2 s_1}{(1+n_1)(s_1+s_2)}k_2 - \frac{s_1 s_2}{(1+n_1)(s_1+s_2)}(t+\varepsilon) < 0 \qquad (28)$$

$$e_1^{wh} = 0,$$

$$\frac{s_1}{s_1+s_2}k_2 - \frac{n_1 s_2}{(1+n_2)(s_1+s_2)}k_1 + \frac{n_1}{1+n_2}k_{12}^u - \frac{s_1 s_2}{(1+n_2)(s_1+s_2)}(t+\varepsilon) < 0 \qquad (29)$$

$$e_2^{wh} = 0.$$

When there is no trade due to trade costs obstructing foreign manufacturers from competing with domestic manufacturers, the equilibrium prices in terms of capacities take the forms

$$p_1^{wh} = a - \frac{1}{s_1}\{n_1 k_1 - n_1 k_{12}^u\} \qquad (30)$$

$$p_2^{wh} = a - \frac{1}{s_2}\{n_2 k_2 + n_1 k_{12}^u\}. \qquad (31)$$

Having solved the firms' second stage maximisation problem, we turn to the decision facing the firms at the first stage. The investment plan of a representative firm in country 1, which identifies how much capacity to invest at home and how much to invest abroad, is obtained by choosing k_{12}^u to maximise expected profits. A firm 1's expected profits, $E[\pi_1]$, may be defined in terms of capacities, using the results generated above on trade and equilibrium prices under different levels of trade costs:

$$
\begin{aligned}
&\mathrm{E}[\pi_1] = \\
&(1-\rho)\cdot\left[\left(p_1^{wh}-c\right)\cdot\left(k_1-k_{12}^u\right)+\left(p_2^{wh}-c-b\right)\cdot k_{12}^u\right] \\
&+\rho\cdot\left[\left(p_1^{wl}-c\right)\left(\frac{s_1}{s_1+s_2}k_1+\frac{(1+2n_2)s_1s_2}{(1+n_1+n_2)(s_1+s_2)}(t-\varepsilon)\right)\right. \\
&+\left(p_2^{wl}-c\right)\left(\frac{s_2}{s_1+s_2}k_1-\frac{(1+2n_2)s_1s_2}{(1+n_1+n_2)(s_1+s_2)}(t-\varepsilon)\right) \qquad (32) \\
&\left.-(\mathrm{t}-\varepsilon)\cdot\left(\frac{s_2}{s_1+s_2}k_1-k_{12}^u-\frac{(1+2n_2)s_1s_2}{(1+n_1+n_2)(s_1+s_2)}(t-\varepsilon)\right)-bk_{12}^u\right]
\end{aligned}
$$

Substituting for the equilibrium prices using equations (24), (25), (30) and (31), and differentiating with respect to k_{12}^u, we obtain the necessary condition for a profit maximum:

$$
\frac{s_2}{s_1+s_2}k_1-\frac{s_1n_2}{2n_1(s_1+s_2)}k_2-\frac{s_1s_2}{2n_1(1-\rho)(s_1+s_2)}b+\frac{\rho s_1s_2}{2n_1(1-\rho)(s_1+s_2)}(t-\varepsilon)-k_{12}^u\le 0
$$

for $k_{12}^u \ge 0$. (33)

If a dispersion of capacities between the two countries is preferred, i.e., optimal $k_{12}^u>0$, the fraction of a firm 1's total capacity, which is invested in production facilities abroad, can be derived from equation (33):

$$
k_{12}^u=\frac{s_2}{s_1+s_2}k_1-\frac{s_1n_2}{2n_1(s_1+s_2)}k_2+\frac{\rho s_1s_2}{2n_1(1-\rho)(s_1+s_2)}(t-\varepsilon)-\frac{s_1s_2}{2n_1(1-\rho)(s_1+s_2)}b. \qquad (34)
$$

Investigation of the expressions (28), (29) and (34) tells us under what circumstances $k_{12}^u>0$ can be verified, i.e., under what circumstances a firm based in the small country 1 will choose to diversify its investments in production facilities internationally. The precise procedure followed in order to derive the conditions for a

positive k_{12}^{u} under trade cost uncertainty is rendered in the appendix. Here we shall concentrate on the main results.

In the certainty case, the additional marginal cost of producing abroad being equal to the expected trade cost ($b = t^E$), was found to be a necessary, but not a sufficient condition for firms to diversify. Hence, in order to derive whether the presence of uncertainty affects a firm's investment behaviour, we may look at the investment strategy adopted by the firm when $b = t^E$ and trade arrangements are subject to uncertainty. Doing this we find that

$$n_1 > 0 \tag{35}$$

and

$$\frac{n_1 k_1}{s_1} > \frac{2n_2 + 1}{3n_2} \cdot \frac{n_2 k_2}{s_2} \tag{36}$$

are necessary as well as sufficient conditions for small country based firms to diversify investments internationally under uncertainty when $b = t^E$.

A closer inspection of expressions (36) and reveals that, the more national competitors a manufacturer has, and the more pronounced his export dependence, i.e., the larger the market 2, the more likely, ceteris paribus, is he to be affected by uncertain trade arrangements and to reallocate part of his production internationally as a response to the uncertainty. Furthermore, from (A13) it can be derived that, if the "relative production capacity" of small country based firms ($n_1 k_1 / s_1$) exceeds that of large country firms ($n_2 k_2 / s_2$), and there is more than one firm based in the small country, then small country firms will always wish to diversify investment internationally when $b = t^E$.

As for the probability of high and low trade costs respectively, we find that provided that foreign investment is conducted, the lower ρ, i.e., the higher the probability of high trade costs "impeding" trade, the larger fraction of his production does a manufacturer choose to locate abroad (see expression (34)).

Comparing the results above with those obtained in the certainty case, we find that the alternative to diversify production — implying foreign direct investment — becomes more attractive to firms in a situation with uncertain trade costs. We remember that in the certainty case, $b \leq t^E$ was a necessary condition for firms to diversify, while $b < t^E$ had to be satisfied for a diversification to be guaranteed. In the case of uncertainty, we have shown that additional marginal costs related to manufacturing abroad equal to expected trade costs, $b = t^E$, may, in fact, be a sufficient condition for firms to diversify geographically. Examining the expressions (35), (36) and (A14), apart from the degree of export dependence, the number of firms based in the small country appears as especially critical to the obtaining of these results. In order to understand why the number of national competitors affects a firm's investment strategy under uncertainty, we need to take a closer look at the model as well as at the assumptions underlying it.

The main characteristics of the model are that capacities have to be decided on ex ante, and that firms based in the same country all conduct the same investment strategy. Let us compare the profit of a small country based firm that diversifies, with that of a firm that concentrates production at home. Given that trade costs turn out low, it follows from (32) that the firm will "lose" $(b-(t-\varepsilon))$ multiplied by the fraction of its capacity that has been located abroad (k_{12}^u). However, if trade costs turn out to be high, having invested in capacity abroad will imply increased profits relative to the alternative of

having concentrated production at home. Why? First, because the strategy of diversification secures access to — and profits from — the large foreign market. Second, since the total capacity of a firm is ex ante fixed, the larger the fraction of a firm's total capacity that is allocated abroad, the less the supply at home, and the higher the price obtained in the home market. The more national competitors with the same strategy as his own a manufacturer has, the larger the gains from diversifying relatively to those obtained from concentration, since the price increase experienced in the home market is positively correlated with the number of firms diversifying. In fact, an increase in the number of national competitors also reduces the price obtained in the foreign market. However, due to the difference in market size, and thereby in the demand functions, the price reduction abroad stemming from the diversification of one more country 1 firm, is less significant than the domestic price increase caused by this firm's behaviour. (See the last section of the appendix, and in particular equations (A16) and (A17).)

It follows that, while the potential losses from diversifying are independent of the number — and behaviour — of national competitors, the potential gains from diversifying are larger, the higher the number of such competitors.

Thus, uncertainty about future policy regimes may encourage geographical diversification of production. Yet, although the presence of trade cost uncertainty may provide firms with an incentive to conduct foreign direct investment, an increased uncertainty, i.e., an increase in the variance in trade costs reduces the optimal amount of capacity allocated abroad (see equation (34)).[10] — A result stemming from the fact that there is a higher degree of flexibility connected to the alternative of concentrating

[10] In statistical theory, an increased variance in trade costs depicts an increased uncertainty. A discussion of this as a concept of increasing risk, is found in Rotschild and Stiglitz (1970).

all production at home versus that of dividing production between the two countries, and that the gains from this flexibility are higher the greater the variance in trade costs.[11] The promoting effect of an increased trade cost variance on the "export alternative", is reinforced by one of the assumptions underlying the model: since export is assumed always to be impeded when trade costs are high, an increased variance does not affect profits in the situation where trade costs resolve as high. But with regard to the situation with low trade costs, an increased variance in trade costs implies reduced costs of exporting, and makes export a more attractive alternative relative to that of diversifying investments geographically.

5. Final remarks

It has been demonstrated that under trade cost uncertainty, an oligopolist with extensive exports may diversify investments and allocate part of his production abroad, in order to secure market access to a large export market. Even though the unit supply cost that he has to take into account when diversifying production internationally, exceeds the expected unit supply cost associated with concentrating all production in his home country, diversification may still occur to him as the more attractive alternative.

It follows that uncertainty about international arrangements and their future may affect firms' investment behaviour, no matter what their attitudes towards risk might be. Even if a firm is risk neutral, its decisions may be affected by uncertainty. Hence, risk

[11] When serving the foreign market from a local plant, a firm earns a constant profit. But serving this market through export, its profit varies with the level of trade costs. An increased variance incurs an increase in profits in the case of low trade costs, which is more substantial than the reduction in profits experienced if trade costs are high.

aversion is not a necessary condition for firms' investment strategy to be influenced by uncertainty. Firms with extensive exports may respond to uncertain international arrangements and trade costs by reallocating production to the export market both in fear of a market foreclosure, and in order to increase profits in the home market by committing to an investment strategy that implies increased sales abroad and reduced sales at home. The consequences for the remaining economic activity in the country they "leave behind", and for the country as a whole, will be serious.

The negative effects on growth and welfare will be further enhanced if external economies of industrial agglomeration are important. If external economies of industrial agglomeration are important, then a reallocation of firms will also reduce the competitiveness of remaining firms and industries, and have dramatic consequences for a country's growth and welfare.

The dependence of firms' foreign direct investments on the perceived probabilities of low and high trade costs (i.e. the size of ρ) also has policy implications. Our results suggest that, a policy aiming at avoiding or minimising the flight of firms and production, should be framed to influence people's perception of ρ. Given relatively fierce competition in the home market and a serious export dependence, it becomes important to convey the impression that there is only a small probability of trade becoming "impeded".

Appendix

Expression (34), i.e.,

$$k_{12}^{u}=\frac{s_2}{s_1+s_2}k_1-\frac{s_1n_2}{2n_1(s_1+s_2)}k_2+\frac{\rho s_1s_2}{2n_1(1-\rho)(s_1+s_2)}(t-\varepsilon)-\frac{s_1s_2}{2n_1(1-\rho)(s_1+s_2)}b,$$

tells us the size of k_{12}^{u}, given that k_{12}^{u} is positive. It follows that $k_{12}^{u}>0$ implies

$$\frac{s_2}{s_1+s_2}k_1-\frac{s_1n_2}{2n_1(s_1+s_2)}k_2+\frac{\rho s_1s_2}{2n_1(1-\rho)(s_1+s_2)}(t-\varepsilon)-\frac{s_1s_2}{2n_1(1-\rho)(s_1+s_2)}b>0. \quad \text{(A1)}$$

(A1) may be rewritten as

$$(1-\rho)\cdot\left[\frac{2n_1}{s_1}k_1-\frac{n_2}{s_2}k_2\right]-\rho\cdot(t-\varepsilon)>b. \quad \text{(A2)}$$

If $b=t^{E}=(1-\rho)\cdot(t+\varepsilon)+\rho\cdot(t-\varepsilon)=t+(1-2\rho)\varepsilon$, then from (A2) follows that

$$\frac{2n_1}{s_1}k_1-\frac{n_2}{s_2}k_2-t>\varepsilon \quad \text{(A3)}$$

must hold for $k_{12}^{u}>0$ to be true.

While (34) provides us with the restriction that has to hold for k_{12}^{u} to be positive, the expressions (28) and (29) create an interval of values within which k_{12}^{u} has to lie, for optimal exports under high trade costs actually to be zero.

From (28) it can be derived that

$$k_{12}^{u} > \frac{s_2}{s_1+s_2}k_1 - \frac{n_2 s_1}{(1+n_1)(s_1+s_2)}k_2 - \frac{s_1 s_2}{(1+n_1)(s_1+s_2)}(t+\varepsilon), \tag{A4}$$

and from (29) it can be derived that

$$k_{12}^{u} < \frac{s_2}{s_1+s_2}k_1 - \frac{(1+n_2)s_1}{n_1(s_1+s_2)}k_2 + \frac{s_1 s_2}{n_1(s_1+s_2)}(t+\varepsilon). \tag{A5}$$

Using (34) to substitute for k_{12}^{u} in (A4) and (A5) respectively, we get

$$\begin{aligned} &n_2(1-\rho)(n_1-1)k_2 + s_2\big(2n_1-\rho(n_1-1)\big)t + s_2\big(n_1(2-3\rho)-\rho\big)\varepsilon \\ &-s_2(1+n_1)b > 0 \end{aligned} \tag{A6}$$

and

$$(1-\rho)(n_2+2)k_2 - s_2(2-3\rho)t - s_2(2-\rho)\varepsilon - s_2 b < 0. \tag{A7}$$

Hence, in order for $k_{12}^{u}>0$ to be true, in addition to the inequality (A2), the inequalities (A6) and (A7) also have to hold.

Examining the specific case where $b = t^{E} = t + (1-2\rho)\varepsilon$, we remember that (A3) had to hold for (A2) to be satisfied. Substituting for $b = t^{E}$ in (A6), we obtain

$$(n_1-1)\big[n_2 k_2 + s_2(t+\varepsilon)\big] > 0 \tag{A8}$$

which holds as long as $n_1 > 1$. Thus, (A6) will always be satisfied for $b = t^E$ provided that there is more than one firm based in the small country. Proceeding by substituting for $b = t^E$ in (A7) we get

$$(n_2 + 2)k_2 - 3s_2(t + \varepsilon) < 0, \tag{A9}$$

which may be rewritten as

$$\frac{2 + n_2}{3s_2} k_2 - t < \varepsilon. \tag{A10}$$

Since (A7) was found always to hold for $n_1 > 1$, we have that $n_1 > 1$ together with the inequalities (A3) and (A10) are necessary as well as sufficient conditions for $k_{12}^u > 0$ to be true for $b = t^E$. (A3) and (A10) define an interval within which ε must lie for $k_{12}^u > 0$ to hold:

$$\underbrace{\frac{2 + n_2}{3s_2} k_2 - t}_{=A} < \varepsilon < \underbrace{\frac{2n_1}{s_1} k_1 - \frac{n_2}{s_2} k_2 - t}_{=B} \tag{A11}$$

Accordingly, we are left with merely two necessary and sufficient conditions for $k_{12}^u > 0$ to hold, namely A<B and $n_1 > 1$. Hence, provided that these two conditions are satisfied, $k_{12}^u > 0$ for certain values of ε. A<B is equivalent with

$$\underbrace{\frac{n_1 k_1}{s_1} - \underbrace{\left(\frac{2n_2 + 1}{3n_2}\right)}_{=C} \cdot \frac{n_2 k_2}{s_2}}_{=D} > 0, \tag{A12}$$

and since $n_2 \geq 1$ implies $C \leq 1$, (A12) will always be satisfied for

$$\frac{n_1 k_1}{s_1} > \frac{n_2 k_2}{s_2}. \tag{A13}$$

A closer inspection of (A12) reveals that

$$\frac{\partial D}{\partial n_1}, \frac{\partial D}{\partial s_2}, \frac{\partial D}{\partial k_1} > 0 \tag{A14}$$

and

$$\frac{\partial D}{\partial n_2}, \frac{\partial D}{\partial s_1}, \frac{\partial D}{\partial k_2} < 0. \tag{A15}$$

The reason why the number of country 1 based firms is decisive for the outcome:

The model presented in this paper relies on the assumption of all firms based in a country conducting the same investment strategy. If we let $E[\pi_1^D]$ depict the expected profit of a small country based firm that chooses to diversify investments internationally, and $E[\pi_1^{CO}]$ depict the profit of a small country based firm choosing to concentrate all its production at home, it can be derived that

$$\begin{aligned} &E[\pi_1^D] - E[\pi_1^{CO}] = \\ &(1-\rho)\left(2s_2 n_1 k_1 - (s_1+s_2)n_1 k_{12}^u - s_1 n_2 k_2 - s_1 s_2 b\right)\frac{1}{s_1 s_2}k_{12}^u - \rho(b-(t-\varepsilon))k_{12}^u \end{aligned} \tag{A16}$$

Differentiating with respect to n_1, we find that

$$\frac{\partial\left(E[\pi_1^D] - E[\pi_1^{CO}]\right)}{\partial n_1} > 0. \tag{A17}$$

In other words, the more "neighbouring" firms that conduct the same strategy as his own, the higher the small country firm's profit from diversifying, relative to concentrating, production.

References

Baldwin, Richard and Paul R. Krugman (1989): "Persistent trade effects of large exchange rate shocks", *Quarterly Journal of Economics* 104 (1989): 635-654

Baldwin, Richard (1994): *Towards an Integrated Europe*, CEPR, 1994

CEPR (1992): *Is bigger better? The economics of the EC enlargement.*, 1992

Cheng, Leonard (1983): "Ex ante plant design, portfolio theory and uncertain terms of trade", *Journal of International Economics* 14 (1983): 25-51

Commission of the European Communities (1988): *The economics of 1992, European Economy,* No. 35. Known as the *Cecchini report.*

Dixit, Avinash (1989a): "Entry and exit decisions under uncertainty", *Journal of Political Economy* 97 (1989): 620-638

Dixit, Avinash (1989b): "Hysteresis, import penetration, and exchange rate pass-through", *Quarterly Journal of Economics* 104 (1989): 205-228

Dumas, Bernard (1980): "The theorems of international trade under generalised uncertainty", *Journal of International Economics* 10 (1980): 481-498

Ethier, Wilfred J. (1986): "The multinational firm", *Quarterly Journal of Economics* 101 (1986): 805-834

Grossman, Gene and Assaf Razin (1985): "The pattern of trade in a Ricardian model with country-specific uncertainty", *International Economic Review* 26 (1985): 193-202

Haaland, Jan I. (1993): "EF eller EØS?", *Dagens Næringsliv* 08.12.93

Helpman, Elhanan and Paul R. Krugman (1985): *Market structure and foreign trade*, Cambridge, Massachusetts: The MIT Press, 1985

Helpman, Elhanan and Assaf Razin (1978a): "Uncertainty and international trade in the presence of stock markets", *Review of Economic Studies* 45 (1978): 239-250

Helpman, Elhanan and Assaf Razin (1978b): *A theory of international trade under uncertainty*, Academic Press, New York 1978

Hoel, Michael and Karl Ove Moene: *Produksjonsteori*, 2.utg., Oslo: Universitetsforlaget, 1993

Kreps, David M. (1990): *A course in microeconomic theory*, New York: Harvester Wheatsheaf, 1990

Krugman, Paul R and Anthony J. Venables (1990): "Integration and the competitiveness of peripheral industry", in Bliss, Christopher J. and Jorge Braga de Macedo (eds.): *Unity with diversity in the European economy: the Community's southern frontier*, Cambridge: Cambridge University Press, 1990

Krugman, Paul R. (1991): *Geography and Trade*, Cambridge, Massachusetts: MIT Press, 1991

Krugman, Paul R. and Anthony J. Venables (1993): "Integration, specialization and adjustment", NBER Working paper No. 4559, 1993

Mai, Chao-Cheng (1984): "Location and the theory of the imperfectly competitive firm under demand uncertainty", *Southern Economic Journal* 50 (1984): 1160-1170

Park, Keehwan and Vijay K. Mathur (1990): "Uncertainty in location", *Journal of Urban Economics* 27 (1990): 362-380

Rothschild, Michael and Joseph E. Stiglitz (1970): "Increasing risk: I. A definition", *Journal of economic theory* 2 (1970): 225-243

Sandmo, Agnar (1971): "On the theory of the competitive firm under price uncertainty", *American Economic Review* 61 (1971): 65-73

Timmermann, Manfred L. (1973): "Standort-Diversifikation" als Instrument der Unternehmenspolitik", *Die Unternehmung* (1973): 41-49

van Wijnbergen, Sweder (1985): "Trade reform, aggregate investment and capital flight", *Economics Letters* 19 (1985): 369-372

Venables, Anthony J. (1990): "International capacity choice and national market games", *Journal of International Economics* 29 (1990): 23-42

Venables, Anthony J. (1993): "Equilibrium locations of vertically linked industries", CEPR Discussion paper No. 802, London, 1993

Essay 3

Technological spillovers, industrial clusters and economic integration

by

Karen Helene Midelfart Knarvik*

Introduction

National and international trade is characterised by regional specialisation and disparities. Differences in factor endowments, like in the Heckscher-Ohlin model, may be one reason for regional specialisation resulting in regional agglomeration within certain industries. An alternative or complementary explanation, however, is the existence of localised external economies. The most cited concept of external economies is developed by Marshall.[1] Marshall identifies three distinct reasons for the localisation of particular industries, in other words three bases for regional divergence: First, industrial agglomeration offers a pooled market for workers with industry-specific skills, ensuring a lower probability of unemployment as well as labour shortage. Second, the availability of specialised inputs and services, i.e. vertically linked industries, encourage concentration of upstream and downstream firms. Finally, knowledge spillovers, such as technological spillovers, which are generated in industrial centres, leave clustered firms with a more efficient production than

* I would like to thank Jan I. Haaland, Victor D. Norman and Anthony J. Venables as well as Diego Puga and Øystein Thøgersen for valuable comments and suggestions as well as inspiring ideas.

[1] See Marshall (1920).

"isolated" firms.[2] Characterising the latter kind of external economies encouraging industrial concentration, is that knowledge acquired by one firm can be used by others, provided that they are localised within the same area. Famous examples of clusters presumed to rely on the existence of technological spillovers, are Silicon Valley in California, the Route 128 near Boston, and the cluster along the M4 west of London.[3]

Economic geography and localised external effects as a source of economic effects have during the last years experienced increasing attention among economists. The emerging literature on this field, which is often referred to as the "new" location theory, studies the geographic clustering of economic activity and regional evolution in terms of convergence and divergence.[4] It is characterised through models with *endogenous* agglomeration economies.[5] In these models agglomeration is modelled without any direct assumptions about the existence and size of the external effects generating concentration. The externalities leading to clustering emerge as a consequence of market interactions alone.

Relying on the concept and tools of the "new" location theory, different models where agglomeration is caused by "visible" external effects have been developed. "Visible" in the sense that they arise from such tangible phenomena as for instance vertically linked industries. In other words, the two first kinds of external economies identified by

---

[2] Following Glaeser et al. (1992) technological spillovers are understood as some kind of knowledge spillovers.
[3] See Dicken and Lloyd (1990), pp. 173-174.
[4] See CEPR (1992), p 31.
[5] For examples of such models see e.g. Krugman (1991a), while Baldwin (1994) may be consulted for a brief and good survey of the literature.

Marshall are both examples of such "visible" externalities.[6] But although technological spillovers are recognised as one of the most important reasons for agglomeration in space, they have been somehow neglected by the economists working with the "new" location theory.[7] — Presumably because of the "invisibility" of such spillovers, and the general lack of knowledge of where technological spillovers go, their form and significance.

The aim of this paper is, within the context of the "new" location theory, to shed some light on the "invisible" technological spillovers, and how economic integration affects industrial structure when such spillovers are present. Focusing on intra-industry labour mobility as one possible source of such spillovers, we build a model with ingredients from trade theory and the "new" location theory, as well as from labour economics.[8] The model allows us to study the geographic agglomeration of firms in an industry with technological spillovers, and to elucidate reasons for regional divergence and convergence when countries become economically integrated.

Our approach to the phenomenon of technological spillovers was inspired by what seems like a discrepancy between a substantial share of the models within the labour market literature and empirical facts: The models we refer to, are based on an assumption about firm-specific human capital.[9] Working in a firm leads to an accumulation of job(firm)-specific human capital over time, so that a worker's productivity increases with job-tenure. One main prediction that comes out of these

---

[6] The distinction between visible and invisible externalities was proposed by Krugman in Krugman (1991b). Examples of models that focus on different kinds of "visible" external economies are found in Krugman (1991a), Krugman and Venables (1993), and Venables (1993).

[7] See e.g. the evidences found by Jaffe, Trajtenberg and Henderson (1993).

[8] It is important to recognise the difference between the kind of model we shall develop and those considering a pooled labour market as centripetal force. Krugman (1991b) and Rotemberg and Saloner (1990) provide a good description of the latter kind of external economies.

[9] Typical examples of such models are the search models. See for instance Pissarides (1992).

models, is that wages rise with job-tenure, leaving the workers with a stake in the value of the employment relationship. Reviewing the labour market data, this prediction is confirmed. But due to the purely firm-specific human capital, whether or not a worker possesses experience, and what kind of experience he has, does not matter to another employer, and does not influence a person's wage perspectives in a new firm. This characteristic of certain labour market models forms a sharp contrast to the data. In the data there is evidence that experience and earnings are positively correlated, and that job-to-job changes lead to wage-jumps.[10] If all human capital is firm specific, there is no reason why changing jobs should incur a wage increase.

Relaxing the assumption about human capital being completely firm-specific, we hope to be able to develop a model more consistent with empirical facts, and where job-changing is an essential component of a worker's career planning. However, job-mobility is not only of importance to the workers. It may be just as important to firms, as the mobility of skilled personnel between firms represents a vehicle for the spread of knowledge among firms. A common view, supported by empirical evidence, is that geographical proximity facilitates transmission of knowledge. Hence, industrial clusters and cities are institutions reckoned as providing an environment in which ideas are quickly disseminated among firms and enhance the overall productivity. Glaeser et al. (1992) mention three good examples of how inter-firm movement enables ideas to flow among neighbouring firms and thereby improving the firms' productivity: Among the microchips manufacturers in Silicon Valley, the fashion designers in New York, and in the Bangladeshi shirt industry, the physical proximity has promoted information transmission in the form of employees moving between firms.

[10] See e.g. Topel and Ward (1992). The authors find that wage gains at job changes account for at least a third of early-career wage growth.

In accordance with the empirical evidence, we shall assume that firms benefit from other firms' knowledge through workers' movement among firms. From a worker's point of view, geographic mobility is expensive, not just in a pecuniary sense, so firms cannot attract experienced workers unless they locate in an industrial centre.

1. Basic elements of the model

We introduce a general equilibrium model of intra-industry trade following the line of work on geography and trade starting with the article by Krugman and Venables (1990). Some of the tools we shall employ here to analyse the location of economic activity, were developed by the same authors and adopted by numerous other economists working on the "new" location theory.[11] Internal economies of scale and trade costs are typical of this new direction within trade theory, and represent the reasons why location matters to firms. The scale economies force firms to decide where to locate — a question that could be ignored as long as constant returns to scale and perfect competition were assumed. Since there are trade costs, the firms' location will have implications for sales and profits.

