

13 October 2008



KUNGL.  
VETENSKAPSAKADEMIEN  
THE ROYAL SWEDISH ACADEMY OF SCIENCES



Scientific background on  
the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2008

# Trade and Geography – Economies of Scale, Differentiated Products and Transport Costs

Compiled by the Prize Committee of the Royal Swedish Academy of Sciences

## Introduction

Over the centuries, international trade and the location of economic activity have been at the forefront of economic thought. Even today, free trade, globalization, and urbanization remain as commonplace topics in the popular debate as well as in scholarly analyses. Traditionally, trade theory and economic geography evolved as separate subfields of economics. More recently, however, they have converged become more and more united through new theoretical insights, which emphasize that the same basic forces simultaneously determine specialization across countries for a given international distribution of factors of production (trade theory) and the long-run location of those factors across countries (economic geography).

As of the mid-1970s, trade theory was based on the notion of comparative advantage. Countries were assumed to trade with each other because of differences in some respect – either in terms of technology, as assumed by David Ricardo in the early 19<sup>th</sup> century, or in terms of factor endowments, according to the Heckscher-Ohlin theory developed in the 1920s. The latter was explicated by Bertil Ohlin in his 1933 monograph *Interregional and International Trade*; Ohlin was awarded the 1977 Economics Prize for his contributions to trade theory.

These theories provided good explanations of the trade patterns in the first half of the 20<sup>th</sup> century. But as many researchers began to observe, comparative advantage seemed less relevant in the modern world. Today, most trade takes place between countries with similar technologies and similar factor proportions; quite similar goods are often both exported and imported by the same country. At least among the richer countries, *intra-industry trade* – whereby, for instance, a country both exports and imports textiles – came to dominate relative to *inter-industry trade* – whereby, for instance, a country exports textiles and imports agricultural products. Under such circumstances, how could intra-industry trade be explained? The traditional view, that a given country would have a comparative advantage in terms of technology or factor endowments when producing a particular type of textile, seems far-fetched as an explanation.

Many trade theorists came up with interpretations of the observed patterns of intra-industry trade by referring to economies of scale. In an influential book, Grubel and Lloyd (1975)

documented the large amount of intra-industry trade and argued that it could be explained by economies of scale. If the average cost of producing a given good (for instance, a particular make of car) would decline with total production, then it could be optimal to split up production so that countries specialize in different makes of cars. Such specialization would make sense even without differences in factor proportions and technology. This idea was however not entirely new in 1975; in fact, the role of economies of scale in explaining trade patterns already been recognized by Ohlin:

"[T]he advantages of producing a large quantity of a single commodity instead of a little of all commodities must lead to interregional trade ... insofar as the market for some articles within each region is not large enough to permit the most efficient scale of production, division of trade and labor will be profitable. Each region will specialize on some of these articles and exchange them for the rest ... The tendency toward specialization because of differences in factor endowments is reinforced by the advantages of large-scale production. The location of an industry in one region and not in another might simply be due to chance ... Thus, all interregional trade, whether due to the one cause or the other, might be regarded as a substitute for geographical mobility of productive factors."<sup>1</sup>

It was not until the late 1970s, with the development of what is now known as the *new trade theory*, that these insights were integrated into a stringent and unified theoretical framework. Such a framework is a prerequisite for systematic empirical work, in turn necessary for studying the relative importance of different determinants of trade, as well as for systematic evaluation of policy proposals.

In the field of economic geography, the key questions concern migration flows of individuals and firms across the geographic landscape, how urban agglomerations arise, and how cities themselves are spatially organized (urban economics). Here as well, it had long been recognized that economies of scale are decisive role for the location of economic activity. As of the 1950s, there was a substantial literature on the effects of the trade-off between increasing returns in production and costs of transportation on agglomeration and the growth of cities (Harris, 1954; Myrdal, 1957; Hirschman, 1958; Pred, 1966). But as in the field of trade theory, these insights were not supported by well-articulated models – especially models that allowed a general-equilibrium analysis, where the location of both consumers and firms was explained in the model.

---

<sup>1</sup> Scattered phrases from Chapter III of Ohlin (1933), quoted from Krugman (1999).

By the late 1980s, researchers had begun to integrate economies of scale into general equilibrium models of location and trade, thereby giving precision to the verbal analyses of earlier researchers and adding important new insights. In the resulting work, now commonly known as the *new economic geography*, economic geographers made use of the new tools, along with economists who took a renewed interest in the field. Several researchers took part in these developments, but the most influential contributions were made by Paul Krugman.