Our model exhibits a tension between two kinds of forces: "centripetal" forces that tend to pull economic activity into spatial agglomerations, and "centrifugal" forces that tend to push such agglomerations apart. Concentration in space is encouraged by internal economies of scale and the existence of knowledge spillovers, the latter being ensured by labour turnover. Trade costs and spatially dispersed consumers, which

[11] See Krugman (1991a), the first of a series of influential papers where the mentioned "tools" appear. The model we shall develop is mainly based on Venables (1993), and Krugman and Venables (1993).

both discourage concentration, work in the opposite direction. Hence, the geographical location of economic activity and the degree of industrial agglomeration in space, are determined by the interaction of trade costs, internal increasing returns to scale, market access and access to technological spillovers.

There are two economies, i=1,2, two sectors in each economy and international trade. Economies and sectors are labelled by subscripts. One sector, sector *a*, is perfectly competitive with production subject to constant returns to scale. The homogenous goods produced in this sector are tradeable, and the trade of these goods is nationally and internationally costless. We choose this good as numeraire.

In sector *b*, production is subject to increasing returns to scale internal to the firm. Firms are heterogeneous, produce differentiated goods, and there are knowledge spillovers between firms. In contrast to the homogenous products produced in sector *a*, international trade of differentiated goods incurs trade costs.[12] The market structure in industry *b* is monopolistically competitive. In equilibrium all firms have the same size and earn zero profits due to free entry and exit. We model this industry using the monopolistic competition model developed by Dixit and Stiglitz (1977). Note that the forces driving the model are scale economies and trade costs, *not* the prevailing kind of competition.

[12] These trade costs have *almost* the same interpretation as the "trade costs" in Krugman and Venables (1990). Thus, trade costs should be thought of as a synthetic measure of a wide range of barriers to international trade, including trade taxes, transport costs, and costs of frontier formalities and differing product standards. Typical of the geography and trade models developed by Krugman and Venables (and adopted by numerous other economists), is that costs related to intra-regional (or domestic) trade are ignored. An exception is found in Martin and Rogers (1994), who provide a model where both intra-regional and international trade are associated with costs.

To simplify we assume that there is only one factor of production, labour.[13] Like Krugman and Venables (1993) we assume labour to be geographically immobile. This assumption characterises the so called "European" economic geography model within the "new" location theory. It is based on the fact that labour in the European Union and the European Economic Area is culturally and emotionally tied to its origins, and therefore reluctant towards migrating. One should be aware that the proceeding arguments are all based on geographical labour immobility. In the succeeding paragraph we describe the labour market in one economy.

Each economy is endowed with a certain number of labour units. The representative consumer in each country receives labour income only. Consumers are identical across countries, i.e. have the same utility function, and use the same share of their income on industry *b* products.

Ultimately, one should note that the model adopts a static general equilibrium approach. We concentrate on describing the range of equilibria, without elaborating how the economies evolve during the adjustment to a new equilibrium.

2. The labour market

The modelling of each country's immobile population is based on Blanchard's model of "Perpetual youth".[14] To simplify, we shall assume that everyone starts working the

[13] As we shall see later, not one, but in fact two factors of production emerge: Skilled and unskilled labour. But for the time being we identify only the more general factor "labour", since we mainly wish to make the point that capital, land a.o. input factors are ignored in the model.
[14] See Blanchard et al. (1989). The "Perpetual youth" model is a continuous time version of an overlapping generations model.

day they are born, and that their working careers last for the rest of their lives, so that a country's labour stock exactly equals its population. Time is continuous and each individual, independent of age, faces a constant probability of death per unit of time, δ. (As Blanchard et al. (1989) points out, the assumption of the probability of dying being independent of age "..is clearly made for convenience rather than realism." (p.149)) The expected lifetime of all individuals at any point of time is thus δ^{-1}.

The instantaneous probability δ can take on any value between 0 and infinity. At each instant of time a new cohort, consisting of people with constant probability of death, δ, is born. We let the size of the new cohort be δL_i. Each cohort is large enough, so that the fraction which dies each instant is also δL_i. Thus, the size of the labour stock in country i at any time is constant and equal to L_i, i.e.,

$$L_i = \bar{L}_i. \tag{1}$$

In equilibrium labour is fully employed, so that

$$L_i = L_{ai} + L_{bi}. \tag{2}$$

We allow for intra- as well as inter-market labour mobility. Inter-market labour mobility occurs when firms migrate from one country to another, thereby altering the industrial structure of the countries. In steady state there is, however, no such inter-market mobility, and the size of the labour stock employed in each sector is constant, i.e.,

$$\dot{L}_{ai} = \dot{L}_{bi} = 0, \tag{3}$$

and there is a constant stream into each sector (δL_{ji}, $j = a,b$) exactly equal to the constant stream out of each sector. It follows that the labour market in a country i, i=1,2, can be illustrated graphically as below:

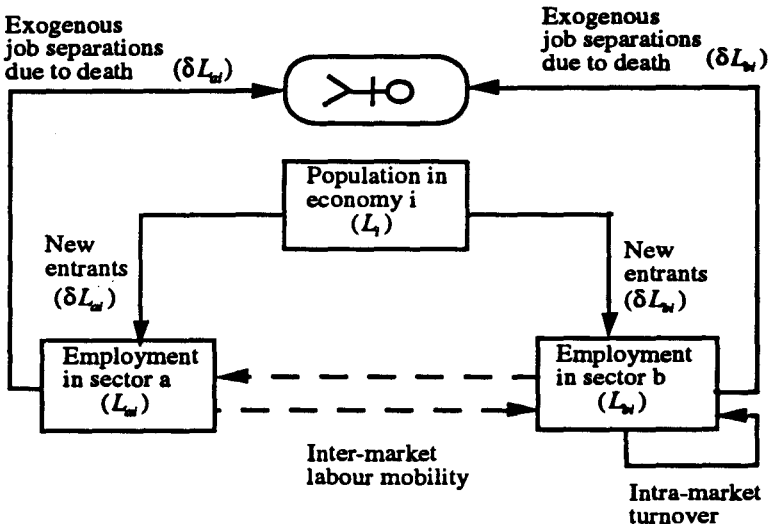


Figure 1

Tenure, experience and productivity

The work in sector *a* does not require any specific skills, any employee in this industry will have the same productivity, regardless of skills or experience. Nor is it possible to acquire any firm, industry or general skills through the work in industry *a*.

But in the sector *b*, things look quite different. Production here requires specific skills. These skills are obtained through learning-by-doing. The knowledge the employees acquire through their work is partly firm-specific and partly industry-specific. However, it is not possible to distinguish clearly between what knowledge is firm and what is industry specific. What we observe is an accumulation of human capital over

time. The human capital of a representative worker is a function of job-seniority (tenure), i.e., how long the person has been working for his current employer, and of job-experience from other firms within the industry. The accumulating human capital implies an increasing productivity over time. Knowing the time an employee has spent with his current employer, as well as the time he has spent in the industry, it is possible to deduce his current personal marginal product.

Ex ante, all individuals are homogenous — all are born unskilled. But from the moment they start working, they become heterogeneous. Those who are employed in industry *a* remain unskilled, while those who get a position in industry *b* start acquiring skills. So, it appears that there are actually two factors of production: skilled and unskilled labour, but each industry only employs one of these factors.

While working within a firm in industry *b*, a person's productivity increases until a certain level is reached, which happens at a tenure of τ^*. Provided that the worker stays in the *same* firm, his marginal product can be expressed as a function of tenure, $h(\tau)$, with $h'(\tau)>0$ for $\tau\in[0,\tau^*)$, and $h'(\tau)=0$ for $\tau\in[\tau^*,\infty]$. Consistent with empirical evidence, $h(\tau)$ is an S-formed function, increasing in τ until a certain tenure τ^* is reached.[15] We let $h(0)=h^m$, and $h(\tau)=h^*$ for $\tau\in[\tau^*,\infty]$, and illustrate $h(\tau)$ graphically:

[15] See Kostiuk and Follman (1989) a.o.

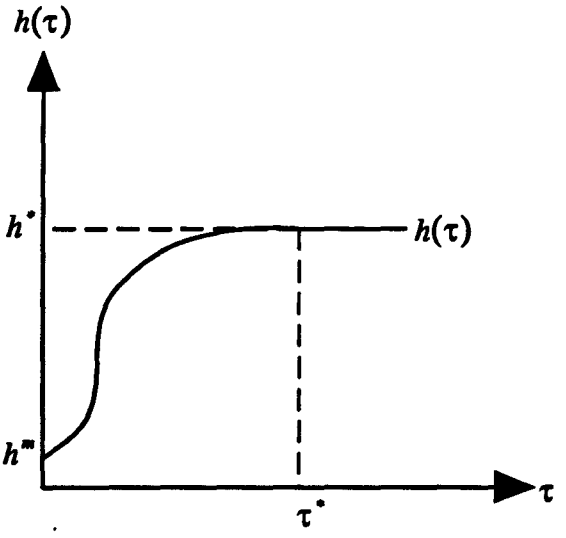


Figure 2

The knowledge a worker acquires while working in a firm k can be utilised not only in the manufacturing of his current employer, but also in the manufacturing of other employers in the industry. If $h(\tau)$ illustrates how his ability to produce k-products increases over time, it is possible to draw another curve, showing how his ability to produce l-products with the manufacturer l evolves over time. $H(\tau)$ depicts the latter curve and is illustrated below:

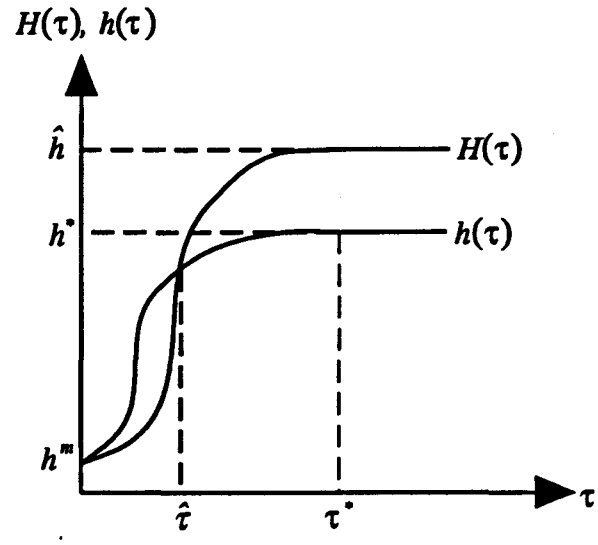


Figure 3

$H(\tau)$ can be explained as follows: When a person starts working in a specific firm, he starts acquiring industry- and firm-specific knowledge. In the beginning his knowledge will be fairly poor and firm-specific. The longer he stays in the firm, the more insights he gains and the better overview he attains. Consequently, his skills become more general. He is able to apply what he has learned into the solving of a wider range of problems. Until a certain tenure, $\hat{\tau}$, is reached, he is still of higher value to his current employer than he would be to any other firm in the industry. Due to his firm specific skills, his contribution to the production by his current employer is higher than it would be with any other firm in the industry.

However, having worked for a firm for a period equivalent to the tenure $\hat{\tau}$, he actually becomes more valuable to another firm in the industry, than to his current employer. Why? Because we assume the accumulation of a person's human capital to be affected

by the number of employers he has had, in other words, by the diversity of his experience. Such an assumption is consistent with newer theories on personnel management and organisational psychology, that stress the importance of diverse training for the development of effective organisations.[16] Since the manufacturers in industry *b* produce heterogeneous goods, the knowledge acquired through the work with one manufacturer is never equal to the knowledge acquired with another manufacturer in the industry. Observing two employees working in the same firm, one of whom possesses experience from a competitor and the other not, the former worker will be able to obtain a maximum level of productivity beyond that of the employee who has stayed with this particular employer all his life.

The assumption about the importance of "experienced"[17] workers that underlies function $H(\tau)$, is also based on the previously cited empirical evidence of job changes and wage growth.[18] From this evidence we draw the conclusion that experience from another firm in the industry matters to an employer. If this were not true, why should firms that advertise for experienced employees, be willing to offer an applicant a higher wage than he can get from his current employer?

Readers might object that we are likely to see a small downward shift in a person's productivity curve, just as he changes jobs. But we argue that when starting a new job, due to the learning process he has accomplished, the "experienced worker" learns

[16] See for instance Herriot and Pemberton (1995).
[17] The notion "experienced worker" is used to depict a person who has been employed by another firm in industry *b* before starting the job with his current employer in this industry. An "apprentice" depicts a worker in his first job in industry *b*.
[18] See Topel and Ward (1992) to which we refer in the introduction.

faster than an "apprentice", and is able to attain the necessary firm-specific knowledge more rapidly.[19]

In order to simplify the model, and focus on how a job-change affects an individual's marginal product, we redefine the $h(\tau)$ and $H(\tau)$ functions as step-functions. These functions reflect the way skills evolve over time less precisely than the original functions, though sufficiently for our purpose.

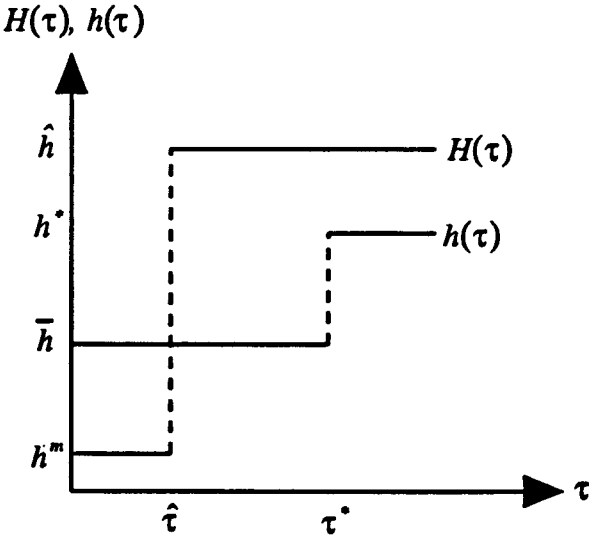


Figure 4

[19] Our argument is based on Norman (1972), who points out that cross-section evidence suggests that education facilitates rapid learning.

In accordance with the graphics above,

$$h(\tau)=\bar{h}, \quad \text{for } \tau\in[0,\tau^*\rangle, \; h(\tau)=h^*, \quad \text{for } \tau\in[\tau^*,\infty],$$

$$H(\tau)=h^m \text{ for } \tau\in[0,\hat{\tau}\rangle, \; H(\tau)=\hat{h} \text{ for } \tau\in[\hat{\tau},\infty], \tag{4}$$

and we note that $h^m < \bar{h} < h^* < \hat{h}$.

If there is just one industry *b* firm in a location *i*, intra-industry labour mobility certainly not occurs. However, when more than one firm is located in the economy, intra-market turnover is generated, and reaching the tenure $\hat{\tau}$, everyone changes jobs. To simplify the analysis and ensure that everyone stays in their second job throughout their lives, we assume that an increase in productivity cannot be obtained by a second job change within the industry. The workers' *incentive* to change jobs is elaborated in the next section. According to their individual marginal products we may divide workers in industry *b* in location *i* into three groups:

- $\bar{q}_i$ = apprentices.
- q_i^* = workers in their first job, who have reached the tenure $\hat{\tau}$, but are unable to change jobs because there is just one *b* firm in the country.
- $\hat{q}_i$ = experienced workers in their second job in the industry.

In steady state the distribution of the industry *b* labour force between these three groups is fixed. The table below displays how the labour stock in a country *i* is distributed across the three groups, depending on the number of firms in the country(n_i):

	$n_i=1$	$n_i>1$
$\bar{q}_i$	$\int_0^{\tau^*} \delta L_{bi} e^{-\delta\tau} d\tau = \left(1-e^{-\delta\tau^*}\right) L_{bi}$	$\int_0^{\hat{\tau}} \delta L_{bi} e^{-\delta\tau} d\tau = \left(1-e^{-\delta\hat{\tau}}\right) L_{bi}$
q_i^*	$\int_{\tau^*}^{\infty} \delta L_{bi} e^{-\delta\tau} d\tau = e^{-\delta\tau^*} L_{bi}$	0
$\hat{q}_i$	0	$\int_{\hat{\tau}}^{\infty} \delta L_{bi} e^{-\delta\tau} d\tau = e^{-\delta\hat{\tau}} L_{bi}$

Table 1

Finally, we add the assumption that the knowledge capital accumulation of a firm's labour force is characterised by economies of scope. Consistent with management theories that stress the importance of diverse recruitment, a firm is assumed to gain from having workers with qualitatively different experiences.[20] It follows that, not only does a firm prefer experienced workers, but it also prefers their respective backgrounds to differ, i.e., a firm has preferences for a variety of experienced workers.[21] Since the manufacturers in industry *b* produce heterogeneous goods, the skills acquired in one firm differ from those learned in another firm. Hence, the only way to obtain a variety of workers, is to attract people, who each has been employed by a different firm.

---

[20] It is the same management theories that stress the importance of diverse training. See footnote16 for references.

[21] This assumption resembles that of Abdel-Rahman and Fujita (1990); Krugman and Venables (1993); and Venables (1993). In these models the downstream industry has the desire to employ a variety of *intermediates*. In our model the industry also has preferences for differentiated inputs, but the inputs we focus on are people, not intermediates. (Although not an industry, a university provides a good example of the kind of institution that seeks to obtain a variety of employees.)

The production of an industry b firm involves the use of apprentices as well as experienced workers. We represent this by letting the firms employ the input "efficiency units of labour". λ_i is thereby defined as the total number of efficiency units of labour employed by industry b in a country i:

$$\lambda_i = \begin{cases} \bar{h} \cdot \bar{q}_i + h^* \cdot q_i^*, & \text{for } n_i = 1 \\ \bar{h} \cdot \bar{q}_i + \left(\sum_{k=1}^{n_i} \left(\frac{\hat{q}_i}{n_i} \hat{h}_k \right)^{\varepsilon} \right)^{\frac{1}{\varepsilon}}, & \text{for } n_i > 1 \end{cases} \tag{5}$$

with $0 < \varepsilon < 1$.

We assume that the value of experience is the same, regardless of in which firm it has been obtained, i.e. $\hat{h}_k = \hat{h} \quad \forall \; k$, which allows us to rewrite equation (5) as

$$\lambda_i = \begin{cases} \bar{h} \cdot \bar{q}_i + h^* \cdot q_i^*, & \text{for } n_i = 1 \\ \bar{h} \cdot \bar{q}_i + n_i^{\frac{1-\varepsilon}{\varepsilon}} \hat{h} \cdot \hat{q}_i, & \text{for } n_i > 1 \end{cases}. \tag{6}$$

λ_i is increasing and concave in the number of firms in the location i, reflecting the manufacturers' preferences for employees with divergent experiences: the marginal product of an experienced worker is higher the more diverse the knowledge of his experienced colleagues, i.e., the more diverse the labour stock, the higher the productivity of each worker. And the higher the geographical concentration of firms, the higher the chances of each firm to obtain a diversified labour stock.

The value of ε tells us the degree of economies of scope characterising the firms' knowledge capital. If ε were equal to 1, the miscellany of backgrounds is of no

importance to the firms. But if ε approaches 0, there is a maximum degree of economies of scope. Firms benefit substantially from having a varied workforce, and the proximity to several other firms is very important to a firm.

In general, we have that as long as $h^* < \hat{h}$, location matters in the sense that whether or not a firm has "neighbours" affects the productivity of its own labour force. — And provided that $0 < \varepsilon < 1$, the larger the industrial agglomeration the higher the productivity of the firms within the agglomeration.

Wages

In the constant returns to scale (CRS) sector, workers are paid a wage w^a. Since the trade of a-goods is assumed costless, this ensures that the wage rate in sector a and the price paid for a-goods are constant and the same across countries. Choosing units such that one unit of labour produces one unit of output gives

$$w_a = p_a = 1. \tag{7}$$

The labour market in the increasing returns to scale (IRS) sector is perfectly competitive. Workers are paid a wage equal to the value of their marginal product, which is exactly the wage that any other manufacturer in the sector would be willing to offer him. Given such a wage structure, it is obvious that at a critical tenure $\hat{\tau}$, all workers find it profitable to start looking for a new job (see figure 4). Perfect information in the labour market ensures that, 1) a worker does not have any incentive to change jobs before he has reached this critical tenure; 2) having reached the tenure

$\hat{\tau}$ he will always get a new job immediately, due to the employers' preferences for experienced employees.[22]

The workers' productivity evolution, the manufacturers' preferences, and the wage structure in industry *b*, generate an instantaneous intra-market turnover. However, since a second job change is assumed not to increase a person's marginal product (see the previous section), all workers only have the incentive to change employers in industry *b* once.

If we differentiate equation (6) with respect to the apprentices and experienced workers, and multiply by the price paid for one efficiency unit of labour, we obtain group and country specific wages. Letting the location specific price of one efficiency unit of labour be ω_i, the apprentices, i.e. the workers who have not reached the critical tenure, earn a wage

$$\overline{w}_i = \omega_i \cdot \overline{h}\,, \tag{8}$$

while the experienced workers who have not been able to change jobs, earn the wage

$$w_i^* = \omega_i \cdot h^*, \tag{9}$$

and the experienced workers in their second job are paid

---

[22] Perfect information is defined as both workers and manufacturers recognising the form of the functions $h(\tau)$ and $H(\tau)$. Knowing the time he has spent in the industry, a worker can tell what value he has to his current employer as well as his value to an alternative employer. The manufacturers are ex ante able to perceive an applicant's level of skills, and distinguish between those who have, and those who have not, reached the critical tenure $\hat{\tau}$.

$$\hat{w}_i = \omega_i \cdot n_i^{\frac{1-\varepsilon}{\varepsilon}} \cdot \hat{h}. \tag{10}$$

Our model relies on the assumption that, in equilibrium, individuals ex ante do not have any preferences for the work in one sector or the other, i.e., in steady state the expected total income over the life time of an employee in sector *a* equals that of an employee in sector *b*. Using (7)-(10), and assuming a zero discount rate, we derive the condition which ensures an ex ante indifference:

$$\int_0^{\bar{\tau}} e^{-\delta\tau} d\tau = \omega_i \left[\int_0^{\tau^*} \bar{h} e^{-\delta\tau} d\tau + \int_{\tau^*}^{\bar{\tau}} h^* e^{-\delta\tau} d\tau \right], \quad \text{for } n_i=1 \tag{11}$$

$$\int_0^{\bar{\tau}} e^{-\delta\tau} d\tau = \omega_i \left[\int_0^{\hat{\tau}} \bar{h} e^{-\delta\tau} d\tau + \int_{\hat{\tau}}^{\bar{\tau}} n_i^{\frac{1-\varepsilon}{\varepsilon}} \hat{h} e^{-\delta\tau} d\tau \right]. \quad \text{for } n_i>1 \tag{12}$$

From equations (11) and (12) it can be deduced, that having started to work in a particular industry, one will always prefer to stay there for the rest of one's life. The reluctance to move to another industry is generated by differences in wage structures across industries, from which follows that a change of industry implies a reduction in total life-time income.

3. The market for goods

All consumers, regardless of nationality, are assumed to share the same tastes. Their utility function is

$$U = C_a^{1-\theta} C_b^{\theta}, \qquad 0 < \theta < 1 \tag{13}$$

with C_a depicting the consumption goods produced in sector a, and C_b is the aggregated consumption of goods produced in industry b in both countries. (A description of both industries is given in the first paragraph, and will not be repeated here.) C_b is a constant-elasticity-of-substitution function of consumption of the differentiated goods, taking the form

$$C_b = \left(\sum_{k}^{n_1} c_k^{\left(\frac{\sigma-1}{\sigma}\right)}\right)^{\frac{\sigma}{\sigma-1}}, \tag{14}$$

with $\sigma > 1$, $n = n_1 + n_2$,

and may be thought of as a sub-utility function or alternatively as a quantity index. We assume the elasticity of substitution, σ, to be constant and independent of number of firms.

The number of varieties produced is large enough to make oligopolistic interactions negligible, i.e.,we are left with a constant elasticity of demand equal to σ. The large number of potential varieties and economies of scale in production imply that the production of each variety will only be undertaken by one firm, since a potential

entrant can always do better by introducing a new product variety than by sharing in the production of an existing product type.

International trade of goods produced in industry b involves trade costs. These take the Samuelson's "iceberg" form, so that in order to deliver one unit of a good from one country to another, $t \geq 1$ units must be shipped. In other words, if a good is exported, only $1/t$ of the good reaches its destination. If the countries are completely integrated, then $t = 1$, and trade costs equal zero.[23]

Dual to the quantity index we define location specific price indices:

$$P_1 = \left[n_1 p_1^{1-\sigma} + n_2 (p_2 t)^{1-\sigma} \right]^{\frac{1}{1-\sigma}} \tag{15}$$

$$P_2 = \left[n_1 (p_1 t)^{1-\sigma} + n_2 p_2^{1-\sigma} \right]^{\frac{1}{1-\sigma}}, \tag{16}$$

which are in fact nothing but the location specific unit expenditure functions for industry b products.

Due to the Dixit-Stiglitz framework we have employed, a two-stage budgeting procedure is valid. In stage one, maximising the Cobb-Douglas utility function allocates expenditure between sectors a and b, so that a fraction θ of the total income

[23] We define free trade as a situation with t=1, which is equal to zero trade costs. Reviewing the definition of "trade costs", these were defined as: "...a synthetic measure of a wide range of barriers to international trade..." According to our interpretation a situation with "zero trade costs", is a situation where the costs related to trade are independent of whether the trade is national or international. There are still costs associated with the trade of goods, but these are primarily transport costs determined by geographic distance. In order to include such costs, we would have to specify exactly where in a country industry b is located — how far from borders and cities. Such an extension of the model is, without doubt, of great relevance to the study of economic geography, trade and economic integration, but beyond the scope of this paper.

is spent on industry *b* products. In stage two, we derive the demand for a particular variety, which can be expressed as a function of the price indices. The first subscript refers to where the good is produced, the second to the location where it is consumed, and the third to the manufacturer. Expenditure at location *i* is denoted by E_i:

$$x_{iik} = p_i^{-\sigma} P_i^{\sigma-1} \theta E_i \tag{17}$$

$$x_{ijk} = t^{1-\sigma} p_i^{-\sigma} P_j^{\sigma-1} \theta E_j \tag{18}$$

Expenditure in a country is by assumption, determined purely by the labour income in the country. We can derive the expenditure in country i in steady state by using equations (7)-(10) and the information provided in table 1.

$$E_i = \begin{cases} L_{ai} + \overline{w}_i \cdot \overline{q}_i + w_i^* \cdot q_i^*, & \text{for } n_i = 1 \\ \\ L_{ai} + \overline{w}_i \cdot \overline{q}_i + \hat{w}_i \cdot \hat{q}_i, & \text{for } n_i > 1 \end{cases} . \tag{19}$$

We substitute for the expressions for wages, solve equations (11) and (12) with respect to ω_i, substitute further, and find that the expenditure in an economy in steady state is, moreover, equal to the economy's total labour force:

$$E_i = L_i \tag{20}$$

We remember that production was assumed subject to economies of scale. This is reflected by the linear production function below, which takes a form typical of the trade and geography models:

$$\lambda_{ik} = \alpha + \beta x_{ik} \tag{21}$$

The production of each variety k requires α efficiency units of labour as a fixed cost and β per unit output thereafter. The cost function dual to the production function can easily be derived, and the cost function of a firm located in country i is defined as

$$C(\omega_i, x_{ik}) = \omega_i \cdot (\alpha + \beta x_{ik}) \tag{22}$$

with $x_{ik} \equiv x_{iik} + x_{ijk}$, and ω_i defined as the price of one efficiency unit of labour.

The profits of a representative firm in country i take the form:

$$\pi_{ik} = p_i \cdot (x_{iik} + x_{ijk}) - C(\omega_i, x_{ik}) \tag{23}$$

The profit maximisation condition of a firm located in country i, is given by the equality of marginal revenue and marginal cost:

$$p_i = \left(\frac{\sigma}{\sigma - 1}\right) \cdot \beta \omega_i. \tag{24}$$

It follows that in equilibrium, where there is a constant number of firms in each country, the price is location specific and a constant mark-up over marginal cost.

If there are free entry and exit, profits will be driven to zero. The zero profits condition together with the condition for profit maximisation (equation (24)) give us the equilibrium output level:

$$x_{ik} = x = \frac{\alpha(\sigma-1)}{\beta} \tag{25}$$

As it appears, the equilibrium output level is independent of price, level of costs, market size and number of firms. Hence, the size of the firm determined by imposing the free entry and exit condition, is the same for all firms in the industry, and independent of their location.

Substituting for the equilibrium output in the production function, we obtain total demand for efficiency units of labour in country *i* in equilibrium:

$$\lambda_i = n_i \alpha \sigma \tag{26}$$

4. Solving the model

We specify the equations for country 1 necessary to solve for the equilibrium, and just note that the equations for country 2 are symmetric. Without loss of generality, and in order to save notation, we choose $\alpha = 1/\sigma$ and $\beta = (\sigma-1)/\sigma$, so that $x = 1$ and $p_i = \omega_i$.[24]

Product market equilibrium requires that the total demand for a single variety must equal the units produced. The equilibrium condition is derived using equations (15)-(18), (20) and (25):

[24] α, β and σ are all constants and equal across locations. As long as there is only one IRS industry and no inter-industry comparisons are undertaken, suppressing these constant terms does not affect our results.

$$1 = \omega_1^{-\sigma} \cdot \theta\left[\left\{n_1\omega_1^{1-\sigma} + n_2(\omega_2 t)^{1-\sigma}\right\}^{-1} L_1 + \left\{n_1(\omega_1 t)^{1-\sigma} + n_2\omega_2^{1-\sigma}\right\}^{-1} t^{1-\sigma} L_2\right] \quad (27)$$

Labour market equilibrium is characterised by the workers ex ante having no preferences for the work in one sector or the other, and the labour stock being fully employed. Integrating and rewriting the equations (11) and (12), we have that

$$\omega_1 = \begin{cases} \left[\bar{h} \cdot \left(1 - e^{-\delta\tau^*}\right) + h^* \cdot e^{-\delta\tau^*}\right]^{-1}, & \text{for } n_1 = 1 \\ \left[\bar{h} \cdot \left(1 - e^{-\delta\tilde{\tau}}\right) + \hat{h} \cdot n_1^{\frac{1-\varepsilon}{\varepsilon}} \cdot e^{-\delta\tilde{\tau}}\right]^{-1}, & \text{for } n_1 > 1 \end{cases} \quad (28)$$

From equation (28) it can easily be derived that the price per efficiency unit of labour is decreasing and convex in the number of firms in a location. Labour market clearing additionally requires that the sum of demand for labour equals the total supply of labour in the economy (see equation (2)). Using equations (6), (26) and (28) the equilibrium labour demand in sector *b* can be expressed as

$$L_{b1} = \omega_1 n_1, \quad (29)$$

and we may rewrite the labour market clearing condition (equation (2)) using equation (29) to substitute, as

$$L_{a1} = L_1 - \omega_1 n_1. \quad (30)$$

In order to allow for complete specialisation we ultimately add the assumption that,

$$L_i \geq \theta \cdot (L_1 + L_2), \quad \text{for } i=1,2, \tag{31}$$

to ensure that each country is large enough to be able to "host" the complete industry *b*. The four equations (27), (28), (29) and (30) together with the symmetric equations for country 2 simultaneously solve for the equilibrium n_1, n_2, ω_1, ω_2, L_{b1}, L_{b2}, L_{a1} and L_{a2}. Through substitution, we can subsequently derive equilibrium wages and prices in sector *b* in both countries, and the equilibrium number of apprentices and experienced workers in both countries.

5. Equilibrium configurations

We shall consider three different cases: First, the situation where trade costs are infinitely high so that trade is completely impeded, in other words, the autarky case. Next, we focus on the possibility of the two countries becoming fully economically integrated, and trade costs diminishing to zero. Finally, a situation of partly integrated countries and intermediate trade costs is considered.

(i) Autarky

Proposition 1:

Under autarky ($t \to \infty$) there is a unique diversified equilibrium. The share of the labour force engaged in each industry is the same across locations. A difference in country size will be magnified by a productivity difference, so that the relative number of firms in industry b does not equate the relative size of the countries.

In other words, if country 1 is twice as big as country 2, then there are more than twice as many *b*-firms in country 1 as in country 2.

Proof: From the conditions for the product market equilibrium, we get that $n_1 = \theta L_1/\omega_1$ and $n_2 = \theta L_2/\omega_2$. Substituting for the number of firms in equation (29) and the symmetric equation for country 2, we obtain the equilibrium labour demand in industry *b*: $L_{b1} = \theta L_1$, $L_{b2} = \theta L_2 \Rightarrow L_{b1}/L_{b2} = L_1/L_2$, which reflects that consumers spend the same share of their income on *b* products invariant of their country of origin. Assuming that the size of each country is large enough to support more than one firm, and differentiating equation (28) with respect to the number of firms in the economy to find that $\partial\omega_1/\partial n_1 < 0$, it is clear that $L_1 \neq L_2$ implies $(|L_1 - L_2|/L) < (|n_1 - n_2|/n)$ in equilibrium.

The autarky result does not only hold for infinitely high trade costs, but for relatively high trade costs in general. Characterising the result, is that the location of firms is primarily determined by demand considerations and not by cost considerations.

(ii) Complete integration

Under the assumption of complete economic integration and zero trade costs (i.e. $t = 1$) most of the economic geography models developed within the new trade theory ("new" location theory), reach the following results: When there are zero trade costs, the forces driving the agglomeration of firms seize to exist, and location becomes irrelevant. The circular causation that drives the industrial agglomeration is no longer active. The location of firms is primarily determined by cost considerations, and firms

may, for instance, spread out in search of low-cost immobile factors, such as land.[25] These results apply to several kinds of external economies, but not to the case of technological spillovers (or knowledge spillovers in general): ***Regardless of the existence of trade costs, as long as there are spillovers, firms will ceteris paribus always gain from the proximity to other firms within the same industry. Other things equal, clustered firms will always be more competitive than isolated firms.***

Proposition 2:

A complete integration (t=1) implies three possible equilibria: Complete concentration of industry b in country 1, complete concentration in country 2, and equal division of firms between the two countries. The diversified equilibrium implies factor price equalisation, but unlike the two first equilibria, it is an unstable equilibrium: Through migrating, firms will be able to reduce their own costs as well as those of their new "neighbours".

Which country gets the core of industry *b* firms, will depend on the initial distribution of firms ("history"), or possibly on people's expectations about where the core is going to be. If the industry *b* firms are initially unequally distributed between the two countries, concentration is self-reinforcing, and the country that — for some reason — gets a head start will end up with the complete industry *b*. Formally expressed: If $n_1 > n_2$ initially, then from (28) it is obvious, that producing in economy 1 implies a lower average cost than producing in economy 2, $\omega_1 < \omega_2$. Firms located in country 2 will always find migrating to the other location profitable.

[25] See for instance Krugman (1991b), Venables (1993), Baldwin (1994), Amiti (1995).

But if the economies start out with an equal division of industry b, then there is a significant scope for self-fulfilling prophecy. Independent of the relative size of the countries, this is the only division of production between countries which constitutes an equilibrium. A diversified equilibrium at $t = 1$ requires that $\omega_1 = \omega_2$ (see equation (27) and the symmetric equation for country 2), and it follows from equation (28), that this condition can only be satisfied for $n_1 = n_2$. But although $n_1 = n_2$ is an equilibrium, it is unstable. Unstable — because it is always profitable for any firm to shift its production to the other country. From equation (28) it can be derived that the price the manufacturer would have to pay per efficiency unit of labour in location 2 will, in the short run, never exceed, and in the long run (steady state), always be lower than, the price he is currently paying in location 1.

The question is but, in which direction do the firms move? The outcome will depend on factors not incorporated in our model. The more important the economies of scope in the accumulation of knowledge capital (in other words: the stronger the external economies), and the more rapidly firms move, the higher the chances of self-fulfilling expectations determining the outcome.

We observe that one country specialises in the production of a-goods, and one country specialises in the production of b-goods. In the appendix we show that, for $t = 1$ concentration of all manufacturing in industry b in one location is always a locally stable equilibrium. While the production of b-goods ends up completely concentrated in one country, depending on consumers' preferences for a relative to b goods and the size of the countries' populations, we may still have production of a-goods in both countries.

Somewhat surprisingly, we note that the relative size of a country does *not* determine whether it gets the core. Decisive is but the relative number of *b*-firms located in the country, and this does not necessarily equate the relative size of the country. Yet, a country's size may indirectly affect whether it ends up with the high-skill industry: If the initial situation is similar to that of autarky, and preferences are stable over time, it is obvious that the larger country will also have the larger share of industry *b* initially, and — according to our model — therefore obtains the core of *b*-firms.

But the fact that the relative size of the industry (n_i/n_j) rather than the relative size of the country, determines a country's opportunities of getting the core, has an important consequence: Industrial policies supporting and protecting certain industries become very attractive to governments. Having succeeded in the establishing of an efficient cluster, the sustainability and growth potential of the cluster increase when borders become eliminated and the countries integrated.

(iii) Intermediate trade costs

In order to analyse a situation characterised by such levels of trade costs, it is useful to distinguish between different country size relations. We define

$$\eta \equiv L_1/L_2 \tag{32}$$

a) The two countries are fully symmetric, i.e. $1 < t < \infty$ and $\eta = 1$.
Under the assumption that the countries have the same size, we obtain results similar to those under free trade. There are multiple equilibria, in which the specialised

equilibria are stable. In the appendix we prove that an equal division of manufacturers between the two locations is an equilibrium. Analytically it is difficult to prove that this is the only possible diversified equilibrium. Exploring the model numerically, does, however, unveil that production in both countries may only take place if the manufacturers are initially equally divided between the two locations. (See figure 5 for an illustration of a numerical example.) But as in the case of zero trade costs, such a diversified equilibrium is unstable (see the appendix for proof).

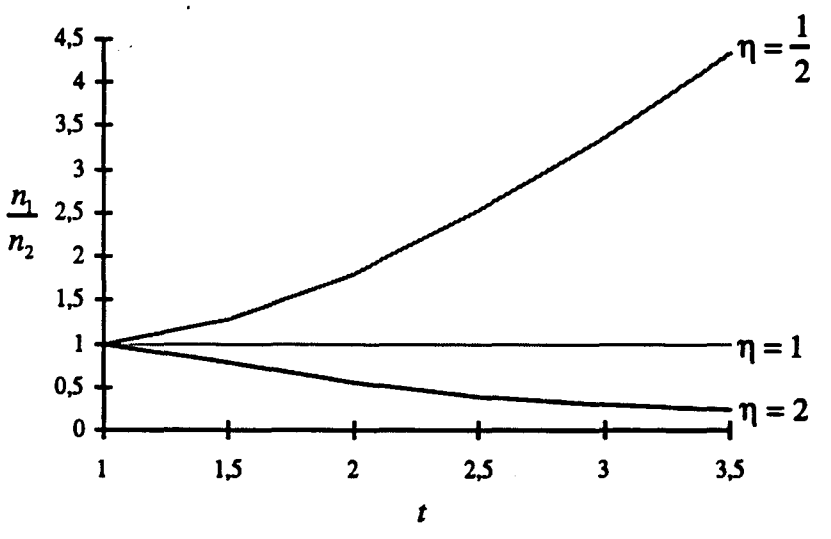


Figure 5

Figure 5 provides an illustration of the diversified, but unstable, equilibria which might occur for $1 < t < \infty$. We have chosen this form of presentation in order to emphasise the dependence of an equilibrium on trade costs. The central line along which $\eta = 1 \Leftrightarrow L_1 = L_2$ illustrates the fact that if the countries are of the same size, there is only one possible diversified equilibrium, namely $n_1 = n_2$.

b) There is one large and one small country, i.e. $1 < t < \infty$ and $\eta < 1$.
There are multiple equilibria, with two stable, specialised equilibria. If $n_1 < n_2$ initially, then from (28) it follows that the average cost of production in country 2 is lower than that taken into account by firms in the small country. The cost difference together with the market access effect, will induce a migration of *b*-firms from the small to the large country. Provided that the trade cost is below a critical level of t, there will be a unique specialised equilibrium where, influenced by costs as well as by demand considerations, all the manufacturers in industry *b* agglomerate in the large country.

If the firms are equally distributed across countries initially, $n_1 = n_2$, firms' average cost is the same across countries (see equation (28)), but the firms in the smaller country are placed at a disadvantage by their inferior market access. In the appendix, we show that an equal division of firms between the countries is never an equilibrium. Manufacturers located in country 1 have incentives to move to the larger country, and in equilibrium the complete industry is agglomerated in the large country.

Finally, we assume that due to for instance a government actively conducting trade policy, a larger share of the industry *b* is initially located in the smaller country ($n_1 > n_2$). These firms are faced with a lower average cost than their competitors in country 2. But the firms located in the big country possess a superior market access. If the number of firms in the small country is so large, that their higher productivity exactly compensates for the superior market access of their competitors, the initial distribution of firms across countries is an equilibrium. The upper line in figure 5 along which $\eta = \frac{1}{2} \Leftrightarrow L_1 < L_2$ illustrates the possible diversified equilibria in dependence of trade costs. But as above, the diversified equilibrium is an unstable

equilibrium (see the appendix for proof). Who gets the core is a question subject to indeterminacy.

Reviewing the results, it appears that, as in the case of complete integration ($t = 1$), the complete industry *b* will end up agglomerated in one location, and the big country will always get the core, unless $n_1 > n_2$. If there are more firms in the smaller country initially, the small country may actually get the core. But it will only obtain the core given that economies of scope in knowledge accumulation enhancing the external economies in industry *b* are very important, and trade costs are low.

One question remains to be answered: for what specific levels of trade costs does the result above apply. "Intermediate" trade costs are, without doubt, a vague formulation. However, it appears impossible to define an absolute level of trade costs, beyond which the results derived here do not apply, and a unique, stable diversified equilibrium emerges. The critical level of *t*, below which an asymmetric equilibrium is stable, depends on the relative size of the core market, the extent of internal economies of scale in production, and the degree of external economies. In the appendix we show that it is possible to define a level of *t* dependent upon these variables, below which a stable asymmetric equilibrium is always ensured.

6. Final Remarks

The existence of technological spillovers implies that location in the sense of proximity to other firms, always matters, regardless of the existence of trade costs. Applying a

general equilibrium model, a divergent development and geographic agglomeration of industries seem to be an inevitable result of economic integration.

The scope for policy actions is significant, because the *absolute* size of a country's industrial cluster prior to a (closer) economic integration is decisive for the country's chances of sustaining and enhancing the cluster after a completed integration. Our analysis provides strong arguments for the support of high-technology industries, as well as subsidies to research and development (R&D). Getting a head start within a new industrial field appears to be decisive for a country's prospects of sustaining and extending an industry.

We further note that the more trade costs are reduced as a consequence of integration, the better the prospects of a small country ending up with the core. But for a larger country the opposite is, in fact, true. The implications for the trade policy of both small and large countries follow immediately, and should not need any further explanation. Noteworthy is the fact, that the recommendations derived from the analysis above resemble those that follow from traditional trade theory based on a prevalence of perfect competition.

Although a consumer is always a worker and vice versa, in order to be able to say something about welfare implications, it is convenient to separate between consumers and workers. A complete economic integration, inducing firms to agglomerate, increases the number of varieties and reduces prices (see appendix for proof), and is, without doubt, beneficial to all consumers. As for situations with intermediate trade cost, the implications are more ambiguous. Consumers in the country with immigration of firms gain from the integration, whereas those in the "deserted"

country may either benefit or lose from the integration: The number of varieties increases, and prices are reduced due to higher productivity. On the other hand, however, consumer prices are increased due to trade costs.

From the workers' point of view, integration may be bad or good, depending on which country they inhabit, and whether or not they are born at the time when the integration is completed. In the long run, i.e., in a new steady state, all workers earn the same nominal life time income. Their real income, however, is dependent on whether or not the countries have become completely integrated. If trade costs equal zero after the integration, location is irrelevant for workers' welfare. But if integration reduces trade costs without eliminating them, the workers inhabiting the core market will experience a higher real income than those in the "deserted" country.

In the short run, there may, however, be severe adjustment problems. In the country which experiences an emigration of knowledge intensive firms, the workers left behind will have to seek employment in the other sector. Due to differences in wage structures across industries, their life time income will decrease compared to what they would have earned if they could have stayed in the same industry throughout their lives. Additionally, we have reason to believe that, before they get employed in the other sector, they may even have to cope with a period of unemployment.

The framework presented here suggests that integration leads to geographic consolidation of industries and increased international specialisation. Hence, the economic geography of regions that become integrated is likely to be severely altered. Losses as well as gains are certain outcomes of an integration process, and the scope for policy actions significant.

Appendix

Determinants of asymmetric, locally stable equilibria

An allocation where production in industry b is completely concentrated in, say, location 1, is locally stable only if it is unprofitable for any firm to deviate and shift its production to country 2. We let $n_2 = 0$, and derive the maximum price per efficiency unit of labour that a deviant firm can accept in such a situation without suffering losses from equation (27) and the symmetric equation for country 2:

$$\omega_2 = \left(\frac{\eta t^{1-\sigma} + t^{\sigma-1}}{\eta + 1} \right)^{\frac{1}{\sigma}} \omega_1. \tag{33}$$

In country 2, in a new steady state, the firm would have to pay a price $\omega_2 = \left[\bar{h} \cdot \left(1 - e^{-\delta \tau^*} \right) + h^* \cdot e^{-\delta \tau^*} \right]^{-1}$ per efficiency unit, while in the short run, the price that will have to be paid in order to induce workers to abandon sector a and start working in industry b, is even higher. Thus, an allocation with industry b agglomerated in economy 1 is a locally stable equilibrium if

$$\omega_2 < \left[\bar{h} \cdot \left(1 - e^{-\delta \tau^*} \right) + h^* \cdot e^{-\delta \tau^*} \right]^{-1}. \tag{34}$$

We may rewrite the condition above as

$$V \equiv \left(\frac{\eta t^{1-\sigma} + t^{\sigma-1}}{\eta + 1} \right) \cdot \left(\frac{\bar{h} \cdot \left(1 - e^{-\delta \tau^*} \right) + h^* \cdot e^{-\delta \tau^*}}{\bar{h} \cdot \left(1 - e^{-\delta \hat{\tau}} \right) + \hat{h} \cdot n_1^{\frac{1-\varepsilon}{\varepsilon}} \cdot e^{-\delta \hat{\tau}}} \right)^{\sigma} < 1. \tag{35}$$

From expression (35) it can be derived, that

$$\eta > \frac{1-\gamma^{\sigma}t^{\sigma-1}}{\gamma^{\sigma}t^{1-\sigma}-1} \tag{36}$$

with

$$\gamma \equiv \left(\frac{\overline{h}\cdot\left(1-e^{-\delta\tau^{*}}\right)+h^{*}\cdot e^{-\delta\tau^{*}}}{\overline{h}\cdot\left(1-e^{-\delta\hat{\tau}}\right)+\hat{h}\cdot n_{1}^{\frac{1-\epsilon}{\epsilon}}\cdot e^{-\delta\hat{\tau}}} \right) \tag{37}$$

is a sufficient condition for an agglomeration to be sustainable. A closer inspection of the expression (35) or (36) reveals that for $t=1$ an agglomeration is always sustainable invariant of its geographic location, while for $t\rightarrow\infty$ it is never sustainable. Provided that there is a positive number of inhabitants in each economy $0<\eta<\infty$,

$$t \leq \gamma^{-\left(\frac{\sigma}{\sigma-1}\right)} \tag{38}$$

ensures a locally stable core-periphery equilibrium invariant of the relative size of the core market. The more significant the economies of scope in knowledge accumulation, the more vital is the proximity to other firms, the smaller is γ, and the less likely is it that a firm will find it profitable to "defect". A reader familiar with the recent literature on economic geography and trade, will be aware that the elasticity of demand, σ, can also be interpreted as an inverse index of the importance of internal increasing returns, since one measure of economies of scale is the ratio of average cost to marginal cost, which in equilibrium is equal to $\sigma/(\sigma-1)$. From expression (38) it can then be deduced, that — in general — the less elastic the demand or the greater the internal economies of scale, the more sustainable is a core.

Differentiating V with respect to η, we find that

$$\frac{\partial V}{\partial \eta} = \gamma^{\sigma}\left(\frac{t^{1-\sigma} - t^{\sigma-1}}{(\eta+1)^2}\right) < 0. \tag{39}$$

i.e., the larger the relative size of the core market, the stronger the forces sustaining the core. A bit surprising perhaps, and unlike the results obtained in other economic geography models it appears that the share of income spent on industry *b* products, does not affect the forces for agglomeration.[26]

Turning to trade costs, these are found to have an ambiguous effect on the "survival" of the core,

$$\frac{\partial V}{\partial t}\begin{Bmatrix} > \\ = \\ < \end{Bmatrix} 0 \quad for \quad t\begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \eta^{\frac{1}{2(\sigma-1)}}, \tag{40}$$

and are illustrated for a numerical example in figure 6 below.

[26] See for instance Krugman (1991b).

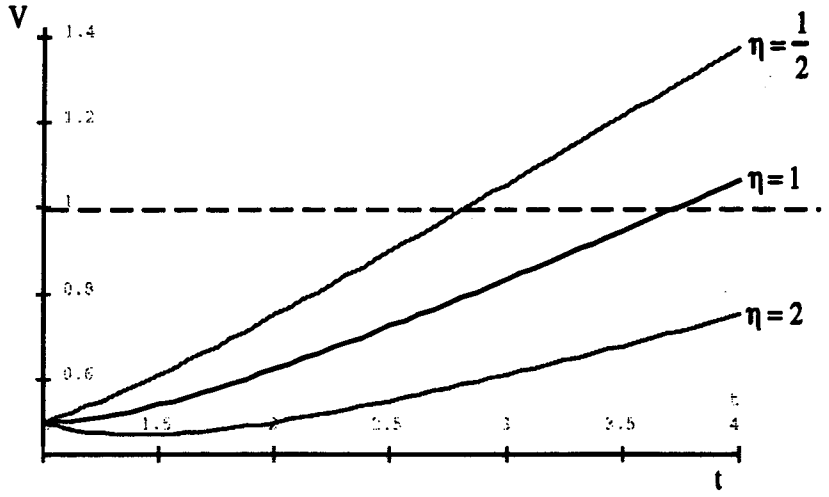


Figure 6

Equilibrium configurations for intermediate trade costs

Define

$$A \equiv \left(n_1 \omega_1^{1-\sigma} + n_2 (t\omega_2)^{1-\sigma}\right)^{-1}, \; B \equiv \left(n_1 (t\omega_1)^{1-\sigma} + n_2 \omega_2^{1-\sigma}\right)^{-1}. \tag{41}$$

Using (27) and the symmetric equation for country 2 it can be derived that for $t = 1$, $A = B$ always, and for $t > 1$

$$n_1 \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} n_2 \Leftrightarrow A \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} B \tag{42}$$

Employing equation (41) it is convenient to rewrite the product market equilibrium conditions as

$$\frac{A\eta t^{1-\sigma}+B}{A\eta+Bt^{1-\sigma}}=\left(\frac{\omega_2}{\omega_1}\right)^{\sigma}. \tag{43}$$

(a) Show that $n_1 = n_2$ is an equilibrium for $t > 1$, $\eta = 1$, though unstable.

$n_1 = n_2$ implies that $\omega_1 = \omega_2$ (see equation (28)) and $A = B$ (see equation (42)). Thus, equilibrium requires that

$$\frac{A\eta t^{1-\sigma}+A}{A\eta+At^{1-\sigma}}=1, \tag{44}$$

implying that $\eta = 1$, which is true.

The maximum price per efficiency unit of labour that a deviant firm migrating from, say, location 1 to 2 can afford to pay without suffering any losses, is

$$\omega_2=\left(\frac{A\eta t^{1-\sigma}+B}{A\eta+Bt^{1-\sigma}}\right)^{\frac{1}{\sigma}}\omega_1. \tag{45}$$

The price that he actually will have to pay in location 2 is $\omega_2=\left(\bar{h}\cdot\left(1-e^{-\delta\hat{\tau}}\right)+\hat{h}\cdot\left(n_2+1\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{\tau}}\right)^{-1}$, it follows that

$$K\equiv\left(\frac{A\eta t^{1-\sigma}+B}{A\eta+Bt^{1-\sigma}}\right)^{\frac{1}{\sigma}}\left(\frac{\bar{h}\cdot\left(1-e^{-\delta\hat{\tau}}\right)+\hat{h}\cdot\left(n_2+1\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{\tau}}}{\bar{h}\cdot\left(1-e^{-\delta\hat{\tau}}\right)+\hat{h}\cdot n_1^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{\tau}}}\right)<1 \tag{46}$$

must be true for the equilibrium to be stable. Using that the diversified equilibrium implies $n_1 = n_2$ and $A = B$, the condition (46) is not satisfied. Thus, we have proved that an equal division of firms between locations will never be a stable equilibrium.

(b) Show that $n_1 = n_2$ is never an equilibrium for $\eta \neq 1$, $t>1$.

$n_1 = n_2$ implies that $\omega_1 = \omega_2$ (see equation (28)) and $A = B$ (see equation (42)). Thus, equilibrium requires that

$$\frac{A\eta t^{1-\sigma} + A}{A\eta + At^{1-\sigma}} = 1, \tag{47}$$

implying that $\eta = 1$ or/and $t = 1$ which is a contradiction.

(c) Show that $n_1 > n_2$ cannot be rejected as an equilibrium for $\eta < 1$, $t>1$.

$n_1 > n_2$ implies that $\omega_1 < \omega_2$ (see equation (27)) and $A < B$ (see equation (42)).

$$A\eta t^{1-\sigma} + B > A\eta + Bt^{1-\sigma} \tag{48}$$

is obviously a necessary, though not a sufficient, condition for $n_1 > n_2$ to be an equilibrium. Solving the inequality, we get that $A\eta < B$, which is always true, i.e. the condition (48) is always satisfied.