Krugman has published a large number of important articles and monographs in the fields of both trade and geography. In particular, he made the initial key contributions. He wrote the first article in the trade theory, soon followed by another influential paper that extended his initial analysis (Krugman, 1979a and 1980). Further, Krugman (1991a) is commonly viewed as the starting point of new economic geography. In fact, the seeds of the new economic geography can already be found in his 1979 (a) article which, in its final section, argues that patterns of migration can be analyzed within the same framework as the new trade theory. While this article had an immediate impact on the trade literature, it would take more than ten years for the final section, on migration and agglomeration, to have an influence on the geography literature – kindled by Krugman himself, in the 1991 (a) paper.

In what follows, we discuss Krugman's contributions to trade and geography. We begin by considering the key common elements – economies of scale, monopoly power, and demand for variety – of both theories, using the setting of Krugman's 1979 (a) paper. We then examine, separately and in more detail, his work on trade and geography, with special emphasis in each case on how Krugman's research transformed the literature.

### **The Basic Model: Economies of Scale and Monopoly Power**

In the late 1970s, several researchers – Krugman (1979a, 1980), Dixit and Norman (1980, Chapter 9) and Lancaster (1980) – independently formalized the idea that economies of scale and imperfect competition can give rise to trade even in the absence of comparative advantage.<sup>2</sup> In related contributions, Ethier (1979, 1982) developed models of intra-industry trade based on economies of scale in intermediate rather than final goods. It was Paul

---

<sup>2</sup> Chapter 9 in Dixit and Norman's book builds on unpublished notes by Norman from 1976 and 1978.

Krugman who most clearly and forcefully articulated the revolutionary nature of this new approach for the theory of international trade. His short paper in the *Journal of International Economics*, entitled “Increasing Returns, Monopolistic Competition and International Trade” (1979a), is twofold. It contains not only a new trade theory that allows us to explain observed patterns of intra-industry trade, but also the seeds of a new economic geography where the location of production factors and economic activity can be stringently analyzed within the framework of a general-equilibrium model. Remarkably, the paper achieves all of this in only ten pages, and in a very simple and transparent fashion. The model is extremely simple. There is only one production factor, and returns to scale are represented by a linear cost function with fixed costs. But due to its simplicity, it illustrates the key mechanisms in a particularly clear way.

The central feature in Krugman’s approach is economies of scale that are internal to the firm, i.e., the firm itself can reduce its own average cost by expanding production. Under such conditions, markets cannot be perfectly competitive. Models of imperfect competition had often been shunned in trade theory because of their analytical complexity. But Krugman made use of a recent model of monopolistic competition due to Dixit and Stiglitz (1977) that turned out to be well suited for the analysis of trade.<sup>3</sup> In that spirit, he assumed that there are  $n$  different goods, and that consumers have a taste for variety that can be expressed by the following utility function:

$$U = \sum_{i=1}^n v(c_i), \quad (1)$$

where  $c_i$  is consumption of the  $i$ th good and where  $v(\cdot)$  is an increasing, concave function. Concavity is crucial as it expresses the taste for variety. Absent price differences between goods, it implies that the consumer would spread his resources evenly across as many goods as possible rather than only consume one good. In the following, we assume that  $v(\cdot)$  takes a particular form:

---

<sup>3</sup> Trade models with external economies of scale had earlier been developed by Matthews (1949), Kemp (1964), Melvin (1969), Negishi (1969) and Chipman (1970).

$$v(c_i) = \frac{c_i^\gamma}{\gamma}, \quad 0 < \gamma < 1. \quad (2)$$

Although this functional form was not used in Krugman's original 1979 (a) paper, it has subsequently been widely used (not least by Krugman himself) owing to the simple and elegant analytical solutions that it provides.<sup>4</sup> In the following, we refer to preferences represented by (1) and (2) as "Dixit-Stiglitz preferences".

Krugman assumes further that there is only one factor of production, namely labor (thus giving all countries identical factor proportions by definition) and that all goods are produced with identical cost functions. He represents increasing returns to scale by assuming that, for any good  $i$  produced, the labor requirement  $l_i$  is given by

$$l_i = \alpha + \beta x_i, \quad \alpha, \beta > 0, \quad (3)$$

where  $x_i$  is the output of good  $i$ , and  $\alpha$  is a fixed cost. When all markets clear and there is free entry then all monopolistically competitive firms will have zero profits in equilibrium. The model can be solved for the three unknowns: the price of each good relative to labor,  $p_i/w$  (where  $w$  is the wage rate), the output of each good,  $x_i$ , and the number of goods produced,  $n$ .

Let us first discuss some aspects of the monopolistic equilibrium of this economy and then introduce a two-country version of the model from the perspective of trade and geography. Consumers maximize their utility subject to their budget constraint ( $\sum_{i=1}^n p_i c_i = w$ ). This leads to the first-order condition  $v'(c_i)/p_i = v'(c_j)/p_j = \lambda$  for any two goods  $i$  and  $j$ , where  $\lambda$  denotes the shadow value of income: the consumer's marginal rate of substitution between the goods should equal the relative price. We then obtain the individual's demand function for good  $i$  as  $c(p_i) = (\lambda p_i)^{1/(\gamma-1)}$ . Indirect utility can easily be computed and is proportional to  $(w/\bar{p})^\gamma$ , where  $\bar{p} = (\sum_i p_i^{\gamma/(\gamma-1)})^{(\gamma-1)/\gamma}$  is a price index across goods. We can thus think of  $w/\bar{p}$  as the real wage in this economy.