(d) Show that $n_1>n_2$ is an unstable equilibrium for $\eta < 1$, $t>1$

Proceeding in the same way as we did under (a), we find that

$$K \equiv \left(\frac{A\eta t^{1-\sigma}+B}{A\eta+Bt^{1-\sigma}}\right)^{\frac{1}{\sigma}}\left(\frac{\bar{h}\cdot\left(1-e^{-\delta\hat{t}}\right)+\hat{h}\cdot\left(n_2+1\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{t}}}{\bar{h}\cdot\left(1-e^{-\delta\hat{t}}\right)+\hat{h}\cdot n_1^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{t}}}\right)<1 \tag{49}$$

must be true for firms not to move to location 2, while

$$K \equiv \left(\frac{A\eta t^{1-\sigma}+B}{A\eta+Bt^{1-\sigma}}\right)^{-\left(\frac{1}{\sigma}\right)}\left(\frac{\bar{h}\cdot\left(1-e^{-\delta\hat{t}}\right)+\hat{h}\cdot\left(n_1+1\right)^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{t}}}{\bar{h}\cdot\left(1-e^{-\delta\hat{t}}\right)+\hat{h}\cdot n_2^{\frac{1-\varepsilon}{\varepsilon}}\cdot e^{-\delta\hat{t}}}\right)<1 \tag{50}$$

must be true for firms not to move in the opposite direction from 2 to 1. Using equation (43) to substitute for the first part of the expressions for K, it is clear, that neither of the conditions (49) and (50) can be true.

Number of firms in a core

We define:

ω_1^c = the price per efficiency unit of labour if all b firms are concentrated in location 1,

ω_1^d = the price per efficiency unit of labour if firms are equally divided between the two locations.

Setting $n_2 = 0$ and solving the product market equilibrium equation (27), we find that

$$n_1 = \frac{\theta(L_1+L_2)}{\omega_1^c}, \tag{51}$$

always, regardless of the level of trade costs. Compared with the number of firms at a diversified equilibrium at $t = 1$,

$$n_1 = n_2 = \frac{\theta(L_1 + L_2)}{2\omega_1^d}, \text{ and } \omega_1^d = \omega_2^d \ , \tag{52}$$

it is obvious that the total number of firms is larger if all firms are clustered in one location. Why? Because the price of one efficiency unit of labour is decreasing in the number of firms in a location: $\omega_1^c < \omega_1^d$.

The price to the consumers is a constant mark up over the price of one efficiency unit of labour, and is accordingly lower if all firms are agglomerated geographically.

Exploring equations (27) and (28) reveals that in a situation without complete specialisation the total number of firms is always lower than it is if all firms in the industry were agglomerated.

Parameter values

The simulation of the figure 5 set $\bar{h} = 1$, $\hat{h} = 2$, $\delta = 0.05$, $\hat{\tau} = 10$, $\theta = 0.5$, $\varepsilon = 0.6$, $\sigma = 4$; while that of figure 6 set $\gamma = 0.7$, $\sigma = 2$.

References

Abdel-Rahman, Hesham. and Masahisa Fujita (1990): "Product variety, Marshallian externalities and city sizes", *Journal of Regional Science* 30 (1990): 165-183

Amiti, Mary (1995): "Economic integration, industrial structure and trade patterns", Centre for Economic Performance, London School of Economics, Discussion paper no. 225, London, 1995

Arthur, W. Brian (1990): "Positive feedbacks in the economy", *Scientific American* 262 (February): 92-99

Baldwin, Richard E. (1994): *Towards an integrated Europe*, London: CEPR, 1994

Blanchard, Olivier Jean and Stanley Fischer (1989): *Lectures on macroeconomics*, Cambridge, Massachusetts: The MIT Press, 1989

Caballero, Ricardo J. and Richard K. Lyons (1990): "Internal versus external economies in European industry", *European Economic Review* 34 (1990): 805-830

CEPR (1992): *Is Bigger Better? The economics of the EC enlargement*, London:: CEPR, 1992

Dicken, Peter and Peter E. Lloyd (1990): *Location in space. Theoretical perspectives in economic geography*, 3rd ed., New York: HarperCollins Publishers, 1990

Dixit, Avinash K. and Joseph E. Stiglitz (1977): "Monopolistic competition and optimum product diversity", *American Economic Review* 67 (1977): 297-308

Glaeser, Edward L., Hedi D. Kallal, José A. Scheinkman, and Andrei Schleifer (1992): "Growth in cities", *Journal of Political Economy* 100 (1992): 1126-1152

Herriot, Peter and Carole Pemberton (1995): *Competitive advantage through diversity: Organizational learning from difference*, London: Sage Publications Ltd., 1995

Jaffe, Adam B., Manuel Trajtenberg and Rebecca Henderson (1993): "Geographic localization of knowledge spillovers as evidenced by patents citations", *Quarterly Journal of Economics* 108 (1993): 577-598

Kostiuk, Peter F. and Dean A. Follmann (1989): "Learning curves, personal characteristics, and job performance", *Journal of Labor Economics* 7 (1989): 129-146

Krugman, Paul R. and Anthony J. Venables (1990): "Integration and the competitiveness of the peripheral industry", in Bliss, Christopher J. and Jorge Braga de Macedo (eds.): *Unity with diversity in the European economy: the Community's southern frontier*, Cambridge: Cambridge University Press, 1990

Krugman, Paul R. (1991a): "Increasing returns and economic geography", *Journal of Political Economy* 94 (1991): 483-499

Krugman, Paul R. (1991b): *Geography and Trade*, Cambridge, Massachusetts: MIT Press, 1991

Krugman, Paul R. and Anthony J. Venables (1993): "Integration, specialization and adjustment", NBER Working Paper no. 4559, Cambridge, Massachusetts, 1993

Krugman, Paul R. and Maurice Obstfeld (1994): *International economics: Theory and policy*, 3rd ed., New York: HarperCollins College Publisher, 1994

Marshall, Alfred (1920): *Principles of Economics*, London: Macmillan, 1920

Martin, Philippe J. and Carol Ann Rogers (1994): "Industrial location and public infrastructure", CEPR Discussion Paper no. 909, London, 1994

Norman, Victor D. (1972): *Education, learning and productivity*, Dissertation, MIT, Cambridge, Massachusetts, 1972

Pissarides, Christopher A. (1985): "Short-run equilibrium dynamics of unemployment, vacancies, and real wages", *American Economic Review* 75 (1985): 676-690

Pissarides, Christopher A. (1992): "Search theory at twenty-one", Centre for Economic Performance, London School of Economics, Discussion paper no. 90, London, 1992

Pissarides, Christopher A. (1994):" Search unemployment with on-the-job search", *The Review of Economic Studies* 61 (1994): 457-476

Rotemberg, Julio J. and Garth Saloner (1990): "Competition and human capital accumulation: A theory of interregional specialization and trade", NBER Working Paper no. 3228, Cambridge, Massachusetts, 1990

Thøgersen, Øystein (1994): "Fiscal policy, structural adjustment and intergenerational welfare", Norwegian School of Economics and Business Administration, Institute of Economics Discussion paper no. 22/94, Bergen, 1994

Topel, Robert H. and Michael P. Ward (1992): "Job mobility and the careers of young men", *Quarterly Journal of Economics* 107 (1992): 438-479

Venables, Anthony J. (1993): "Equilibrium locations of vertically linked industries", CEPR Discussion Paper no. 802, London, 1993

Essay 4

International trade, technological development and agglomeration

by

Karen Helene Midelfart Knarvik and Jostein Tvedt*

Introduction

International specialisation in the classical Heckscher-Ohlin setting is determined by relative differences in factor endowments given equal technology. Technology is, however, a dynamic phenomenon. New process innovations are constantly being made, and old technologies are replaced by new and improved ones. Yet, technological progress within a certain industry does not necessarily take place simultaneously world wide. It is often argued that the diffusion of technology takes less time in geographically smaller regions with a high density of production units in the industry in question. The improved access to new knowledge is reckoned as one of the main reasons why manufacturers tend to agglomerate and create clusters — a view that is strongly supported by empirical findings.[1]

In this paper we wish to focus on the hypothesis that the creation of clusters — which provide pools of knowledge and enable rapid transmission of information —

* We would like to thank Jan I. Haaland, Victor D. Norman and Anthony J. Venables for valuable comments.

[1] See e.g Porter (1990).

may cause uneven technological development internationally, and contributes to explaining regional specialisation. To do this, a dynamic setting is chosen, in which technological advantage depends on localised knowledge spillovers, and evolves over time. We study how uneven technological development internationally may influence national industrial structures, trade patterns, trade volumes and welfare. Furthermore, within the same framework, we elaborate on how different levels of integration affect international specialisation, localisation of industries as well as national and world welfare.

Rivera-Batiz and Romer (1991) show that the theoretical treatment of knowledge is decisive for how integration affects growth, and distinguish between two cases: one in which flows of knowledge can be separated from flows of goods, and one in which such a separation is impossible. One should note that our model clearly builds on an assumption of such a separation being possible.

Crucial to our story is, obviously, an assumption of localised external economies. Consistent with traditional endogenous growth theory, we let the degree of innovation in a given region (or cluster) be related to the number of manufacturers. Yet, for positive externalities to be generated, some kind of transmission mechanism is needed. There are different kinds of such mechanisms, which all constitute reasons for agglomeration in space.[2] The model presented in this paper relies on the assumption that transmission of knowledge is due to intra-industry labour mobility: when people change jobs, they bring with them knowledge and impulses from their previous employer that can be utilised in the production by their new employer.[3]

---

[2] See Marshall (1920) a.o.

[3] See Knarvik (1995) for a discussion of the importance of labour mobility for the sustainability of clusters.

Because workers regard geographic mobility as expensive or are impeded by immigration policies, they are mobile among firms if they are located within the same region, but not if they are geographically departed. Hence, there is transfer of knowledge within a cluster, but not within the industry globally.

A reader familiar with the different directions within trade theory, is likely to recognise, that the model we shall present in the proceeding sections contains ingredients from two different directions within trade theory: the so called "new" location theory, as well as the literature on uneven development, less developed countries (LDC) and leapfrogging. As for the former direction most papers occupied with industrial agglomeration (clusters) and trade generally assume some kind of deviation from the hypothesis of constant return to scale and free competition.[4] Here we keep these two assumptions that characterise traditional trade theory. The effect of industrial clusters may therefore be studied in a classical general equilibrium model.

Employing trade models of endogenous growth, the latter direction focuses on the consequences of trade liberalisation among countries with very different characteristics for trade patterns and welfare.[5] These models, and the one we present in the proceeding section have in common the existence of external economies, that causes internationally unequal development — but do not share the way the external economies are modelled.

[4] See e.g. Krugman and Venables (1993) and (1994).
[5] See e.g. Krugman (1981), Grossman and Helpman (1991), and Brezis, Krugman and Tsiddon (1993).

The paper is organised as follows. Section 1 sets out the formal model — a general equilibrium model where perfect competition prevails in all markets. In section 2 we consider the case of free trade in goods, but no factor mobility, while in section 3 we look at a stage of closer integration by allowing for trade in capital as well. Section 4 concludes.

1. The model

We introduce a model of international trade. Our world consists of two regions (countries), region h and region f. Both regions produce two goods, 1 and 2. Thus, the total production of the two goods is given by $x_1 = x_1^h + x_1^f$ and $x_2 = x_2^h + x_2^f$. Further, we assume that each good is produced by a technology exhibiting constant returns to scale in the two production factors n and k. We shall refer to n as labour and k as capital. The total factor input in country i for production of good j is given by $\{n_j^i, k_j^i\}$, where $i = h, f$ and $j = 1,2$. Total factor endowment in each country is given by $\{n^i, k^i\}$, and we assume that $\{n^i, k^i\}$ is fixed in time.

The assumption of constant returns to scale applies at the industry level, but not at the firm level. Each single firm is assumed to face U-shaped average costs, and chooses production quantity in order to minimise average costs. To simplify we let the minimum average cost be reached when a firm produces one unit. It follows that, in each region the total number of firms in an industry is equal to the total amount of output.

We assume perfect competition in goods and factor markets, and free trade in goods. Internally in both regions at any time t, there is costless mobility of both labour and capital between the two industries. Workers are but internationally immobile, due to immigration policies, moving costs, or cultural and emotional ties to their origins.

We let production of good 1 be intensive in the use of labour relative to the production of good 2. We assume that the production technology used in industry 1 is constant in time and equal across locations. The technology in industry 2 may however, change over time. The production in country i at time t of good 2 is given by

$$x_{2t}^{i} = \Psi_{t}^{i} H\left(n_{2t}^{i}, k_{2t}^{i}\right) \tag{1}$$

where Ψ_{t}^{i} is a location specific productivity scalar process.

The development of the productivity scalar is influenced by two factors. We assume that each production unit has a constant innovation trend, ϕ. That is, ϕ represents an incumbent unique improvement to the production technology of each of the production units. However, through intra-industry labour mobility this unit specific knowledge will gradually be shared by the other production units in the region. The stock of not yet shared knowledge in region i at time t is given by I_{t}^{i}, and the diffusion from the stock of knowledge is denoted by μ_{t}^{i}. The increment of the productivity scalar in sector 2 at time t will then be

$$d\Psi_{t}^{i} = \left(\phi + \mu_{t}^{i} I_{t}^{i}\right) dt. \tag{2}$$

That is, the productivity of each production unit in sector 2 increases due to own innovation, ϕ, but also by learning about innovations made by competitors located within the same cluster through new employees, i.e. by $\mu_t^i I_t^i$.

Since the number of production units is equal to aggregated production volume, total instantaneous innovation in the industry be given by $\phi x_{2_t}^i$. From the stock we deduct the shared knowledge at time t, $\mu_t^i I_t^i$. Accordingly, the increment of the stock of unshared knowledge is given by

$$dI_t^i = \left(\phi x_{2_t}^i - \mu_t^i I_t^i\right)dt \tag{3}$$

Innovation in each firm, ϕ, may differ between the two regions, and may depend on the level of education, the entrepreneurial and innovative spirit of the workers, and may vary through the life cycle of the industry. The degree of innovation in a young industry is usually different from that of an old industry, and hence it may be reasonable to let ϕ be time dependent. To simplify, though, we shall assume ϕ to be a constant.

The diffusion variable μ_t^i depicts the diffusion of innovations, i.e., the knowledge spillovers, that is generated by intra-industry labour mobility. Knowledge is assumed sector specific, so that inter-industry labour mobility does not add to the diffusion of knowledge, and is restricted to the firms within a region, because workers do not to move internationally.

Each job-worker pair is faced with an exogenous separation probability due to random shocks to the production units in both sectors.[6] The exogenous separation process causes continuous labour turnover, ensuring the spread of sector specific knowledge.

Agents are assumed so small, and so many, that they perceive the technological development as exogenous, i.e., to be independent of their own actions. Thus, the agents do not act strategically to enhance external economies of scale. The model is based on an assumption of instantaneous adjustment to equilibrium, which implies that all workers in a region earn the same wage regardless of their industry of employment. Hence, unemployed workers are indifferent between working in the two sectors, and accept the first job offer they get, regardless from which industry it comes.

If wages are perceived as equal across sectors and pay is the only criterion for choosing between job offers, then the probability that an industry 2 worker in search of a new job will end up in the same industry , is given by n_2^i/n^i. Accordingly, we have a simple expression for the diffusion rate of knowledge:

$$\mu_t^i = \lambda^i \frac{n_{2t}^i}{n^i}, \tag{4}$$

where λ^i represents the exogenous separation probability per time unit. It may also be thought of as a constant representing the degree of mobility of the labour force in

[6] A reader familiar with labour market theory will recognise the assumption of an exogenous separation rate, which is employed in most models within the search theory and matching literature (see e.g. Pissarides (1985)).

equilibrium. There may be differences in λ^i among regions, due to for instance differences in geographical size or cultural differences.

According to our model, both the absolute and the relative size of the industry influence productivity. An absolutely large industry means that a "lot of talent" is occupied within this industry, which may lead to a high number of improvements to the sector specific technology. In accordance with endogenous growth theory, we shall refer to this as the "growth effect". A relatively large industry implies that the innovations of one firm are rapidly shared by others in the industry, and shall be referred to as the "spillover effect".[7] Due to a high degree of diffusion of innovations, the productivity level may hence be higher in a region with a high density of firms of a given industry, even though this region has only a minor part of total world production of the industry and of overall innovation in the world.

Consistent with the assumption that k^i is fixed in time ($k_t^i = k^i \; \forall \; t$), we let there be no savings, and the preferences of the representative consumer in region i are reflected by an additive separable utility function

$$U_t^i = u\left(c_{1t}^i, c_{2t}^i\right), \tag{5}$$

where c_{jt}^i is consumption of good j in region i at time t. In order to remove scale effects, i.e., to let the marginal rate of substitution be unaffected by changes in income, we assume that preferences are homothetic.

[7] What we refer to as the "spillover effect" here, actually combines two of the types of externalities identified by Marshall as reasons for firms to agglomerate, namely (1) a pooled labour market for workers with specialised skills, as well as (2) information spillovers.

To maximise utility in a region i then reduces to the maximisation of $u(c_{1t}^i, c_{2t}^i)$ given the consumer's budget constraint. The budget restriction

$$w_t^i n^i + r_t^i k^i = p_{1t} c_{1t}^i + p_{2t} c_{2t}^i \tag{6}$$

implies that the total regional labour and capital income must be equal to total regional consumption expenditure where w_t^i is the wage rate, r_t^i is the rate of return on capital, and p_{jt} is the price of good j, all at time t.

2. Free trade in goods

We shall first consider the case where there is free trade in goods, but no factor mobility. In order to simplify the discussion of the dynamics of the model, we restrict ourselves to Cobb-Douglas production and utility functions, and suppress the subscript indicating time.

Let the production function in industry 1 be given by

$$G(n_1^i, k_1^i) = {n_1^i}^{\alpha} {k_1^i}^{1-\alpha}, \tag{7}$$

where $0 < \alpha < 1$, and let the production function in industry 2 be given by

$$\Psi^i H(n_2^i, k_2^i) = \Psi^i {n_2^i}^{\varepsilon} {k_2^i}^{1-\varepsilon} \tag{8}$$

with $0 < \varepsilon < 1$. We have assumed that industry 1 is intensive in the use of labour, from which follows, that $\alpha > \varepsilon$. Adding the assumption that λ^i is equal across locations, i.e. $\lambda^i = \lambda \; \forall \; i = h, f$, the external economies of scale prevailing in sector 2 take the same form across countries.

The utility of the consumer in region i at time t is expressed as

$$u\left(c_1^i, c_2^i\right) = c_1^{i\gamma} c_2^{i1-\gamma}, \tag{9}$$

where $0 < \gamma < 1$, i.e., the agents have identical preferences across regions.

Factor market equilibrium requires that the sum of demands for each factor equals the supply of each factor. With no factor mobility, the equilibrium conditions are

$$k^i = \frac{\partial b_1^i}{\partial r^i} x_1^i + \frac{\partial b_2^i}{\partial r^i} x_2^i \tag{10}$$

and

$$n^i = \frac{\partial b_1^i}{\partial w^i} x_1^i + \frac{\partial b_2^i}{\partial w^i} x_2^i. \tag{11}$$

Let b_j^i depict the unit cost in industry j in region i, and for industry 1 be given by

$$b_1^i\left(w^i, r^i\right) = r^i \left(\frac{w^i}{r^i}\right)^{\alpha} \left(\frac{1-\alpha}{\alpha}\right)^{\alpha} \left(\frac{1}{1-\alpha}\right), \tag{12}$$

and for industry 2 by

$$b_2^i(w^i,r^i)=\Psi^{i-1}r^i\left(\frac{w^i}{r^i}\right)^{\varepsilon}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1}{1-\varepsilon}\right). \tag{13}$$

Profit maximising in the product market implies that $b_1^i(w^i,r^i)=p_1$ and $b_2^i(w^i,r^i)=p_2$, while maximising utility gives the total world demand:

$$c_1^h+c_1^f=\frac{\gamma}{p_1}R \tag{14}$$

and

$$c_2^h+c_2^f=\frac{1-\gamma}{p_2}R \tag{15}$$

where R is the value of total production, $R=w^hn^h+r^hk^h+w^fn^f+r^fk^f$. Product market equilibrium requires that total world demand equals total world supply, i.e., $c_1^h+c_1^f=x_1^h+x_1^f$ and $c_2^h+c_2^f=x_2^h+x_2^f$.

We use good 1 as numeraire, and solve for the equilibrium characterised by diversified production in both countries initially. Employing the information provided above, in equilibrium all endogenous variables may be expressed in terms of parameters only. (The exact procedure followed in order to obtain the equilibrium values of the various variables is found in appendix B.)

The equilibrium price of good 2 is given by

$$p_2=\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\frac{(1-\alpha)}{(1-\varepsilon)}\left(A\frac{N}{K}\right)^{\alpha-\varepsilon}, \tag{16}$$

where $A=\left((1-\varepsilon)(1-\gamma)+(1-\alpha)\gamma\right)/\left(\varepsilon(1-\gamma)+\alpha\gamma\right)$. K is a weighted sum of total world capital, and N is a weighted sum of the total world labour supply. The weights depend on the productivity scalar of industry 2:

$$K=\left(\Psi^h\right)^{\frac{\alpha}{\alpha-\varepsilon}}k^h+\left(\Psi^f\right)^{\frac{\alpha}{\alpha-\varepsilon}}k^f \tag{17}$$

and

$$N=\left(\Psi^h\right)^{\frac{\alpha-1}{\alpha-\varepsilon}}n^h+\left(\Psi^f\right)^{\frac{\alpha-1}{\alpha-\varepsilon}}n^f. \tag{18}$$

Turning to factor prices, equilibrium returns to capital are given by

$$r^i=\Psi^{i\frac{\alpha}{\alpha-\varepsilon}}\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(A\frac{N}{K}\right)^{\alpha}, \tag{19}$$

and equilibrium wage rate is

$$w^i=\Psi^{i\frac{\alpha-1}{\alpha-\varepsilon}}\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(A^{-1}\frac{K}{N}\right)^{1-\alpha}. \tag{20}$$

In equilibrium the production volumes of the two goods are equal to

$$x_1^i=\frac{-\varepsilon r^i k^i+(1-\varepsilon)w^i n^i}{(\alpha-\varepsilon)} \tag{21}$$

and

$$x_2^i=\frac{\alpha r^i k^i-(1-\alpha)w^i n^i}{(\alpha-\varepsilon)p_2}. \tag{22}$$

In the following we seek to elucidate how the dynamic development of technological advantage determines industrial structure and trade patterns. At time

zero, technologies are assumed equal across regions, and industry structure and trade are determined by factor endowments, $\{n^i, k^i\}$. Over time, the technology in sector 2 is but subject to region specific development, and governed by two forces. A large absolute size of the industry in question means that the total number of innovations in a region is high, while a large relative size influences how fast innovations are shared by competitors in a region. Both forces, the growth as well as the spillover effect, enhance the competitiveness of the manufacturers in a cluster.

It is possible to distinguish between three different cases with regard to how trade patterns evolve over time. In the first two cases the region that initially possesses a comparative advantage in the production of good 2, due to relative factor endowments, reinforces its position in industry 2 over time. In the third case, there is a reversion of trade patterns over time, because the region that initially possessed a relative disadvantage in the industry 2 production, gradually gains a leading position in the industry, due to the technological development.

(i) A sustainable cluster

We shall first consider the cases where relative factor endowments are predictive of which region that ends up being the net exporter of what product. Assume that all firms, regardless of initial location, employ the same technology, so that $\Psi_0^h = \Psi_0^f = 1$, and let region h be relatively abundantly endowed with capital. Hence, region h has initially a comparative advantage in the production of good 2, which is relatively intensive in the use of capital. At time zero, region h is a net exporter of good 2, while region f is a net exporter of good 1. It follows that region h initially has a higher percentage of the work force employed in sector 2 than region f, and that the spillover effect is stronger in the former region.

In what region the growth effect is most significant will, however, depend on the international distribution of factor endowments. If region *h* is larger than, or equal to, region *f* in terms of labour supply, not only does region *h* experience a stronger spillover effect, but it experiences a stronger growth effect than region *f* as well. Accordingly, the outcome is unambiguous. The uneven technological development enhances the position of region *h* in industry 2 over time. We observe increased specialisation and a trend towards complete specialisation with industry 2 concentrated in region *h*.

Yet, even if region *h* is slightly smaller than region *f* in terms of labour as well as capital endowment, the same outcome as above will appear. In such a case region *f* actually experiences a more significant growth effect than region *h*, due to its larger size. Because of the stronger spillover effect enjoyed by region *h*, region *f* is, nevertheless, not able to challenge the position of region *h* in industry 2. How much smaller region *h* may be, without the outcome being altered, depends on the significance of the growth effect relative to the significance of the spillover effect, as well as on the share of income spent on sector 2 goods. The smaller this share, the smaller region *h* may be, without having to fear the vanishing of its industry 2 cluster.

To illustrate the dynamics in a situation where the technological development reinforces the trade patterns initialised by relative factor endowments, we employ a numerical example. (For parameter values and endowments, see the appendix A.) Figure 1 shows how trade develops over time. (Negative exports indicate imports.) The cluster of industry 2 in region *h* is steadily growing. Region *h* exports good 2 and imports good 1. The technological development magnifies the comparative

advantage that region *h* initially possessed in the production of good 2. There is increased specialisation and a trend towards complete specialisation.

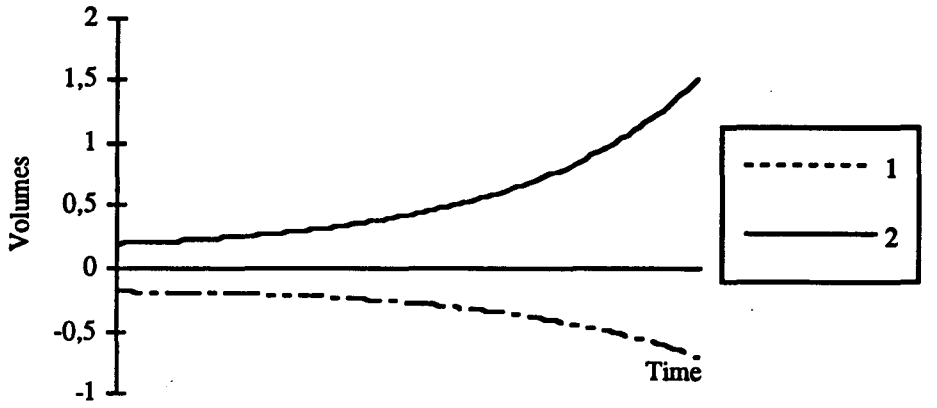


Figure 1: Export from region *h* — case (i)

As is shown in figure 2 below, an increasing part of the work force in region *h* becomes employed in sector 2, and because of the spillover effect, the cluster is thereby strengthened.

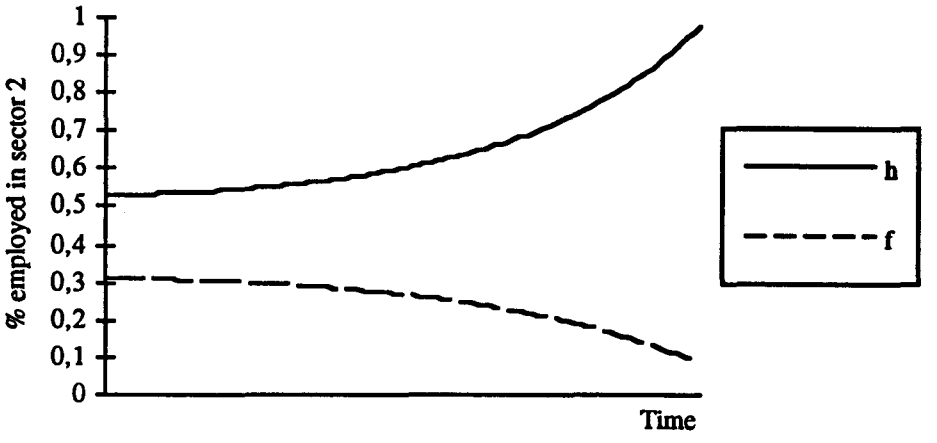


Figure 2: Percent employed in sector 2 — case (i)

Given free trade and equal technologies across regions, as is assumed to be the case at time zero, factor prices are initially equalised. As time passes by, region *h* enjoys a relatively advantageous technological development in the industry that is intensive in the use of capital. As a result, returns to capital increase and the wage rate decreases in region *h*, while in region *f* the effect is the opposite, returns to capital fall and the wage rate rises.

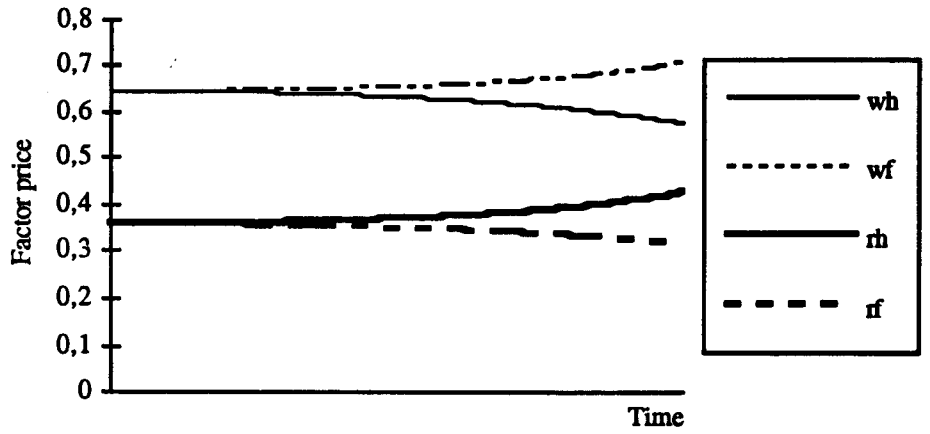


Figure 3: Factor prices — case (i)

The way industrial structure and trade patterns are affected by the uneven technological development, may be analysed within a framework analogue to the Rybczynski and Stolper-Samuelson theorems: one region experiences a relatively stronger technological progress in sector 2 than does the other. In this region production increases in the industry which is subject to technological progress, while the production in the other sector decreases. The price of the factor that is intensively used in the expanding sector, increases, while the price of the other

factor decreases. It follows that the unequal technological development across regions inhibits factor price equalisation internationally.

The technological progress is not only reflected by changes in factor prices, but also by a decrease in the price of good 2. Using per capita utility as a measure for welfare, we find that there is a welfare increase in both regions. The world as a whole, and each nation, gain from the technological advancement. The benefit is, however, more substantial in the region where the industry characterised by external economies gradually becomes concentrated: the uneven technological development causes rising differences in welfare across regions (see figure 4).

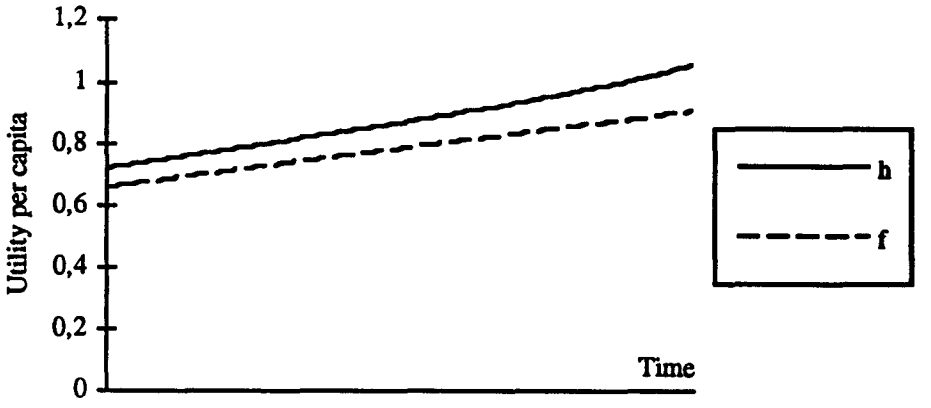


Figure 4: Utility per capita - case (i)

(ii) A non-sustainable cluster

Next, we consider the case where region *h* is still assumed to be relatively abundantly endowed with capital, but where this region is significantly smaller than the other region in terms of capital as well as labour endowments. In such a case, the

growth effect experienced by region *f* is so strong, that although region *h* initially enjoys a stronger spillover effect, its position in the industry 2 will be challenged. Consequently, there is a reversion of trade patterns over time. Region *h* gradually turns into a net importer of good 2, and there is a trend towards complete specialisation, with all industry 2 production concentrated in region *f*.

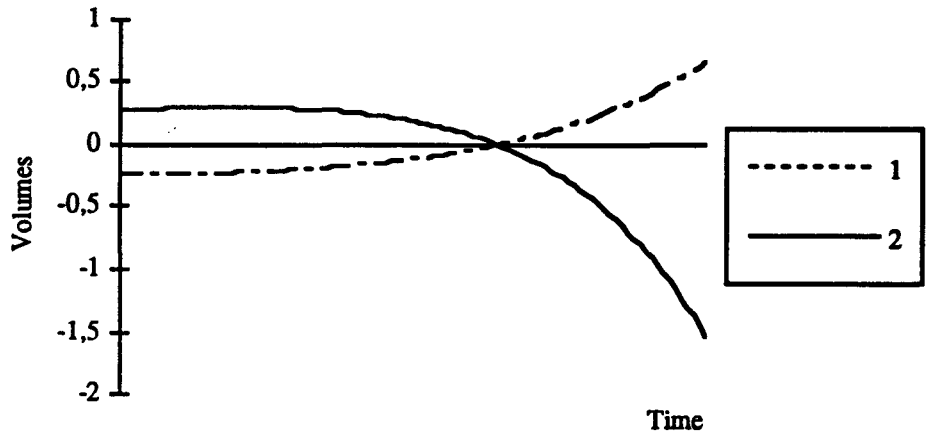


Figure 5: Export from region *h* — case (ii)

Figure 6 shows the development of sector 2 as an employer. Due to the original comparative advantage in capital intensive production, region *h* is initially net exporter of good 2. However, since the region is too small to keep pace with the technological development of the large region *f*, sector 2 eventually disappears from region *h*.

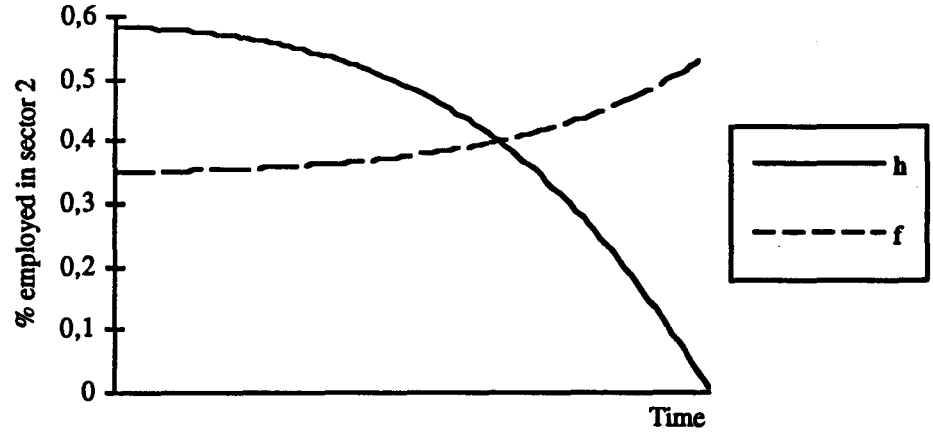


Figure 6: Percent employed in sector 2 — case (ii)

Factor prices evolve similarly to the way they did in case *(i)*, but with opposite signs (see figure 7). The relatively adverse development of the production technology in region *h* implies that the wage rate rises and the returns to capital fall, as the labour intensive industry in this region is enlarged. The opposite is the case in region *f*. Just like above trade does not entail international factor price equalisation.

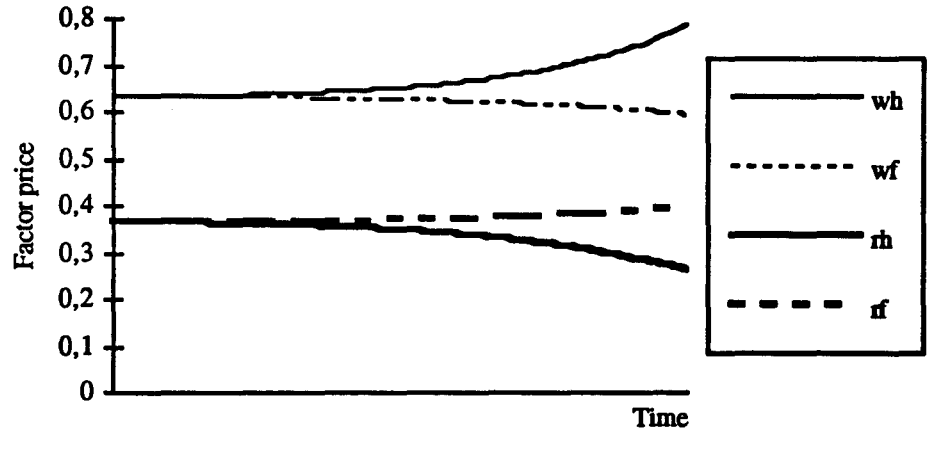
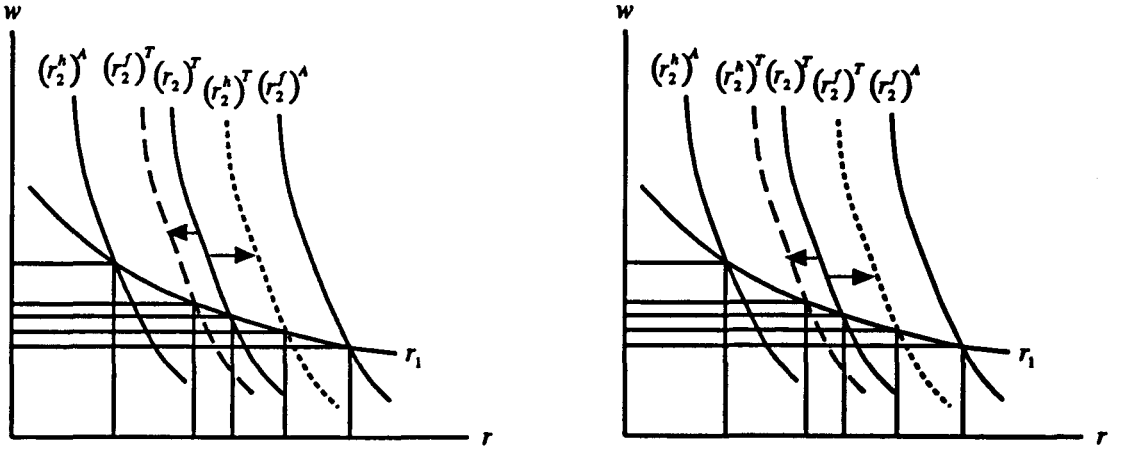


Figure 7: Factor prices — case (ii)

As in the former case, the technological progress caused by positive externalities entails a rise in world as well as in national welfare. During the first period, while region *h* is still a net exporter of good 2, the inhabitants of this region experience a more substantial increase in welfare than the inhabitants of the other region. This period will be followed by an interval characterised by convergence in welfare across regions. Eventually, convergence will be taken over by divergence. As long as there is technological progress, welfare will continue rising in both regions, but the inhabitants of the region that ends up hosting the cluster of industry 2, i.e., region *f*, will enjoy a more significant increase in welfare. The divergent economic development entails rising national disparities in welfare.

In order to understand the difference between cases (i) and (ii), it may be helpful to consider the two diagrams below, where returns to capital are drawn as a function of the wage rate. The superscript "A" denotes the autarky level, while "T" denotes free trade.



Case (i) Case (ii)

Figure 8

In both cases, region *h* initially obviously possesses a comparative advantage in the production of the capital intensive good 2. As we open for trade, factor prices become equalised, and are determined by the intersection of r_1 and $(r_2)^T$. Due to the unequal technological development, factor prices do, however, not stay equalised over time. The technological progress is reflected by: (1) increased returns to capital, inducing a shift to the right in the returns to capital curves, and (2) a decreasing price of good 2, inducing the curves to shift to the left. While the price effect stemming from the technological advancement is the same for all firms regardless of their location, the positive effect on returns to capital depends on the prevailing localised external economies in a country. In case (i) the external economies are relatively more significant in region *h* than in region *f*, while the opposite is true for case (ii). Accordingly, in case (i), the price effect dominates in the region *f*, but not in the region *h*. Until the industry 2 has possibly become concentrated in region *h*, returns to capital will steadily increase and the wage rate decrease in region *h*, while the opposite will happen in region *f*. In case (ii), however, the price effect comes to dominate in region *h* rather than in region *f*, so although trade liberalisation induces an increase in returns to capital originally, the technological development reverses this trend, and causes decreasing returns to capital.

Summarising the results obtained under the assumption of free trade and no factor mobility we find that external economies cause increased international specialisation and concentration of the industry characterised by externalities in one single region. Positive externalities lead to technological progress that benefits all individuals, but especially those living in the region that attains the industrial cluster.

3. Free trade in goods and capital

We have investigated the dynamic development of trade in the case of free trade in goods. The world today is characterised by what one may consider as a "boom in regionalism" [8] — set off by the EU. A number of countries seek to join existing blocs or to create their own. In order to take into account the fact that within these blocs, countries tend to become increasingly integrated, we shall move on to a situation where there is free trade not just in goods, but in capital as well. We do not include an assumption of international mobile labour, because experience has shown that cultural and emotional ties often impede international labour mobility. Note that capital income is assumed to be used for consumption in the region where the capital is initially localised.

If capital is allowed to move freely across borders, and there are no technological disparities between the two countries initially, then non-economic considerations such as firms' expectations about what location is to become the most attractive, will determine industrial structures and trade patterns. But if one country initially possesses a technological advantage, free trade in goods and capital produces three different outcomes with regard to international specialisation and economic geography:

- The country that initially has a technological advantage in production of the capital intensive good, is large enough in terms of labour supply, to produce profitably the total world demand for this good. The smaller the share of the world income spent on sector 2 goods, the smaller the labour stock needed for this country to be able to cover total world demand for these goods.

[8] See Krugman (1993).

- The country with the leading technology is too small to cover the world demand for the capital intensive good, and the production of this good is dispersed between the two countries.
- The country with the leading technology is too small to cover the world demand for the capital intensive good. Relatively to the other country, it is too small to sustain industry 2 production. Consequently, there is a reversion of trade patterns over time.

There may be various reasons for one region's initial technological lead. Initial differences in factor endowments, consumer preferences, trade policy, and economic policy in general may all cause uneven technological development, ensuring one region a head start as trade becomes liberalised. Uncertainty about the irreversibility of policy reforms influencing firms' choice of location, may also be responsible for initial differences in technology.

(iii) Complete concentration

Assume that due to relative capital abundance, region h is initially the technologically leading region, and that it is large enough in terms of labour supply to cover total world demand for the capital intensive good at any time. Allowing for trade in goods and capital, region h specialises in the production of good 2, and from the assumption of the regions' production capacity follows that $x_2^f = 0$, always.

International capital mobility ensures equal returns to capital across borders in equilibrium. Production of good 1 in both regions entails international factor price equalisation with respect to wages as well. Solving for the equilibrium with capital

mobility, we get the following equilibrium prices and production volumes (see the appendix B for the detailed procedure):

The price of good 2 is expressed by

$$p_2 = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\frac{(1-\alpha)}{(1-\varepsilon)}\frac{1}{\Psi^h}\left(A\frac{n}{k}\right)^{\alpha-\varepsilon} \tag{23}$$

with $n = n^h + n^f$ and $k = k^h + k^f$. The international wage rate and rate of returns to capital are given by

$$w = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(A^{-1}\frac{k}{n}\right)^{1-\alpha} \tag{24}$$

and

$$r = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(A\frac{n}{k}\right)^{\alpha}. \tag{25}$$

Since region h is the sole producer of good 2, the total production of this good is given by

$$x_2 = x_2^h = \frac{\alpha rk - (1-\alpha)wn}{(\alpha-\varepsilon)p_2}. \tag{26}$$

The production in region h of good 1 is

$$x_1^h = \frac{-\varepsilon rk + (1-\varepsilon)wn^h}{(\alpha-\varepsilon)} + \frac{\varepsilon(1-\alpha)wn^f}{\alpha(\alpha-\varepsilon)}, \tag{27}$$

while the production in region f is simply given by

$$x_1^f = \frac{wn^f}{\alpha}. \tag{28}$$

Since the total world consumption of the capital intensive good continues to be produced in the initially technologically leading region *h*, it follows that the other region is completely specialised in production of the labour intensive good. According to the model, the region that is completely specialised in the labour intensive good will not manage to improve its technology in sector 2. Hence, region *f* will not, at any stage, challenge the position of region *h* in the production of good 2.

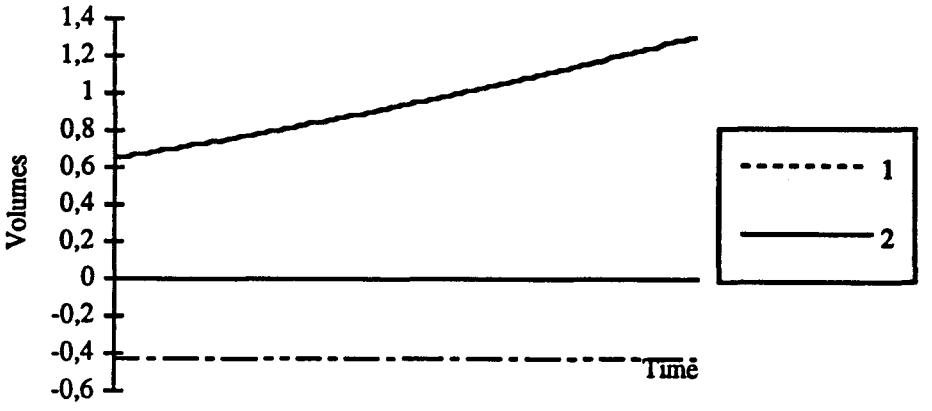


Figure 9: Export from region h — case (iii)

As shown in figure 9, production of good 2 will grow steadily in region *h*, due to the technological progress. As trade become liberalised, the distribution of labour across countries and the mobile capital allow for immediate concentration of the industry subject to external economies in one single country, at the same time as both

countries share in the production of the constant-returns good. The conditions for factor price equalisations are, in other words, satisfied.

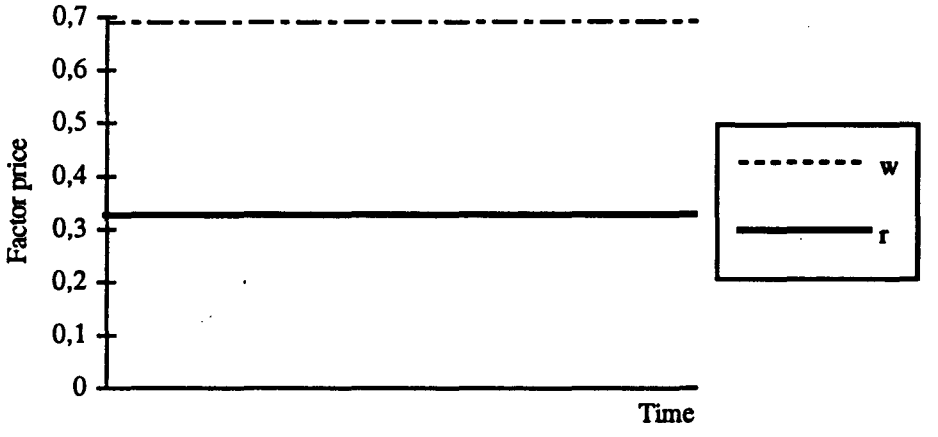


Figure 10: Factor prices — case (iii)

As illustrated in figure 10, factor prices stay constant over time, i.e., the technological progress leading to increased production volumes in sector 2 is purely reflected by a decreasing price of good 2. The world as well as the national welfare rise, but despite the factor price equalisation, location does not turn out to be irrelevant for the individuals' welfare. Yet, the disparities in national welfare that may be observed, are not caused by the technological development. The reason why some regions may experience a larger absolute increase in per capita utility, is due to differences in relative factor endowments. Since we have assumed that capital income is used for consumption by the individuals in the region where it was originally localised, those living in the relatively capital abundant region will enjoy a higher absolute increase in welfare due to trade liberalisation.

(iv) Dispersed production of good 2

We then turn to the case where region h, still assumed to be the region that is initially relatively capital abundant and technologically leading, is too small to profitably produce the total world demand for good 2. Reviewing case (ii), we find that with no capital mobility and insufficient production capacity to host the complete industry 2, there will be a gradual reversion of trade patterns. But with free capital mobility, region h may still be able to sustain an industry 2 cluster, despite its labour stock being insufficient to cover the total production of good 2.

Yet, from the assumption about labour endowments it follows that the production of good 1 will only come to take place in region f, i.e. $x_1^h = 0$. As in case (iii), free capital movements ensure international equalisation of returns to capital. But because both regions take part in the production of good 2, thereby employing unequal technologies, equilibrium wage rates come to differ across countries.

The equilibrium price of good 2 may be expressed as:

$$p_2 = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\frac{(1-\alpha)}{(1-\varepsilon)}\frac{1}{\Psi^f}A^{\alpha-\varepsilon}\left(n^h\left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}}+n^f\right)^{\alpha-\varepsilon}k^{\varepsilon-\alpha}. \qquad (29)$$

Equilibrium returns to capital are in both regions given by

$$r = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)A^{\alpha}\left(n^h\left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}}+n^f\right)^{\alpha}k^{-\alpha}. \qquad (30)$$

Further, we have that the equilibrium wage rate is

$$w^f = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)A^{\alpha-1}\left(n^h\left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}}+n^f\right)^{\alpha-1}k^{1-\alpha} \tag{31}$$

in region *f*, and

$$w^h = \left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}} w^f \tag{32}$$

in region *h*. Total production in region *h* is simply given by

$$x_2^h = \frac{w^h n^h}{\varepsilon p_2}. \tag{33}$$

Region *f* produces both goods, and in equilibrium the production volumes are

$$x_2^f = \frac{\alpha rk - (1-\alpha)w^f n^f}{(\alpha-\varepsilon)p_2} - \frac{\alpha(1-\varepsilon)w^h n^h}{\varepsilon(\alpha-\varepsilon)p_2} \tag{34}$$

$$x_1^f = \frac{(1-\varepsilon)\left(w^f n^f + w^h n^h\right) - \varepsilon rk}{(\alpha-\varepsilon)}. \tag{35}$$

Figure 11 shows the export from — and import to — region *h*. The region's production increases over time because of the technological progress. In a situation like this, the technologically leading region actually experiences the strongest possible spillover effects, because 100 percent of its labour force is engaged in the production of good 2.

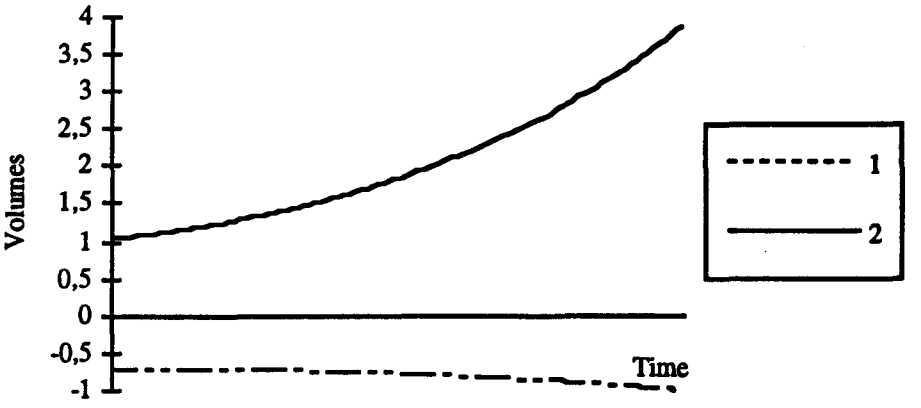


Figure 11: Export from region *h* — case (iv)

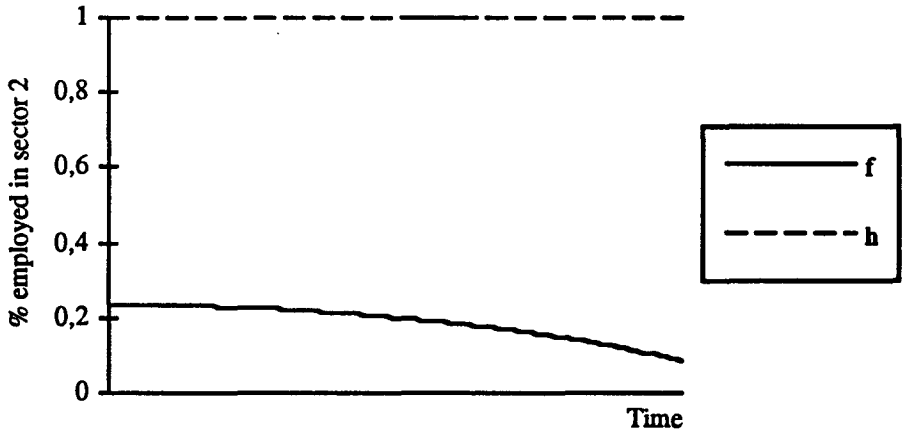


Figure 12: Percent employed in sector 2 — case (iv)

Since the size of the labour stock in region *h* is above a critical level, the spillover as well as the growth effects created in industry 2, are stronger than those prevailing in region *f*. Hence, the technological lead of region *h* is not challenged at any time, and the importance of sector 2 in region *f* is gradually reduced as the technology improvements increase the production potential of region *h*.

Labour is restricted in region h, the technologically leading region. As a consequence, the technological improvements in this region entail an increasing wage rate. As for wages in region f, these are subject to a continuous decrease, caused by the escalating technology gap between the two regions.

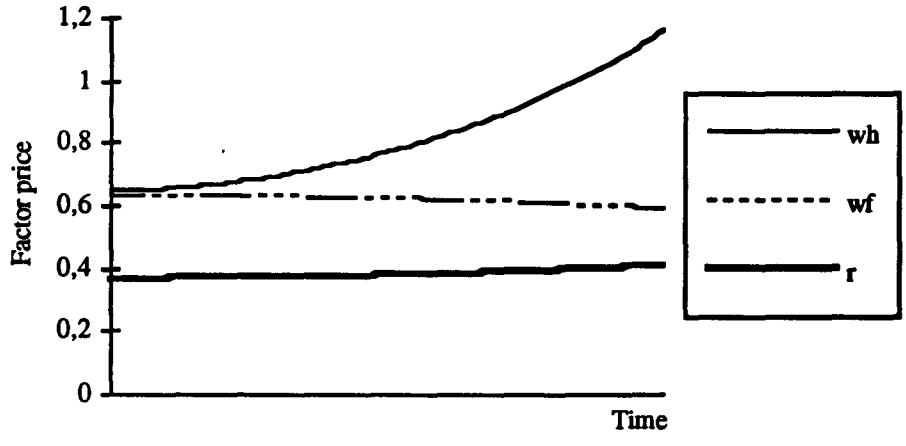
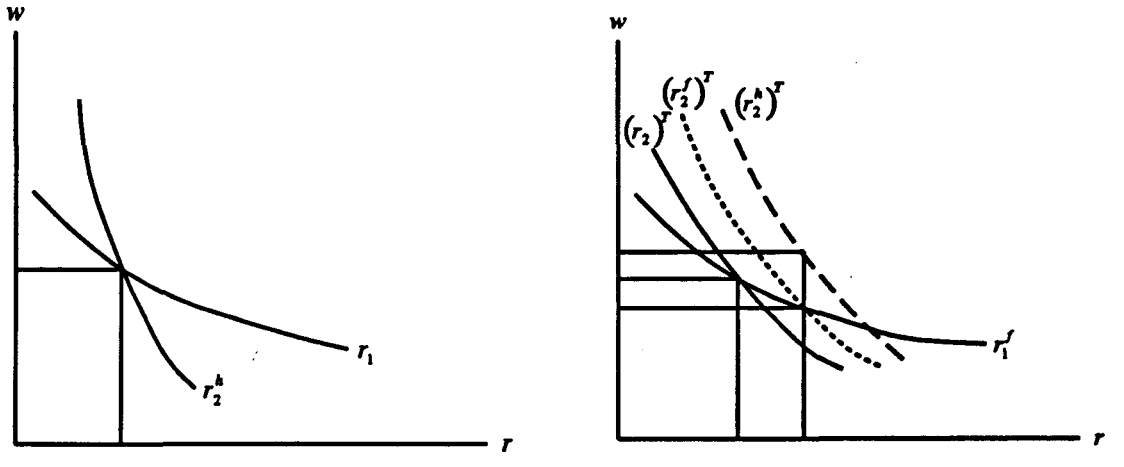


Figure 13: Factor prices — case (iv)

In order to compare case (iv) with case (iii), it may be useful once more to consider returns to capital as a function of the wage rate. Figure 14 illustrates the fact that, in case (iii), concentration of industry 2 in one country and production of good 1 in both countries, entail factor price equalisation. In case (iv), trade causes factor price equalisation in the first place. Because labour endowments inhibit a concentration of industry 2 in the technological leading country, the factor prices will, however, not stay equalised over time. Although free capital mobility ensures equal returns to capital internationally, wages diverge over time since the positive net effect of the technological progress is more significant in region h than in region f. Trade in

capital entails that equilibrium returns to capital are always determined by the intersection of r_1^f and $\left(r_2^f\right)^T$, "regardless of" $\left(r_2^h\right)^T$.



case (iii) case (iv)

Figure 14

The technological development is in case (iv) reflected by increasing returns to capital, a decreasing price of good 2, and wages evolving as described in figures 13 and 14 above. Despite the diminishing wages in region *f*, all individuals in the world experience increasing welfare. The inhabitants of region *h* do, however, gain considerably more over time than their neighbours in region *f*. Not only do returns to capital increase in region *h*, but so does the wage rate, at the same time as the price of good 2 decreases. These benefits are reinforced by the higher per capita capital income in region *h*, due to the initial distribution of endowments internationally. Hence, trade and uneven technological development trigger a highly divergent economic development, that is enhanced by international differences in initial relative factor endowments (see figure 15).

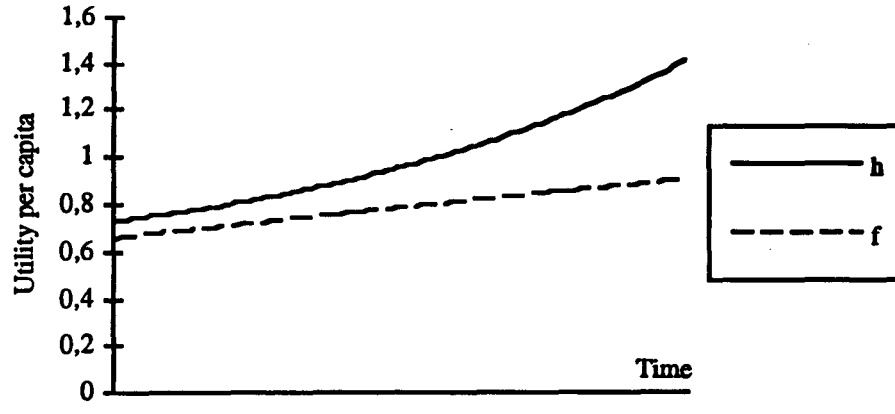


Figure 15: Utility per capita - case (iv)

(v) From dispersion towards concentration

In case (ii) we saw that an increase in factor endowments in region *f* deprived region *h* of its technological lead, and led to a reversion of trade patterns. If we do the same numerical experiment in the case with capital mobility, this does not happen (see case (iv)). Region *h* reinforces its technological leading position as time passes by. Starting with the same labour to capital ratios and the same parameter values as in case (ii), it was necessary to approximately double the size of region *f*, in order to see trade patterns reversed and factor prices equalised over time (see figure 16).

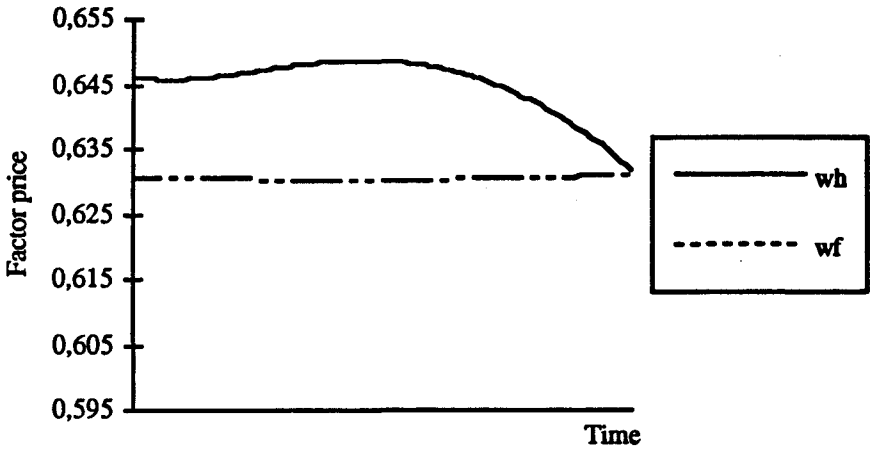


Figure 16: Factor prices — case(v)

Initially, region *h* specialises in sector 2 production, but as in case (iv), the region is not large enough to cover total world demand for good 2. Relative to region *f*, the labour supply in region *h* is below a critical level, so that region *h* is not able to sustain the industrial cluster over time. Although the spillover effect is stronger in country *h* than in country *f*, the growth effect is stronger in the latter country. Gradually, the growth effect experienced by the firms in country *f* outweighs the spillover effect in country *h*, so that the external economies prevailing in region *f* are more significant than those in region *h*. Consequently, the smaller region loses its technological lead. At the time when the regions reach the same technological level, the wage rates become equalised internationally. The wage rate in region *h* will not fall below the level of region *f* as the technological development continues, but the sector 2 industry will move from region *h* to region *f*, while the labour in region *h* will become employed in the production of good 1. Thus, the regions' industrial structures are dramatically altered, and trade patterns are reversed.

As for returns to capital, these decrease until factor prices are equalised internationally. At this stage, industry 2 is completely concentrated in region *f*, and in region *h* one shifts to the producing of good 1. It follows that, in the long run, both regions share in the production of good 1. Figure 17 is elucidative of how factor prices evolve before factor price equalisation is finally attained. The shifts in the returns to capital curves depict the net effects of the technological development on returns to capital. It appears that in this case, the positive effect on returns to capital caused by the technological progress is more than compensated for by the price decrease induced by the same progress.

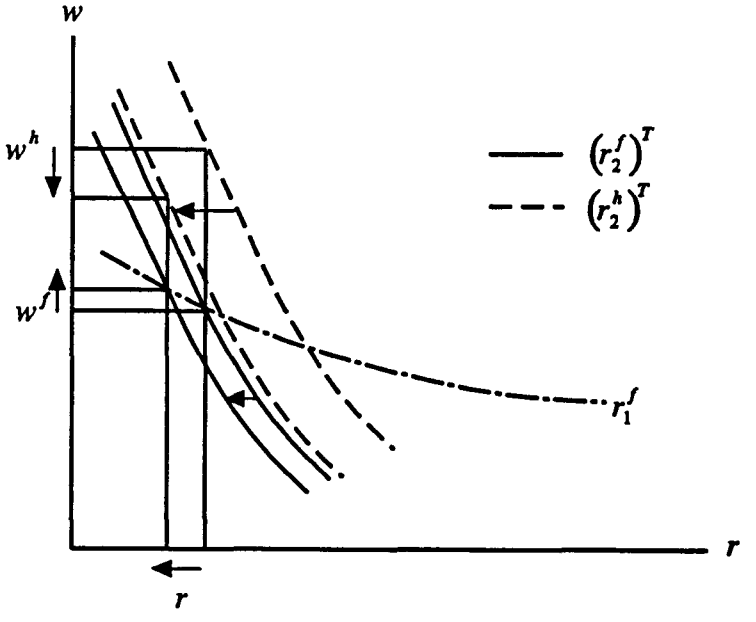


Figure 17

The price of the good 2 is subject to constantly decrease. In the long run, the welfare implications are the same in this case as in case (iii), where free trade in goods and

capital also gave factor price equalisation: world and national welfare increases, and the location of the industrial cluster becomes irrelevant to consumers' welfare, though differences in relative factor endowments cause national disparities.

4. Summary and concluding remarks

Two effects drive the technological development in the model, the innovations in each production unit, and the rate of diffusion of these new innovations among firms in each region. The number of innovations in a region increases with the size of the region's industry, while sector specific knowledge (innovations) is spread as workers change jobs within a sector. The probability that a new employee comes from a specific sector, is given by the number of workers in this sector relative to the total number of workers. Innovations are therefore more rapidly spread if the sector employs a large share of the work force in the region.

Most literature on international trade let the market size depict the size of a country. In this paper we have employed an alternative approach, letting production capacity (labour supply) denote the country size. Our definition of country size, and focus on external instead of internal economies of scale may explain why the results we achieve diverge from well known results in the trade literature on scale economies and trade patterns: various authors find that, under the assumption that two countries have the same composition of demand, the larger country specialises in goods whose production involves economies of scale — a result that is not consistent with ours.[9] We find that if the small country for some reason gets a "head start", the spillover

[9] See e.g. Krugman (1980).

effect experienced in this country may be so strong, that it outweighs a relatively stronger growth effect in the larger country. Hence, despite its size, the small country may specialise in goods whose production is characterised by economies of scale.

It appears, that the size of the different countries in terms of labour and capital supply; relative factor endowments; the share of the income spent on products produced by the industry characterised by external economies; degree of international integration; and the importance of growth and spillover effects — may all be decisive for how industrial structure and trade patterns evolve. The minimum size of a region necessary to sustain a cluster, is substantially reduced as one allows for free trade, not just in goods, but in capital as well. Obviously, our results point in the same direction as the "new" location theory, which argues that a small country is better off with complete integration.

Factor price equalisation is a necessary, but not a sufficient, condition to secure equal consumer welfare across regions. To achieve factor price equalisation, concentration of the industry subject to external economies in one country is required, and both countries must share in the production of the goods produced in the industry without external economies. Location becomes irrelevant to consumers' welfare if factor prices are equalised and there are no differences in relative factor endowments across regions.

As for an efficient use of resources, we find that for the world as a whole to exploit resources most efficiently, an industrial cluster should though be located in the smallest region which is still large enough to cover the total world demand for the

good in question. By locating the cluster in such a region, the strongest spillover effects are obtained, and the most rapid and significant technological progress is ensured.

What is not considered in this paper, is that the substantial changes in the economic geography of nations as described here, imply serious adjustment problems. Temporary unemployment is just one of several aspects associated with such changes.

Finally, it is tempting to consider the results we have obtained in a more historical context: imagine that at some point of time a small country has a comparative advantage in the production of a specific good due to favourable factor endowments. Over time, the transport of inputs such as natural resources, becomes cheaper and requires less time: the trend towards "global sourcing" means that the location of inputs has gradually become less relevant to the localisation of manufacturers. Provided that the small country in the meantime has reached a higher technological level than the rest of the world, because of its long experience in producing these goods, it will be able to sustain a competitive cluster despite the increased factor mobility. In fact, it appears that when confronted with larger regional blocks, the more open the small country is to trade in outputs as well as inputs, the more sustainable is its industrial cluster.

Appendix A

Case (i): figures 1-3; base case assumptions of parameter values and endowments.

p_1	1
α	0.6
ε	0.4
γ	0.5
ϕ	0.008
λ	0.008
n^h	1
n^f	1.5
k^h	2
k^f	2.5

Case (ii): figures 4-6; changes from case (i): $n^f = 3.75$ and $k^f = 6.25$.

Case (iii): figures 7-8; changes from case (i): $\gamma = 0.7$ and trade in capital.

Case (iv): figure 9-11; changes from case (i): $n^f = 3.75$ and $k^f = 6.25$ and trade in capital.

Case (v): figure 12; changes from case (i): $n^f = 8.4$ and $k^f = 14$ and trade in capital.

Appendix B

Free trade in goods

Equilibrium is defined by the equations (A1) - (A8):

$$c_1^i = \frac{\gamma}{p_1}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A1}$$

$$c_2^i = \frac{1-\gamma}{p_2}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A2}$$

$$n^i = \left(\frac{w^i}{r^i}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha-1} x_1^i + \left(\Psi^i\right)^{-1}\left(\frac{w^i}{r^i}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon-1} x_2^i, \quad \forall\, i = h, f \tag{A3}$$

$$k^i = \left(\frac{w^i}{r^i}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha} x_1^i + \left(\Psi^i\right)^{-1}\left(\frac{w^i}{r^i}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon} x_2^i, \quad \forall\, i = h, f \tag{A4}$$

$$x_1^h + x_1^f = c_1^h + c_1^f \tag{A5}$$

$$x_2^h + x_2^f = c_2^h + c_2^f \tag{A6}$$

$$p_1 = r^i\left(\frac{w^i}{r^i}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right), \quad \forall\, i = h, f \tag{A7}$$

$$p_2 = r^i\left(\Psi^i\right)^{-1}\left(\frac{w^i}{r^i}\right)^{\varepsilon}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1}{1-\varepsilon}\right), \quad \forall\, i = h, f. \tag{A8}$$

Equations (A3), (A4), (A7) and (A8) give:

$$w^i n^i = \alpha p_1 x_1^i + \varepsilon p_2 x_2^i, \quad \forall\, i = h, f \tag{A9}$$

$$r^i k^i = (1-\alpha) p_1 x_1^i + (1-\varepsilon) p_2 x_2^i, \quad \forall\, i = h, f. \tag{A10}$$

Solving the four equations (A9) and (A10) and using $p_1 = 1$ to substitute, gives the equilibrium production volumes as functions of factor prices:

$$x_1^i = \frac{(1-\varepsilon) w^i n^i - \varepsilon r^i k^i}{\alpha - \varepsilon}, \quad \forall\, i = h, f \tag{A11}$$

$$x_2^i = \frac{\alpha r^i k^i - (1-\alpha) w^i n^i}{(\alpha - \varepsilon) p_2}, \quad \forall\, i = h, f \tag{A12}$$

where (A11) appears as expression (21) in the text, and (A12) as expression (22).

Equations (A7) and (A8) give:

$$r^i = p_1^{\frac{\varepsilon}{\varepsilon-\alpha}} \left(\frac{\alpha}{1-\alpha}\right)^{\frac{\alpha\varepsilon}{\varepsilon-\alpha}} (1-\alpha)^{\frac{\varepsilon}{\varepsilon-\alpha}} p_2^{-\frac{\alpha}{\varepsilon-\alpha}} \left(\Psi^i\right)^{-\frac{\alpha}{\varepsilon-\alpha}} \left(\frac{\varepsilon}{1-\varepsilon}\right)^{-\frac{\alpha\varepsilon}{\varepsilon-\alpha}} (1-\varepsilon)^{-\frac{\alpha}{\varepsilon-\alpha}}, \tag{A13}$$
$$\forall\, i = h, f$$

$$w^i = p_1^{\frac{\varepsilon-1}{\varepsilon-\alpha}} \left(\frac{\alpha}{1-\alpha}\right)^{\frac{\alpha(\varepsilon-1)}{\varepsilon-\alpha}} (1-\alpha)^{\frac{\varepsilon-1}{\varepsilon-\alpha}} p_2^{\frac{1-\alpha}{\varepsilon-\alpha}} \left(\Psi^i\right)^{\frac{1-\alpha}{\varepsilon-\alpha}} \left(\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{\varepsilon(1-\alpha)}{\varepsilon-\alpha}} (1-\varepsilon)^{\frac{1-\alpha}{\varepsilon-\alpha}}, \tag{A14}$$
$$\forall\, i = h, f$$

Using equations (A1) to substitute in equation (A5), and equations (A2) to substitute in equation (A6), we get two equations that can be solved to give

$$\frac{x_1^h + x_1^f}{x_2^h + x_2^f} = \frac{p_2}{p_1} \cdot \frac{\gamma}{1-\gamma}. \tag{A15}$$

Using equations (A11) and (A12) to substitute with in equation (A15), we get that

$$r^h = \frac{(1-\gamma)(1-\varepsilon)+\gamma(1-\alpha)}{(1-\gamma)\varepsilon+\gamma\alpha} \cdot \left(w^h n^h + w^f n^f\right) \cdot \frac{1}{k^h} - r^f \cdot \frac{k^f}{k^h}. \tag{A16}$$

Equations (A13), (A14) and (A16) can be solved to give p_2 as a function of p_1; and using $p_1 = 1$ to substitute, we get the equilibrium price of good 2

$$p_2 = \left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1-\alpha}{1-\varepsilon}\right)\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\frac{N}{K}A\right)^{(\alpha-\varepsilon)}. \tag{A17}$$

(A17) appears as expression (16) in the text, and the definitions of A, K, and N are found at pages 10 and 11.

Using equation (A17) to substitute in equations (A13) and (A14), we get the equilibrium factor prices:

$$r^i = \left(\Psi^i\right)^{\frac{\alpha}{\alpha-\varepsilon}}\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(\frac{N}{K}A\right)^{\alpha}, \quad \forall\, i = h, f \tag{A18}$$

$$w^i = \left(\Psi^i\right)^{\frac{\alpha-1}{\alpha-\varepsilon}}\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(\frac{K}{N}A^{-1}\right)^{1-\alpha}, \quad \forall\, i = h, f, \tag{A19}$$

where (A18) appears as expression (19) in the text, while (A19) appears as expression (20).

Free trade in goods and capital: case (iii)

By assumption we have that $x_1^h, x_1^f > 0$, $x_2^h > 0$, and $x_2^f = 0$. Equilibrium is accordingly, defined by the equations (A20) - (A28):

$$c_1^i = \frac{\gamma}{p_1}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A20}$$

$$c_2^i = \frac{1-\gamma}{p_2}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A21}$$

$$n^h = \left(\frac{w^h}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha-1} x_1^h + \left(\Psi^h\right)^{-1}\left(\frac{w^h}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon-1} x_2^h \tag{A22}$$

$$n^h = \left(\frac{w^f}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha-1} x_1^f \tag{A23}$$

$$k = \left(\frac{w^h}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha} x_1^h + \left(\frac{w^f}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha} x_1^f + \left(\Psi^h\right)^{-1}\left(\frac{w^h}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon} x_2^h \tag{A24}$$

$$x_1^h + x_1^f = c_1^h + c_1^f \tag{A25}$$

$$x_2^h = c_2^h + c_2^f \tag{A26}$$

$$p_1 = r\left(\frac{w^i}{r}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right), \quad \forall\, i = h, f \tag{A27}$$

$$p_2 = r\left(\Psi^h\right)^{-1}\left(\frac{w^h}{r}\right)^{\varepsilon}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1}{1-\varepsilon}\right). \tag{A28}$$

From equations (A27) it can be derived, that

$$w^h = w^f = w, \tag{A29}$$

and we may rewrite the two equations (A27) as

$$p_1 = r\left(\frac{w}{r}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right). \tag{A30}$$

Equations (A22) - (A24), (A27) and (A28) give:

$$w\, n^h = \alpha p_1 x_1^h + \varepsilon p_2 x_2^h \tag{A31}$$

$$w\, n^f = \alpha p_1 x_1^f \tag{A32}$$

$$r\, k = (1-\alpha)p_1\left(x_1^h + x_1^f\right) + (1-\varepsilon)p_2 x_2^h. \tag{A33}$$

Solving the three equations (A31) - (A33) and using $p_1 = 1$ to substitute, gives the equilibrium production volumes as functions of factor prices:

$$x_1^h = \frac{(1-\varepsilon)wn^h - \varepsilon rk}{\alpha - \varepsilon} + \frac{\varepsilon(1-\alpha)wn^f}{\alpha(\alpha-\varepsilon)} \tag{A34}$$

$$x_1^f = \frac{wn^f}{\alpha} \tag{A35}$$

$$x_2^h = \frac{\alpha r\, k - (1-\alpha)wn}{(\alpha-\varepsilon)p_2}, \tag{A36}$$

where (A34) appears as expression (27) in the text, (A35) as expression (28), and (A36) as expression (26).

Using (A29), equations (A27) and (A30), can be solved to give:

$$r = p_1^{\frac{\varepsilon}{\varepsilon-\alpha}}\left(\frac{\alpha}{1-\alpha}\right)^{\frac{\alpha\varepsilon}{\varepsilon-\alpha}}(1-\alpha)^{\frac{\varepsilon}{\varepsilon-\alpha}}p_2^{-\frac{\alpha}{\varepsilon-\alpha}}\left(\Psi^h\right)^{-\frac{\alpha}{\varepsilon-\alpha}}\left(\frac{\varepsilon}{1-\varepsilon}\right)^{-\frac{\alpha\varepsilon}{\varepsilon-\alpha}}(1-\varepsilon)^{-\frac{\alpha}{\varepsilon-\alpha}} \qquad \text{(A37)}$$

$$w = p_1^{\frac{\varepsilon-1}{\varepsilon-\alpha}}\left(\frac{\alpha}{1-\alpha}\right)^{\frac{\alpha(\varepsilon-1)}{\varepsilon-\alpha}}(1-\alpha)^{\frac{\varepsilon-1}{\varepsilon-\alpha}}p_2^{\frac{1-\alpha}{\varepsilon-\alpha}}\left(\Psi^h\right)^{\frac{1-\alpha}{\varepsilon-\alpha}}\left(\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{\varepsilon(1-\alpha)}{\varepsilon-\alpha}}(1-\varepsilon)^{\frac{1-\alpha}{\varepsilon-\alpha}}. \qquad \text{(A38)}$$

Using equations (A20) to substitute in equation (A25), and equations (A21) to substitute in equation (A26), we get two equations that can be solved to give

$$x_2^h = \frac{p_1}{p_2}\cdot\frac{1-\gamma}{\gamma}\left(x_1^h + x_1^f\right). \qquad \text{(A39)}$$

We use (A39) to substitute for x_2^h in equations (A31) and (A33), add (A31) and (A32) together, and solve this new equation as well as equation (A33), to get

$$w = \frac{k}{n}\cdot\underbrace{\frac{(1-\gamma)\varepsilon+\gamma\alpha}{(1-\gamma)(1-\varepsilon)+\gamma(1-\alpha)}}_{=A^{-1}}r. \qquad \text{(A40)}$$

Equations (A38) and (A40), may now be solved for r, while this new expression for r, and (A37) can be solved to give p_2 as a function of p_1. Using $p_1 = 1$ to substitute, we get the equilibrium price of good 2:

$$p_2 = \left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1-\alpha}{1-\varepsilon}\right)\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\Psi^{h}\right)^{-1}\left(\frac{n}{k}A\right)^{(\alpha-\varepsilon)}. \qquad \text{(A41)}$$

(A41) appears as expression (23) in the text.

Using equation (A41) to substitute in the expression for r that was found by using (A38) and (A40), we get the equilibrium returns to capital:

$$r = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(\frac{n}{k}A\right)^{\alpha}, \qquad \text{(A42)}$$

and substituting back into (A40), we get the equilibrium wage rate:

$$w = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)\left(\frac{k}{n}A^{-1}\right)^{1-\alpha}. \qquad \text{(A43)}$$

(A42) appears as expression (25) in the text, and (A43) as expression (24).

Free trade in goods and capital: cases (iv) and (v)

By assumption we have that $x_1^h = 0$, $x_1^f > 0$, and $x_2^h, x_2^f > 0$. Equilibrium is accordingly, defined by equations (A44) - (A52):

$$c_1^i = \frac{\gamma}{p_1}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A44}$$

$$c_2^i = \frac{1-\gamma}{p_2}\left(w^i n^i + r^i k^i\right), \quad \forall\, i = h, f \tag{A45}$$

$$n^h = \left(\Psi^h\right)^{-1}\left(\frac{w^h}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon-1} x_2^h \tag{A46}$$

$$n^f = \left(\frac{w^f}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha-1} x_1^f + \left(\Psi^f\right)^{-1}\left(\frac{w^f}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon-1} x_2^f \tag{A47}$$

$$k = \left(\frac{w^f}{r}\cdot\frac{1-\alpha}{\alpha}\right)^{\alpha} x_1^f + \left(\Psi^h\right)^{-1}\left(\frac{w^h}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon} x_2^h + \left(\Psi^f\right)^{-1}\left(\frac{w^f}{r}\cdot\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon} x_2^f \tag{A48}$$

$$x_1^f = c_1^h + c_1^f \tag{A49}$$

$$x_2^h + x_2^f = c_2^h + c_2^f \tag{A50}$$

$$p_1 = r\left(\frac{w^f}{r}\right)^{\alpha}\left(\frac{1-\alpha}{\alpha}\right)^{\alpha}\left(\frac{1}{1-\alpha}\right) \tag{A51}$$

$$p_2 = r^i\left(\Psi^i\right)^{-1}\left(\frac{w^i}{r}\right)^{\varepsilon}\left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1}{1-\varepsilon}\right), \quad \forall\, i = h, f. \tag{A52}$$

Equations (A46) - (A48), (A51) and (A52) give:

$$w^h n^h = \alpha p_2 x_2^h \tag{A53}$$

$$w^f n^f = \alpha p_1 x_1^f + \varepsilon p_2 x_2^f \tag{A54}$$

$$r\,k\; = (1-\alpha) p_1 x_1^f + (1-\varepsilon) p_2 \left(x_2^h + x_2^f \right). \tag{A55}$$

Solving the three equations (A53) - (A55) and using $p_1 = 1$ to substitute, gives the equilibrium production volumes as functions of factor prices:

$$x_1^f = \frac{(1-\varepsilon) w^f n^f - \varepsilon r k\; + (1-\varepsilon) w^h n^h}{\alpha - \varepsilon} \tag{A56}$$

$$x_2^h = \frac{w^h n^h}{p_2 \varepsilon} \tag{A57}$$

$$x_2^f = \frac{\alpha r k - (1-\alpha) w^f n^f}{(\alpha - \varepsilon) p_2} - \frac{\alpha (1-\varepsilon) w^h n^h}{\varepsilon (\alpha - \varepsilon) p_2}, \tag{A58}$$

where (A56) appears as expression (35) in the article, (A57) as expression (33), and (A58) as expression (34).

Equations (A51) and (A52) can be solved to give:

$$r = p_1^{\frac{\varepsilon}{\varepsilon-\alpha}} \left(\frac{\alpha}{1-\alpha} \right)^{\frac{\alpha\varepsilon}{\varepsilon-\alpha}} (1-\alpha)^{\frac{\varepsilon}{\varepsilon-\alpha}} p_2^{-\frac{\alpha}{\varepsilon-\alpha}} \left(\Psi^f \right)^{-\frac{\alpha}{\varepsilon-\alpha}} \left(\frac{\varepsilon}{1-\varepsilon} \right)^{-\frac{\alpha\varepsilon}{\varepsilon-\alpha}} (1-\varepsilon)^{-\frac{\alpha}{\varepsilon-\alpha}} \tag{A59}$$

$$w^f = p_1^{\frac{\varepsilon-1}{\varepsilon-\alpha}}\left(\frac{\alpha}{1-\alpha}\right)^{\frac{\alpha(\varepsilon-1)}{\varepsilon-\alpha}}(1-\alpha)^{\frac{\varepsilon-1}{\varepsilon-\alpha}}p_2^{\frac{1-\alpha}{\varepsilon-\alpha}}\left(\Psi^f\right)^{\frac{1-\alpha}{\varepsilon-\alpha}}\left(\frac{\varepsilon}{1-\varepsilon}\right)^{\frac{\varepsilon(1-\alpha)}{\varepsilon-\alpha}}(1-\varepsilon)^{\frac{1-\alpha}{\varepsilon-\alpha}} \quad \text{(A60)}$$

$$w^h = w^f\left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}} \quad \text{(A61)}$$

Equation (A61) should be recognised as the expression (32) in the article.

Using equations (A44) to substitute in equation (A49), and equations (A45) to substitute in equation (A50), we get two equations that can be solved to give

$$\frac{x_1^h}{x_2^h + x_2^f} = \frac{p_2}{p_1}.\frac{\gamma}{1-\gamma}. \quad \text{(A62)}$$

We use equations (A56) - (A58) to substitute for production volumes in equation (A62), and solve for r to get

$$r = A\frac{1}{k}\left(w^h n^h + w^f n^f\right). \quad \text{(A63)}$$

Equations (A59) - (A61) and (A63), may now be solved to give p_2 as a function of p_1. Using $p_1 = 1$ to substitute, we find the equilibrium price of good 2:

$$p_2 = \left(\frac{1-\varepsilon}{\varepsilon}\right)^{\varepsilon}\left(\frac{1-\alpha}{1-\varepsilon}\right)\left(\frac{\alpha}{1-\alpha}\right)^{\alpha}\left(\Psi^f\right)^{-1}A^{\alpha-\varepsilon}\left(n^f + \left(\frac{\Psi^h}{\Psi^f}\right)^{\frac{1}{\varepsilon}}n^h\right)^{\alpha-\varepsilon}k^{\varepsilon-\alpha}. \quad \text{(A64)}$$

(A64) appears as expression (29) in the article.

Using equation (A64) to substitute in equation (A59), we get the equilibrium returns to capital:

$$r = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)A^{\alpha}\left[n^{f}+\left(\frac{\Psi^{h}}{\Psi^{f}}\right)^{\frac{1}{\epsilon}}n^{h}\right]^{\alpha}k^{-\alpha}, \qquad \text{(A65)}$$

while we get the equilibrium wage rate in country by using (A64) to substitute in equation (A60):

$$w^{f} = \left(\frac{\alpha}{1-\alpha}\right)^{\alpha}(1-\alpha)A^{\alpha-1}\left[n^{f}+\left(\frac{\Psi^{h}}{\Psi^{f}}\right)^{\frac{1}{\epsilon}}n^{h}\right]^{\alpha-1}k^{1-\alpha}. \qquad \text{(A66)}$$

(A65) appears as expression (30) in the article, and (A66) as expression (31).

References

Braun, Martin, (1983): "The Spread of Technological Innovations", Chapter 6 in Braun, Coleman and Drew (eds.): *Modules in Applied Mathematics - Differential Equation Models*, Springer-Verlag, 1983

Brezis, Elise S., Paul R. Krugman, and Daniel Tsiddon (1993): "Leapfrogging in international competition: A theory of cycles in national technological leadership", *American Economic Review* 83 (1993): 1211-1219

Dixit, Avinash and Victor D. Norman, (1980): *Theory of International Trade*, Cambridge: Cambridge University Press, 1980

Grossman, Gene and Elhanan Helpman (1991): *Innovation and growth in the global economy,* Cambridge, Massachusetts: MIT Press, 1991

Helpman, Elhanan and Paul R. Krugman (1985): *Market structure and foreign trade*, Cambridge, Massachusetts: MIT Press, 1985

Knarvik, Karen H. Midelfart, (1995): "Technological spillovers, industrial clusters and economic integration", Institute of Economics, Norwegian School of Economics and Business Administration Discussion paper 6/95, Bergen, 1995

Krugman, Paul R. (1980): "Scale economies, product differentiation and the pattern of trade", *American Economic Review* 70 (1980): 950-959

Krugman, Paul R. (1981): "Trade, accumulation, and uneven development", *Journal of Development Economics* 8 (1981): 149-161

Krugman, Paul R. (1993): "Blockheaded trading blocks", in "The World in 1994", *The Economist*, 1993.

Krugman, Paul. R. and Anthony J. Venables (1993): "Integration, specialization and adjustment", NBER Working Paper No. 4559, Cambridge, Massachusetts, 1993

Krugman, Paul. R. and Anthony J. Venables (1994): "Globalization and the inequality of nations", CEPR Discussion Paper No. 1015, London, 1994

Mansfield, Edwin, (1961): "Technical Change and the Rate of Imitation", *Econometrica* 29 (1961): 741-765

Marshall, Alfred. (1920): *Principles of economics*, London: Macmillan, 1920

Norman, Victor D. and Anthony J. Venables, (1993): "International trade, factor mobility and trade costs", CEPR Discussion Paper No. 766, London, 1993

Pissarides, Christopher A. (1985): "Short-run equilibrium dynamics of unemployment, vacancies, and real wages", *American Economic Review* 75 (1985): 676-690

Porter, Michael E. (1990): *The competitive advantage of nations*, New York: Free Press, 1990

Rivera-Batiz, Luis A. and Paul M. Romer (1991): "Economic integration and endogenous growth", *Quarterly Journal of Economics* 106 (1991): 531-556

Essay 5

Does it matter where you are? The importance of being localised in the academic world.

by

Karen Helene Midelfart Knarvik*

Introduction

In a time of globalisation, one might expect that "[i]nstantaneous global telecommunications, television and computer networks will soon overthrow the ancient tyrannies of time and space".[1] As a natural consequence of such a development, geography (localisation) would cease to matter, and one would tend towards a smooth dispersion of people, skills, economic competence, and manufacturing across the continents. But reality proves to be otherwise. Even the newest industries are obeying the old rule of geographical localisation, namely to concentrate in regions, cities or towns. And there are no signs of this "clustering trend" declining.

It is therefore not surprising, that during the last years Alfred Marshall's classic economic analysis of industrial localisation and concept of external economies have

* I would like to thank Jan I. Haaland, Victor Norman, Frode Steen, Jostein Tvedt and Anthony J. Venables for valuable comments and suggestions.

[1] The Economist, July 30th 1994, p. 11.

experienced a renaissance: in 1990 Michael Porter introduced the phenomenon "industrial clusters" to businessmen and academicians, thereby emphasising the link between such clusters and international competitiveness. At approximately the same time Paul Krugman and Anthony Venables started developing what is often called the "new" location theory with ingredients from trade theory as well as from traditional economic geography.[2] They address two fundamental questions: why and where do firms concentrate geographically; and offer microeconomic tools to analyse problems of localisation.

There are several reasons why firms should wish to locate close to other firms. Marshall recognised three, which are still considered some of the most important: 1) access to a large variety of intermediates at a low cost; 2) a pooled labour market for workers with specialised skills; and 3) knowledge spillovers. Porter emphasises the importance of an innovating, stimulating environment, proximity to customers and suppliers, and rivalry for a firm's competitiveness — which (according to Porter) are all typical characteristica of an efficient industrial cluster.

While *firms'* choice of location and the impact of this on productivity and competitiveness have caught the attention of a number of economists, little has been written about the relevance (or irrelevance) of academicians' location. How important is an academician's, say, an economist's, location to his academic performance? May one compare an economist with a firm in the sense that, also to an economist the proximity to good colleagues who offer a stimulating, rival environment with a rapid transmission of new ideas and knowledge, is of great importance? Is he able to compete better — meet the requirements of the academic world better — given that he

[2] See e.g. Krugman (1991), and Krugman and Venables (1993).

is located in an efficient academic cluster? And if this is so, does the faculty of the department of economics where he is employed, to some extent determine the quantity and quality of his research, or is an economist's research productivity independent of his geographic location? Alternatively, did the site of employment once matter, but has become irrelevant due to the rapid technological development and expanding international networks?

Closely related to the question about the determinants of individual academic performance and the incentives for academic agglomeration, is the following question: what determines the performance of a department of economics? Why is it that some departments are "good", and remain good over a long period of time, while some perform pretty poorly? One plausible answer is that good departments attract competent people. However, an alternative explanation is that good departments "produce" good people. The latter explanation is consistent with the school of thought that emphasises the importance of social climate and knowledge spillovers for the creation of inventions, which suggests that an economist's environment — and thereby location — is decisive for his productivity.[3] Neither of the two explanations does, however, exclude the relevance of the other.

Provided that location matters, and your "surroundings" are decisive for further development of skills as well as quantity and quality of research output, an additional question arises: how does a change of places of employment, i.e., mobility, affect skills and productivity? Obviously, no department of economics is equal to another. Faculty and the range of fields covered by the faculty members differ across departments. In other words, the "surroundings" offered by one department are never

[3] See e.g. Arrow (1962), and Grossman and Helpman (1992).

equal to those offered by another. Within the management and organisational psychology literature, the importance of diverse experience is often pronounced. Hence, according to the theories and findings within this literature, a mobile academician might become more successful than someone less mobile.

In this paper we aim to take a closer look at possible determinants of individual and departmental performance, and thereby focus in particular on the importance of localisation in the academic world. Our main hypothesis is that, being at a department ranked as good, improves an economist's performance, so that an academic cluster might be just as important to its incumbents as an industrial cluster to firms.

The paper is organised as follows: in section 1 the notion "an academician's productivity" is defined. In section 2 data sets and methodology are described, while in section 3 possible indicators of academic performance are discussed and tested. Section 4 concludes.

1. Defining the productivity of an academician

The greater institutional emphasis on research and the increasing pressure to publish in well-known, refereed journals are probably responsible for the fact that we often tend to equate academic performance with research productivity. "Publish or Perish" — the title of an article by Zivney and Bertin (1992) — does, perhaps, quite accurately reflect the way academicians have come to be evaluated. So do the lines below, which were seen at a noticeboard at the London School of Economics and Political Science:[4]

[4] See London School of Economics and Political Science (1995).

Why God never received tenure at the LSE:

1. because He had only one major publication;
2. and it wasn't published in a refereed journal;
3. and it had no references;
4. and it wasn't in English;
5. and verification of his results has proved impossible;
6. and it is an all-time best-seller.

Although one might object that academic productivity has more dimensions, in this paper we shall focus solely on research productivity. Research productivity is most commonly measured by a person's number of published articles, by page counts of published articles, or by number of citations. There is a considerable literature employing and discussing these different methods.

Although the second approach to the measuring of productivity is widely applied, we doubt whether this is a good measure of research quality and quantity. Our main objection is that, to our knowledge, there is no evidence that the length of an article is correlated with the value and uniqueness of the author's intellectual contribution. Furthermore, as long as there is no such thing as a standard length of an article, number of pages might not convey a correct impression of the number of articles, i.e., ideas, published. Left with the first and the third approach, we find the former the most appropriate for our purpose. The reason being that we wish to focus on an academician's "output" (productivity) over a specific period of time, which would be much more complicated if the latter method were to be employed. Furthermore, instead of counting all published articles, we shall make a selection of major journals, and concentrate on articles published in these. Through such a selection, the strict refereeing process pursued by top journals guarantees a high overall standard of the counted publications. In other words, we let the refereeing process function as a quality control mechanism.

2. Data sets and methodology

In order to shed some light on possible determinants of individual and departmental productivity, a cross-sectional study is conducted. 175 economists are surveyed with regard to age, university of graduation, university of employment, job mobility and publication record. The sample was drawn from the American Economic Association Directory of Members.[5] The surveyed economists satisfy two conditions: (1) Ph.D. completed at a United States university during the years 1970-75; (2) employment at a United States economics department at least from 1979 until 1991. The latter period was chosen this long in order to obtain a sample consisting of people that may be regarded as "typical" academicians, in the sense that their stay in academia is more or less permanent.

We recorded where each academician received his Ph.D., as well as where the economists were employed during the period 1979-81 — at which time all (or at least most) tenure decisions had most likely been made.[6] Second, an appropriate ranking of U.S. economics departments was searched for. There is a large literature on rankings of these departments. The studies aimed at ranking economics departments apply different ranking criteria, the three following being some of the more common:

- Opinion surveys directed at department heads and senior professors.
- Studies where departmental rankings are based on page counts of articles published in top journals by faculty members in the respective departments.
- Studies where departments are ranked by the number of citations attributed to their faculty.

---

[5] This on-line database is a service provided jointly by the American Economic Association and the University of Texas at Austin.

[6] According to Hamermesh (1992) tenure decisions are usually made approximately 5 years after people have completed their Ph.D. thesis and started their first full-time academic job.

While the opinion surveys usually are regarded as subjective and less reliable, departmental ratings based on one of the two latter criteria may be used as a proxy for the quality of the research environment. Such ratings are, by some authors, also reckoned indicative of the quality of dissertation research, i.e., indicative of the "quality" or acumen of graduate students.[7] Due to the previously pronounced scepticism to page counts as a measure of research quality, we decided to employ a ranking list based on citations. Davis and Papanek (1984) provide such rankings, where citations include references to journals, books, and monographs.[8] 122 Ph.D.-granting departments are included in the rankings, which are based on the average number of citations for 1978 and 1981, and the departments are ranked by total number of citations as well as by per capita citations.

A number of the surveyed economists did not work at any of the 122 universities ranked by Davis and Papanek. Ranking lists including non-Ph.D.-granting institutions show, however, that according to total and per capita citations, few of these departments were able to "compete" with the Ph.D.-granting departments.[9] Hence, non-Ph.D.-granting departments occurring as surveyed economists' place of employment, were given the rank 130. As a consequence about, 11 % of the sample are ranked a bit lower than they should have been.

Using the Davis-Papanek rankings, the universities where the people in the sample (1) graduated (sometime between 1970 and 1975), and (2) were employed 1979-81, are evaluated. One might object that the same rankings cannot be employed to evaluate universities at two different points of time. However, published rankings of

[7] See e.g. Graves et al. (1982).
[8] See table A3 in the appendix.
[9] Compare Davis and Papanek (1984) with Blair et al. (1986).

economics departments have proved remarkably consistent even though they cover different periods of time.[10]

As for the surveyed economists' research productivity, we chose to focus on two different periods: (1) an interval of three years, starting with the year they completed their Ph.D. work, and (2) the years 1979-83. All articles published in 50 top journals within the two periods were recorded.[11] The latter period was chosen this long for two main reasons: first, in order to avoid any bias due to fluctuations in research output over the years. Second, to account for the length of time from submission to a possible publication.[12] Consequently, we hope to cover all high standard research output produced during the period 1979-81, for which the economists' employers are recorded. One should note that, our selection only includes journal articles published as "article" or "note". "Replies", "comments" and "discussions" are not counted, since the "[c]reative essence" is small in such efforts compared to that of articles or notes.[13]

---

[10] See e.g. Gibbons and Fish (1991).

[11] To select 50 top journals, two different rankings were employed. When recording publications for the period 1979-83, we used the well known ranking list by Liebowitz and Palmer (1984), which is based on 1980 citations to articles published 1975-79. (See table A2 in the appendix.) As for the econonomists' early publications (the three first years after the completion of their Ph.D.), Laband and Piette (1994) table 1, rank *a*, which is based on 1970 citations to articles published 1965-1969, was used. (See table A1 in the appendix.)

[12] See Hamermesh, D. S. (1992) and (1994) for information on the time it takes to get articles published. According to Hamermesh, maximum time from submission to a possible publication is approximately 2 years.

[13] Zivney and Bertin (1992)

3. Explaining differences in research productivity

3.1 The model

We assume that there is a linear relationship between research output published during the years 1979-83 and (i) age, (ii) university of graduation, (iii) university of employment from 1979 until 1981 (or the main part of this period), (iv) mobility, and (v) faculty size. Expressed more formally,

$$PUBL_i = f(AGE,\ UNI,\ EMPL,\ MOBILITY,\ FSIZE) \quad (1)$$
$$\qquad\qquad -\quad -\quad -\quad +\quad ?$$

where

$PUBL_i$ = publications in 50 top journals 1979-83 by economist *i*;

AGE = age in 1979;

UNI = university of graduation ranked according to Davis-Papanek table 1, rank *b*, which ranks universities by per capita citations in descending order, the best university thereby given the rank "one";

EMPL = university of employment during the years 1979-81 ranked according to Davis-Papanek table 1, rank *b*;

FSIZE = size of the faculty of the department where the academician was employed from 1979 until 1981.

MOBILITY = mobility after the dissertation was completed and up to 1981. People are either "mobile" or "immobile". A mobile person is defined as someone who has changed employers between graduation and 1981, while an immobile person is in 1981 still with his first employer after graduation. To indicate which category each person fits into, a dummy variable is used (1 = mobile; 0 = immobile).

The expected signs of the coefficients are shown below each variable. Our expectations about the impact of different independent variables are based on various hypotheses, at which we shall take a closer look before turning to the results of the regressions run.

According to findings by Tuckman and Leahey (1975) a.o., the returns to academic publishing in terms of benefits such as direct salary increments, promotion-related salary increments, and career-related options, are diminishing. Typically, an assistant professor is found to have higher returns to publishing than a full professor. Thus, we would expect age to have a negative effect on publication. Related to the variable "age", however, are a few problems. First, it seems to be a widely shared view that the relationship between publishing and age is a highly non-monotonic one, where the number of publications is increasing during the first years of an academician's career, for later to start decreasing. Assuming, as we do, a linear relationship between publications and age, may, in other words, be a troublesome assumption, that requires careful interpretation of the results. Second, one may question if the sample here is suitable for testing of the influence of age on publishing. Since all surveyed economists completed their Ph.D. between 1970 and 1975, one would not expect them to differ that much in age. But although a closer look at the data on age shows a pretty low standard deviation, almost 25 % of the sample are found to lie outside the interval defined by mean ± standard deviation. Hence, there is a dispersion in age that opens for a test of how age influences publishing.

The signs below UNI and EMPL may seem rather counterintuitive, but this is simply due to the descending order in which the departments are ranked. The lower the rank of a university, the higher the number of its ranking, and the fewer articles graduates

and faculty members are expected to publish. As regards the university of graduation, we expect an academician's productivity to be higher, the better the university at which he completed his Ph.D. thesis, since faculty ratings are often reckoned indicative of the "quality" of graduate students. Consequently, we presume the coefficient of the variable "university of graduation" to have a negative sign.

Consistent with the hypotheses that "good departments attract competent people", and "good departments "produce" competent people", one would, moreover, expect the rank of the university at which a surveyed economist was employed from 1979-81, to be indicative of his research output from 1979-83. Thus, the higher a department is ranked, the more high standard research is presumably published by its members. In other words, the variable EMPL is assumed to have a negative impact on publications.

Associated with the impact of the department of employment on publishing, however, is an endogeneity problem: departments are ranked according to mean number of citations attributed to their faculty members, and one may presume that the number of citations they attain, depends on their number of publications. To what extent EMPL constitutes an independent variable may, in other words, be questioned. But we argue that only 13 % of the sample are employed in departments with less than 10 members, implying that how the individuals in the survey performs, is not likely to significantly affect the rank of their department. Thus, one may consider the endogeneity problem as a less serious one.

Turning to the impact of mobility, this variable is expected to be positively correlated with research output. The assumption is based on what seems to be a prevailing view, namely that mobility is important for the achievement — and further development —

of skills.[14] It is also motivated by management and organisational psychology theories, as well as by the evidence presented by Howard Marshall (1964), and Skeels and Fairbanks (1968). Based on surveys conducted in the United States during the 1960s, the three authors conclude that academic economists in general, and major publishers in particular, are highly mobile.

One should, however, be aware that the exact period of time, during which the individuals' mobility was observed, may generate odd results: within this period all tenure decisions are likely to have been made. A person's change of places of employment may therefore not necessarily reflect voluntary mobility, but instead the fact that he did not receive tenure. Consequently, the results about the impact of mobility on publishing, should be interpreted carefully, as the phenomenon "tenure" may lead to distortions.

Data on the first four independent variables described on page seven, were gathered for all 175 economists. The fifth independent variable, faculty size (FSIZE), did, however, cause some problems, as it appeared pretty complicated to obtain the size in about 1980 of the faculty of a number of smaller departments. In order to test for the influence of faculty size, a sub-sample with 131 observations was composed of those academicians for whom all data were available.

Employing the two different samples, the two equations

$$PUBL_i = \alpha + \beta_1 AGE + \beta_2 UNI + \beta_3 EMPL + \beta_4 MOBILITY, \qquad (2)$$

and

[14] See e.g. Kyvik and Tvede (1994) and the Research Council of Norway (1995).

$$PUBL_i = \alpha + \beta_1 AGE + \beta_2 UNI + \beta_3 EMPL + \beta_4 MOBILITY + \beta_5 FSIZE \qquad (3)$$

were estimated.

Tables 1 and 2 give means, standard deviations, and the simple correlation matrix for the complete as well as the "reduced" sample. When turning to the results, one should be aware that these should be interpreted with care. We have only included four and five exogenous variables respectively, and we may have omitted variables which are just as decisive for academicians' publishing activity. Omitted variables may result in biased and inconsistent estimators as well as unreliable t- and F-tests.

Table 1 — Summary Statistics (Sample size: 175 observations)

Variable	Mean	Standard Deviation			
PUBL	1.63	2.91			
AGE	36.13	4.52			
UNI	29.69	33.01			
EMPL	87.83	47.67			
Correlations:					
PUBL	1				
AGE	-.39	1			
UNI	-.32	.28	1		
EMPL	-.54	.35	.51	1	
MOBILITY	.002	-.01	-.14	-.12	1

Table 2 — Summary Statistics (Sample size: 131 observations)

Variable	Mean	Standard Deviation			
PUBL	2.14	3.20			
AGE	35.38	4.25			
UNI	22.73	28.06			
EMPL	73.67	47.30			
FSIZE	26.72	15.60			
Correlations:					
PUBL	1				
AGE	-.39	1			
UNI	-.29	.26	1		
EMPL	-.47	.31	.51	1	
MOBILITY	-.02	-.08	-.14	-.13	1
FSIZE	.42	-.31	-.40	-.75	.04	1

3.2 Results

It appeared that a substantial share of the sample had not published any articles in major journals between 1979 and 1983, i.e., the database contained numerous zero observations for the dependent variable. One may argue that, zero publications cannot be given a clear interpretation. We do not know whether a person was about to publish an article by the end of the period we observe, or whether he did any research during this period at all. In other words, the dependent variable may in our case be what we call *censored*, meaning that information is missing for the dependent variable, whereas the corresponding information for the independent variables is present.

However, it is also possible to argue that in our model the value zero does not differ from any other value. For instance, we do not know whether a person who published three articles during the surveyed period, was just about to publish his fourth by the end of period. Hence, whether or not we should regard the dependent variable as censored, may be subject to discussion.

The problem of censored variables was first recognised by James Tobin (1958), who showed that the use of ordinary least squares (OLS) for models with such variables results in biased and inconsistent estimates. Instead of ordinary least squares, one should therefore preferably employ a two-stage estimation process developed by James Heckman, that yields unbiased and consistent estimates of the parameters. At the first stage, the so-called *hazard rate* is estimated by utilising a probit model. At the second stage, one proceeds by adding the hazard rate as an additional explanatory variable to the original model. The "extended model" is then estimated by OLS.[15]

Considering the difference between one publication and zero publications, i.e., the difference between publishing activity and no publishing activity, as more serious than that between, e.g., 3 and 4 publications, we found that the existence of a censored variable problem could not be ruled out. Consequently, the two-stage estimation process devised by Heckman was employed. But in addition we also conducted a normal OLS estimation, the regression results from which are reported in table A5 in the appendix. As these results were not found to differ substantially from those obtained using the two-stage estimator, they will not be commented on further.[16]

---

[15] For a more detailed description of the two-stage estimator developed by Heckman, Pindyck and Rubinfeld (1991) may be consulted.

[16] With a few exceptions, type of estimation procedure did neither affect the signs of the coefficients, nor the effect exerted by the different variables in terms of significance. At a first sight, there seems to be a discrepancy with respect to the variable UNI since, unlike in table 3, in table A5 the variable

Unfortunately, the two stage estimator involves heteroscedasticity, implying that the t-tests are biased. To detect heteroscedasticity and the form of heteroscedasticity, we employed a Breusch-Pagan test. Second, the method of weighted least squares (WLS) was used to correct for the heteroscedasticity. Two alternative functional forms of equations (2) and (3), where heteroscedasticity has been corrected for, are reported below in table 3.[17]

The linear-log form is obviously superior to the ordinary linear form, and improves the fit of the model with more than ten percentage points. With regard to the three first variables in the table, functional form does not influence the effects exerted by the various independent variables in terms of significance. But as for the dummy variable indicating mobility and the variable "faculty size", the linear and linear-log cases generate different results with respect to significance.

The calculated F value does, in all four reported cases, exceed the critical F value at a one percent level of significance. This means that we may reject the null hypothesis that research output is not related to the four (five) variables in the table above. We may, accordingly, assume that the independent variables jointly have an influence on research output that is significantly different from zero.

UNI appears to have an insignificant effect on publications. However, this was found to be due to a multicollinearity problem.

[17] The regression results where heteroscedasticity has not yet been corrected for, are found in table A4 in the appendix.

Table 3 — Regression results explaining research output published in major journals

Independent variable:	Linear Case		Linear-Log Case	
	PUBL	PUBL	PUBL	PUBL
AGE	-.98131	-1.2025	-17.739	-20.833
	(-6.23; -.432)	(-4.75; -.392)	(-4.52; -.328)	(-3.76; -.320)
UNI	-.041402	-.054813	-.69516	-.77660
	(-4.19; -.307)	(-3.35; -.288)	(-3.23; -.241)	(-2.63; -.230)
EMPL	-.099619	-.09689	-2.7511	-2.9135
	(-7.22; -.486)	(-4.66; -.387)	(-6.57; -.451)	(-4.95; -.406)
MOBILITY	-.74459	-1.2226	-.25385	-.63782
	(-1.80; -.137)	(-2.22; -.196)	(-.68; -.052)	(-1.33; -.118)
FSIZE		-.012404		-.90533
		(-.49; -.044)		(-2.14; -.189)
Constant	37.193	44.156	73.517	87.945
	(7.18; .483)	(5.14; .419)	(5.08; .364)	(4.20; .353)
R^2	.43	.37	.48	.45
$\overline{R}^2$	.42	.34	.46	.42
F	25.93	12.37	31.14	16.64
Estimated equation	(2)	(3)	(2)	(3)
Number of observations	175	131	175	131

Note: Numbers in parentheses are t-values and partial correlation coefficients respectively.

Consistent with the a priori suspicion, age turns out to have a significant negative impact on publishing, suggesting that an older faculty member publishes fewer articles in major journals than younger colleagues.[18] In the preceding section we mentioned diminishing returns to academic publishing as one reason for this. Another reason may be that as the faculty member gets older, he faces an increasing number of duties in form of, for instance, administrative work and supervision.

[18] We note that, as long as the level of significance is not explicitly specified, a 5 % level of significance is employed.

Mobility was presumed to have a positive impact on high quality publishing. But the regression results show, in three out of four cases mobility does not have any significant impact on publishing at all; and that in the linear case with a reduced sample where mobility has a significant impact on publishing, it actually has a negative impact. Taking a closer look at the data, our findings are confirmed: only 27.4 % changed jobs during the years that mobility was observed, i.e., within a period of 6 - 11 years. (The length of the period being determined by the year the person completed his Ph.D.) In other words, academicians do not appear as remarkably mobile. Second, approximately 54 % of those who did change their jobs, did not publish at all in major journals during the years 1979-83. Third, although diverse experience from academic as well as non-academic institutions is reckoned as important for the improvement and further development of skills, we find no evidence that a diverse background increases research output. 4.6 % of the academicians in the sample worked in non-academic institutions before they entered academia at some time before 1979. Only 12.5 % of these people did actually publish in a major journal within the period 1979-83.

Obviously, our results on mobility have little in common with previous evidence, organisational psychology theories or prevailing views. One reason for the unexpected results may be that the period we observe is too short. But more important are, perhaps, the aspects related to tenure decisions, which were briefly approached in the previous section. Since all tenure decisions are likely to have been made within the observed interval, mobility might be the result of a person failing to get tenure. Furthermore, before tenure decisions are made, people may have few incentives to change jobs. Hence, the negative impact of mobility on publishing behaviour, as well as the high degree of immobility observed here, may be explained by issues related to the phenomenon "tenure". However, on examining the data on mobility, we found

that only 37.5 % of the "mobile" persons are likely to have changed their place of employment "involuntarily". In other words, because they did not receive tenure. Moreover, left with the "voluntary" mobility, it appeared that a "mobile" person on average published less than an "immobile" person. Thus, explaining away the negative impact caused by mobility becomes difficult. One should not neglect, however, that there are several kinds of mobility, and we have only surveyed one of these. Other kinds of mobility that may affect research productivity positively are, for instance, temporary mobility, like visiting scholar/professor arrangements, as well as part time posts.

Turning to the impact of faculty size, we find that in the linear-log case, faculty size has a significant negative effect on research output — suggesting that the more members a department has, the less each faculty member is likely to publish in major journals. At first, this result appears as surprising, since the simple correlation matrix and our intuition made us presume faculty size to have a positive influence on research productivity. One would expect that a larger number of colleagues, and possibly a larger range of fields covered by the faculty staff, would have a positive impact on a person's publication record. In order to explain the negative influence of a larger faculty, one may argue that faculty size may be correlated with variables such as number of students, teaching load, secretarial support, and research assistance — all of which are likely to be decisive for research productivity. If, for instance, a large faculty is equivalent to a large number of students, a heavy teaching load, and little secretarial support and research assistance on average, the results obtained here seem less surprising.

Consistent with earlier evidence, our results suggest that the research environment of a department is a reliable predictor of the research acumen and publishing behaviour of its graduates.[19] But comparing the variable "university of graduation" (UNI) with the variable "university of employment" (EMPL), the partial correlation coefficients show that university of graduation is relatively less important than university of employment. In other words, a larger proportion of the variation in research output is explained by where people work than by where they completed their Ph.D.s. Calculating the elasticities at mean for the two variables, this result is confirmed: place of employment has a much more substantial influence on people's publication record than do their university of graduation.[20]

The departments of economics where the economists were working from 1979-81 (EMPL), turn out to a clear indicator of research output published in major journals from 1979-83. The results suggest that the higher the rank of the department where an economist is employed, the higher the productivity and quality of his research output. Bearing in mind that the universities are ranked according to mean number of citations, it appears that the "better" his average colleague, the more successful is an economist, and vice versa.[21] But although the results suggest that there is a relationship between an economist's research output and the evaluation of his place of employment, one serious question remains to be answered. *Why* is the rank of the department where a person is employed, suggestive of his publication record? — A question referring to

---

[19] See e.g. Hogan (1986) and Niemi (1975).

[20] The elasticities at mean are -.60 for UNI and - 4.37 for EMPL in the linear case where heteroscedasticity has been corrected for. The calculation is straight forward: the elasticity at mean is defined as $\frac{\partial x}{\partial y}\frac{\bar{y}}{\bar{x}}$. Table 1 provides mean of the variables, while the estimated coefficients are taken from table 3.

[21] We have previously mentioned, that Davis and Papanek also offer a ranking list based on total number of citations attributed to a department's faculty (rank *a*). Comparing the two rankings, they prove to be almost perfectly correlated. We may therefore draw the conclusion that, not only the average individual expertise of a person's colleagues, but also the overall "pool" of current expertise at the department, is indicative of an economist's productivity.

the two hypotheses presented in the introduction of this paper and mentioned in section 3.1. One plausible reason, is that good departments mainly employ economists reckoned as competent, while departments performing less impressively attract less productive people. Another, and complementary, reason may be that a good department offers a stimulating and encouraging environment for its employees, i.e., is capable of "producing" good researchers. Accordingly, if two academicians can be considered ex ante in position of equal acumen, the quality of the research environment at the department where they become employed, will be decisive for how each of them performs.

In order to elucidate the relevance of the latter explanation of why departmental rankings are suggestive of faculty members' publication record, we aim to find a characteristic of economists early in their career, that can be considered a reliable indicator of their qualifications and chances of succeeding as researchers. There is a number of such characteristics, and which one is the better, will naturally be subject to discussion. In addition, it is naive to believe that one characteristic alone is decisive for a person's academic success.

Both the university where a person graduates, and his early publications, occur to us as possible indicators of research qualifications. In order to consider early publications as an indicator, a new variable is defined: EPUBL depicts an economist's number of publications in 50 major journals during the interval of three years starting with the year he completed his Ph.D. (For information about how data were recorded see section 2.) We proceeded by estimating the equation (4), employing the two-stage estimator devised by Heckman.

$$PUBL_i = \alpha + \beta_1 \text{ UNI} + \beta_2 \text{ EMPL} + \beta_3 EPUBL \qquad (4)$$

A Breusch-Pagan test was utilised to detect heteroscedasticity. However, the test showed that the error variances varied directly with the independent variable EPUBL. As this variable in numerous cases takes on the value zero, correcting for heteroscedasticity using weighted least squares (WLS) turned out to be impossible. Thus, heteroscedasticity may be responsible for biased *t*-values in table 4, where the regresssion results from the estimation of equation (4) are reported.

One might object that it should not be necessary to include both EPUBL and UNI in the model, as it seems likely that university of graduation may affect the number of early publications. However, estimating EPUBL as a function of UNI showed that only 10 % of the deviations in the early publication records could be explained by the variable "university of graduation".

Table 4

Independent variable	Estimated coefficient	*t*-value	Partial corr. coefficient
UNI	-.0031978	-.40	-.03
EMPL	-.022069	-2.46	-.185
EPUBL	.72013	-2.61	.197
Constant	2.8731	3.37	.250
R^2 =.38	$\overline{R}^2$ =.36	F = 25.62	
Number of observations: 175			

The results suggest that early publications are a reliable indicator of later publishing performance, while the university of graduation is not. However, the latter result contradicts the evidence obtained earlier in the paper (see table 3). The insignificant impact of university of graduation may be explained by the presence of heteroscedasticity.

In order to test the hypothesis that good departments "produce" good people, we consider the partial regression coefficient and partial correlation coefficients related to the variable "university of employment" (EMPL).[22] From the definitions of partial regression coefficient and partial correlation coefficient, it follows that if persons with the same acumen (measured in terms of university of graduation and/or early publications) get employed at differently ranked universities, place of employment may explain a significant proportion of variation in publishing.

The hypothesis that a good department not only attracts good people, but also "produces" good people cannot, in other words, be rejected. Even if a group of people ex ante seems to have the same potential for succeeding as academicians, their success as such may come to differ substantially. The results here indicate that among other possible determinants of success, department of employment has a significant impact on academic productivity and success. Hence, working in a good department does not only entail prestige and possibly a higher salary, it may also mean:

[22] A partial regression coefficient measures the change in the mean value of the dependent variable per unit change in the respective independent variable, holding all other independent variables constant. A partial correlation coefficient measures the degree of association between the dependent variable and an independent variable, holding all other independent variables constant.

- access to an active and stimulating environment with highly qualified and productive colleagues and favourable working conditions;[23]
- proximity to a pool of knowledge;
- a rapid transmission of new ideas and forthcomings within different fields of economics.

All of which enhance an academician's productivity.

An alternative interpretation of the results reported above is that, utilising other characteristics than just university of graduation and early publications, good universities are better at predicting academicians' success. Being better at predicting, however, does not exclude the possibility that such universities might also offer a more stimulating and favourable environment for their researchers, thereby enhancing their productivity.

Thus, one might presume that the observed variation in publishing may partly be ascribed to good universities being able to attract good academicians, partly to the same universities facilitating efficient "academic clusters", and partly to these universities' better predicting abilities.

[23] The interpretation of a good department is based on the fact that the ranking list we employ is based on average number of citations. A difficulty with this interpretation relates to the "superstar" who pulls up an entire departmental average. But as long as the faculty of a department counts a certain number of members (about 10 according to preliminary studies), one "superstar" is not enough to pull up the departmental average. In the sample employed here, only a few departments (13 %) have less than ten members.

4. Concluding remarks

Observing the results obtained in the preceding sections, it appears that compared with other indicators of research productivity, location (place of employment) contributes significantly to the explanation of deviations in research output. We have reason to believe that location matters — not only in terms of prestige and income — but also due to spillover effects. The results indicate that academic milieus resemble the "industrial world" in the sense that the phenomenon localised externalities/spillovers is present in both places. These spillovers can be exploited when competent academicians behave just like firms — and agglomerate. In the same way as geographically concentrated firms may become more competitive than "isolated" competitors, the "academic clusters" may enhance the research productivity of their incumbents, and contribute to their academic success.

Appendix

Table A1 — 50 top journals
(Employed when recording the economists' "early publications".)

Amer. Econ. Rev.	J. Ind. Econ.
Amer. Econ. Rev. Papers & Proc.	J. Law. Econ.
Amer. J. Agri. Econ.	J. of Polit. Econ.
Bus. Hist. Rev.	J. Reg. Sci.
Econ. Devel. Cult. Change	J. Roy. Statist Soc. Ser. A—Gen.
Econ. Hist. Rev.	Kyklos
Econ. Inquiry (Western Econ. J.)	Land Econ.
Econ. J.	Lloyd's Bank Rev.
Econ. of Planning	Manch. Sch. Econ. Soc. Stud,
Econ. Record	Nat. Tax J.
Econometrica	Oxford Econ. Pap.
Economia Internazionale	Problems of Econ.
Economica	Public Finance
Explorations Econ. Hist.	Quart. J. Econ.
German Econ. Rev.	Quart. Rev. Econ. Bus.
Ind. Lab. Relat. Rev.	Rev. Econ. Statist.
Ind. Rel.	Rev. Econ. Studies
Indian Econ. J.	Rev. Social Econ.
Int. Econ. Rev.	Scand. J. Econ.
Int. Monet. Fund Staff Pap.	Scottish J. Polit. Econ.
J. Amer. Statist. Assoc.	Soc. Econ. Stud.
J. Bus.	South African J. Econ.
J. Develop. Stud.	Southern Econ. J.
J. Econ. Hist.	Weltwirtsch. Arch.
J. Finance	Yale Essays on Econ.

Source: Laband and Piette (1994): Table 1 — Rankings based on citations per article, rank *a* — 1970 citations to articles published 1965-1969.

Table A2 — 50 top journals
(Employed when recording publications in the period 1979-83.)

Amer. Econ. Rev.	J. Ind. Econ.
Amer. Econ. Rev. Papers & Proc.	J. Int. Econ.
Amer. J. Agri. Econ.	J. Law Econ.
Bell J. Econ (Rand J. Econ)	J. Legal Stud.
Brit. J. of Ind. Rel.	J. Math. Econ.
Brookings Pap. Econ. Act.	J. Monet Econ.
Demography	J. Money Credit Banking
Econ Geogr.	J. of Econ. Lit.
Econ J.	J. Polit. Econ.
Econ. Inquiry	J. Pub. Econ.
Econometrica	J. Reg. Sci.
Economica	J. Roy. Statist. Soc. Ser. A—Gen.
Ind. Lab. Relat. Rev.	J. Roy. Statist. Soc. Ser. B—Meth.
Inquiry	J. Urban Econ.
Int. Econ. Rev.	Michigan Law Rev.
J. Acc. Res.	Nat. Tax J.
J. Amer. Statist. Assoc.	Oxford Econ. Pap.
J. Bus.	Population Stud.
J. Cons. Res.	Public Pol.
J. Econ. Hist.	Quart. J. Econ.
J. Econ. Theory	Reg. Stud.
J. Econometrics	Rev. Econ. Statist.
J. Finan. Econ.	Rev. Econ. Stud.
J. Finance	Urban Stud.
J. Hum. Res.	Yale Law J.

Source: Liebowitz and Palmer (1984): Table 1 — Rankings based on 1980 citations to articles published 1975-1979.

Table A3 — University rankings

University	I	II	Faculty Size	University	I	II	Faculty size
Chicago	1	1	37	Florida state	62	61	24
MIT	2	3	42	Oregon	63	70	20
Harvard	3	2	60	Case Western	64	88	11
Rochester	4	17	16	Connecticut (at Storrs)	65	58	30
Princeton	5	5	54	Wayne State	66	63	25
Stanford	6	4	67	Georgia	67	65	24
Penn	7	7	43	Boston coll	68	69	22
UCLA	8	12	32	Penn State	69	60	27
Columbia	9	9	41	Wyoming	70	83	14
Minnesota (Twin Cities)	10	18	24	SUNY-Stony Brook	71	74	19
NYU	11	13	32	Ohio U.	72	81	16
Yale	12	6	57	Iowa State	73	48	49
Northwestern	13	14	32	Purdue	74	34	79
Wis-Madison	14	8	41	Nebraska Lincoln	75	82	20
UC-Berkeley	15	10	47	American	76	89	17
Duke	16	23	16	LSU (Louisiana State U)	77	90	16
USC (Southern Calif.)	17	16	37	Washington state	78	76	25
Johns Hopkins	18	28	18	Utah	79	78	24
VPI	19	19	28	Georgetown	80	75	30
Michigan	20	11	55	Tennessee	81	79	26
Cal Tech	21	26	23	N. Carolina State Raleigh	82	54	62
Washington	22	20	30	Ill.-Chigaco Circle	83	91	19
Mass.-Amherst	23	24	26	Notre Dame	84	85	23
Cornell	24	27	26	St. Louis	85	104	11
UC-Riverside	25	80	5	Kentucky	86	86	26
Brown	26	39	20	South Carolina	87	77	32
Maryland	27	15	62	N. Illinois	88	92	22
New School	28	42	19	Iowa	89	84	27
Claremont	29	29	29	Temple	90	73	41
Virginia	30	30	29	SUNY-Albany	91	94	24
Texas A&M	31	44	20	Oklahoma State	92	68	52
UC-Santa Barbara	32	35	26	Missouri-Columbia	93	71	54
Arizona	33	37	26	Colorado state	94	66	60
Washington-St. Louis	34	38	26	Georgia state	95	95	27
Carnegie Mellon	35	47	22	Tufts	96	102	16
SUNY-Buffalo	36	59	14	Montana state	97	97	24
Colorado boulder	37	36	30	Illinois State	98	105	16
Ill. -Urbana	38	21	52	Miami	99	117	6
George Washington	39	25	49	Oklahoma	100	106	14
Texas Austin	40	32	36	Bryn Mwar	101	119	4
Houston	41	56	19	Texas Tech	102	103	18
Vanderbilt	42	41	31	Texas Dallas	103	96	40
Boston	43	31	44	Rensselaer	104	87	40
Rice	44	64	16	Catholic	105	98	25
Indiana	45	51	25	Cincinnati	106	99	22
Ohio State	46	40	36	Kansas	107	100	23
UC-Davis	47	57	21	Rutgers Newark	108	107	17

(continued)

SUNY-Binghamton	48	49	28	So. Illinois	109	109	15
Tulane	49	93	7	W. Virginia	110	101	24
New Mexico	50	53	25	Clark	111	116	9
CUNY	51	43	34	New Hampshire	112	110	18
SMU (Southern Methodist)	52	62	19	Mississippi	113	114	17
UC-San Diego	53	33	47	Northeastern	114	112	22
Florida	54	50	30	Utah State	115	108	28
Wis.-Milwaukee	55	72	16	Kansas state	116	115	18
Michigan State	56	22	76	Howard	117	113	23
Arizona State	57	55	29	Arkansas	119	111	33
North Carolina	58	46	38	Rhode Island	119	120	12
Rutgers-New Brunswick	59	52	32	Lehigh	120	118	18
Pittsburgh	60	45	40	Fordham	121	121	18
Syracuse	61	67	20	Alabama	122	122	9

Sources:

I: Davis and Papanek (1984): Table 1 — University ranking, by total and per capita citations (1978 and 1981)", Rank *b* ("Rank by mean number").

II: Davis and Papanek (1984): Table 1 — University ranking, by total and per capita citations (1978 and 1981), Rank *a* ("Rank by total number of citations").

Table A4 — Regression results explaining research output published in major journals without correcting for heteroscedasticity

Independent variable:	Linear Case		Linear-Log Case	
	PUBL	PUBL	PUBL	PUBL
AGE	-.92041	-1.1199	-17.545	-20.718
	(-6.35; -.439)	(-4.84; -.399)	(-4.57; -.332)	(-3.81; -.324)
UNI	-.038321	-.050347	-.68996	-.77587
	(-4.42; -.322)	(-3.44; -.295)	(-3.30; -.246)	(-2.69; -.235)
EMPL	-.094201	-.086827	-2.7581	-2.9441
	(-7.48; -.499)	(-4.88; -.401)	(-6.68; -.458)	(-5.07; -.414)
MOBILITY	-.66958	-1.1164	-.24066	-.62760
	(-1.76; -.134)	(-2.16; -.190)	(-.66; -.051)	(-1.33; -.118)
FSIZE		-.015182		-.89782
		(-.63; -.057)		(-2.16; -.191)
Constant	35.252	41.731	72.837	87.558
	(7.31; .490)	(5.24; .426)	(5.15; .368)	(4.26; .357)
R^2	.44	.38	.48	.45
$\overline{R}^2$	.43	.35	.47	.42
F	26.86	12.92	31.58	16.89
Estimated equation	(2)	(3)	(2)	(3)
Number of observations	175	131	175	131

Note: Numbers in parentheses are t-values and partial correlation coefficients respectively.

Table A5 — Regression results explaining research output published in major journals when a standard OLS procedure is used to estimate the model.

Independent variable:	Linear Case		Linear-Log Case	
	PUBL	PUBL	PUBL	PUBL
AGE	-.13978	-.19540	-5.0614	-7.4029
	(-3.22; -.240)	(-3.24; -.279)	(-3.33; -.247)	(-3.54; -.302)
UNI	-.0029569	-.0041717	-.14736	-.20351
	(-.46; -.035)	(-.42; -.037)	(-1.01; -.077)	(-1.05; -.093)
EMPL	-.027660	-.020966	-1.4512	-1.5665
	(-6.02; -.419)	(-2.53; -.221)	(-7.78; -.512)	(-5.95; -.470)
MOBILITY	-.42007	-.63540	-.32468	-.54085
	(-1.02; -.078)	(-1.19; -.106)	(-.89; -.066)	(-1.12; -.100)
FSIZE		.020300		-.56772
		(.86; .076)		(-1.40; -.124)
Constant	9.3166	10.340	26.242	36.986
	(6.256; .433)	(4.20; .351)	(5.00; .358)	(4.81; .395)
R^2	.34	.30	.45	.42
$\overline{R}^2$	.32	.27	.43	.40
F	21.80	10.74	34.08	18.01
Estimated equation	(2)	(3)	(2)	(3)
Number of observations	175	131	175	131

Note: Numbers in parentheses are t-values and partial correlation coefficients respectively.

References

American Economic Association (1993): AEA-Directory of members (on-line database)

Arrow, Kenneth J. (1962): "Economic welfare and the allocation of resources for inventions", in Nelson, R. R. (ed.): *The rate and direction of inventive activity*, Princeton: Princeton University Press for the NBER, 1962

Blair, Dudley W., Rex L. Cottle and Myles S.: "Faculty ratings of major economics departments by citations: an extension", *American Economic Review* 76 (1986): 264-267

Caballero, Ricardo J. and Richard K. Lyons (1990): "Internal versus external economies in the European industry", *European Economic Review* 34 (1990): 805-830

Davis, Paul and Gustav F. Papanek (1984): "Faculty ratings of major economics departments by citations", *American Economic Review* 74 (1984): 225-230

Gans, Joshua S. and George B. Shepherd (1994): "How are the mighty fallen: Rejected classic articles by leading economist", *Journal of Economic Perspectives* 8 (1994): 165-179

Gibbons, Jean D. and Mary Fish (1991): "Rankings of economics faculties and representation on editorial boards of top journals", *Journal of Economic Education* 22 (1991): 361-372

Golden, John, Fred V.Carstensen, Paul Weiner and Steve Kane (1988): "Modelling the ratings game: Publication performance, departmental characteristics, and graduate faculty ratings in economics", *Quarterly Review of Economics and Business* 28 (1988): 101-109

Graves, Philip E., James R. Marchand and Randall Thompson (1982): "Economics departmental rankings: Research incentives, constraints, and efficiency", *American Economic Review* 72 (1982): 1131-1141

Graves, Philip E., James R. Marchand and Randall Thompson (1984): "Economics departmental rankings: Reply and errata", *American Economic Review* 74 (1984): 834-836

Grossman, Gene M. and Elhanan Helpman (1992): *Innovation and growth in the global economy*, Cambridge, Massachusetts: The MIT Press, 1992

Hamermesh, Daniel S. (1992): "The young economist's guide to professional etiquette", *Journal of Economic Perspectives* 6 (1992): 169-179

Hamermesh, Daniel S. (1994): "Facts and myths about refereeing", *Journal of Economic Perspectives* 8 (1994): 153-163

Heckman, James (1979): "Sample selection bias as a specification error", *Econometrica* 47 (1979): 153-161

Heien, Dale and Cathy Roheim Wessells: "Demand systems estimation with microdata: a censored regression approach", *Journal of Business and Economic Statistics* 8 (1990): 365-371

Hogan, Timothy D. (1986): "The publishing performance of U.S. Ph.D. programs in economics during the 1970s.", *Journal of Human Resources* 21 (1986): 216-229

Hogan, Timothy D. (1984): "Economics departmental rankings: Comment", *American Economic Review* 74 (1984): 827-833

Kennedy, Peter (1993): *A guide to econometrics*, 3 rd ed., Cambridge, Massachusetts: The MIT Press, 1993

Krugman, Paul R. (1991): *Geography and trade*, Cambridge, Massachusetts: The MIT Press, 1991

Krugman, Paul R. and Anthony J.Venables (1993): "Integration, specialization, and adjustment", NBER Working Paper No. 4559, 1993

Kyvik, Svein and Olaf Tvede (eds.) (1994): "Mobilitetsmønstre blant norske forskere", Utredningsinstituttet for forskning og høyere utdanning rapport 14/94, 1994

Laband, David N. and Michael J. Piette (1994): "The relative impacts of economics journals: 1970-1990", *Journal of Economic Literature* 32 (1994): 640-666

Liebowitz, Stanley J. and John C. Palmer (1984): "Assessing the relative impacts of economics journals", *Journal of Economic Literature* 22 (1984): 77-88

London school of Economics and Political Science (1995): *London School of Economics Centenary Review,* 1995

Marshall, Alfred (1920): *Principles of economics*, London: Macmillan, 1920

Marshall, Howard D. (1964): *The mobility of college faculties*, New York: Pageant Press, 1964

Niemi, Albert W. Jr. (1975): "Journal publication performance during 1970-74: The relative output of southern economics departments", *Southern Economic Journal* 42 (1975): 97-106

Piette, Michael. J. and Kevin L. Ross (1992): "An analysis of the determinants of co-authorship in economics", *Journal of Economic Education* 23 (1992): 277-283

Pindyck, Robert S. and Daniel L. Rubinfeld: *Econometric models and economic forecasts*, New York: McGraw-Hill, 1991

Porter, Michael E. (1990): *The competitive advantage of nations,* London and Basingstoke: Macmillan, 1990

Romer, Paul M. (1986): "Increasing returns and long-run growth", *Journal of Political Economy* 94 (1986): 1002-1037

Schweser, Carl (1983): "The economics of academic publishing", *Journal of Economic Education* 14 (1983): 60-64

Skeels, Jack W. and R. P. Fairbanks (1968): "Publish or perish: An analysis of the mobility of publishing and non-publishing economists", *Southern Economic Journal* 35 (1968): 17-25

The Economist, July 30th, 1994

The Research Council of Norway (ed.) (1995): *Strategi for næringsrettet forskning og utvikling*, 1995

Tobin, James (1958): "Estimation of relationships for limited dependent variables", *Econometrica 26 (1958)*: 24-36

Tuckman, Howard P. and Jack Leahey (1975): "What is an article worth?", *Journal of Political Economy* 83 (1975): 951-967

Zivney, Terry L. and William J. Bertin (1992): "Publish or Perish: What the competition is *really* doing", *The Journal of Finance* 47 (1992): 295-329