---

<sup>4</sup> In his 1979 paper, Krugman assumes instead that  $v$  is such that the elasticity of demand decreases with income, which leads to slightly different conclusions and makes the exposition somewhat more complicated.

Turning to the monopolistic competitor, profits are maximized by setting the price so that marginal revenue equals marginal cost. With  $L$  consumers, the firm's total revenue equals  $p_i L c(p_i)$  and its total costs are  $(\alpha + \beta L c(p_i))w$ . Moreover, when the monopolist chooses price, the shadow value of income,  $\lambda$ , can be treated as unaffected by the price, since there is a large number of goods. It is easy to verify that when marginal revenue equals marginal cost, the monopolist chooses a constant markup,  $1/\gamma$ , over marginal cost:  $p_i/(\beta w) = 1/\gamma$ . This symmetry across goods implies that the price index in equilibrium will satisfy

$\bar{p} = n^{(\gamma-1)/\gamma} w \beta / \gamma$ . We see that the price index is decreasing in product variety,  $n$ , and thus that the real wage and equilibrium utility are increasing in  $n$ . In this economy, new firms will enter the market, adding product varieties, until profits equal zero. Since profits equal  $p_i x_i - (\alpha + \beta x_i)w$  and all goods are symmetric and produced in equal quantity, this implies that equilibrium output is given by  $x_i = (\alpha / \beta) \gamma / (1 - \gamma) \equiv \bar{x}$ . With this expression for determining the quantity produced of each good, the variety range of goods,  $n$ , is determined by  $L = (\alpha + \beta \bar{x})n$ , since the  $L$  consumer-workers have to allocate their labor supply to  $n$  different goods.

Now that all equilibrium prices and quantities have been determined, we can investigate how the *size* of the economy, which is given by  $L$  (the number of consumer-workers), influences the equilibrium. First, the production of each good,  $\bar{x}$ , is not affected given our assumption on the shape of the utility function, but the number of goods increases (proportionately) with  $L$ .<sup>5</sup> Second, per-capita consumption of each good decreases since  $cL = \bar{x}$ . Third, per-consumer welfare increases with  $L$  since, as we have seen above, real wages increase in product variety. Thus, the larger economy allows more varieties to be produced, and this is the channel through which increasing returns to scale operate here. More generally, and in Krugman's own analysis, increasing returns to scale may also increase the production of each good, thus lowering unit cost on a good-by-good basis. We are now set to examine the possibility of trade between countries.

---

<sup>5</sup> This is due to the specific utility function (2). Krugman's alternative assumption about the elasticity of the  $v(\cdot)$  function implies that the production of each good increases as the economy enlarges.

Assume that there are two countries which are identical in preferences, technology, and size. We can then compare two cases: “autarchy”, where there is no trade due to, for instance, prohibitive transport costs, and another where the countries can trade freely at no cost. With trade, we can regard the size of the economy as the sum of the sizes of the two countries, and we can determine world production and consumption as if there were only one (large) country. Thus, the number of goods produced is now larger; since the countries are identical, the number of goods is twice what it would be under autarchy. Each country specializes in the production of some goods, thereby exploiting the economies of scale inherent in the fixed-cost production function. Precisely which goods are produced where is not determined, however. Trade will amount to half of GDP of each country (in general, if the countries were of different size, trade would be smaller), and consumers will enjoy a larger number of goods than under autarchy. The opening-up of trade will therefore be welfare-enhancing, despite the fact that both countries have identical technologies and factor proportions, since it increases product variety. Under Krugman’s assumption on the utility function (see footnote 4) it would also lower unit costs good by good. The allocation of goods across countries is not determined in the model; it may well happen that good  $i$  is produced in one country while good  $j$  is produced in the other country, even though goods  $i$  and  $j$  are very similar (for instance, two makes of cars). In that sense, the opening-up of trade generates intra-industry trade.

Thus, Krugman (1979a) showed that consumer preferences represented by (1), and production costs represented by (3), can generate trade patterns consistent with real-world data. In the final section of the paper, he discusses the implications of impediments to trade between the two countries when migration of labor is possible. This section contains the precursor of the “new economic geography” work he would pursue later. Using the simple model above, he argues that in the absence of trade, consumer welfare will be highest in the region with the largest labor force. This is so because with Dixit-Stiglitz preferences, as seen above, the real wage  $w/\bar{p}$  depends positively on the number of products  $n$ , which is greater in the region with the largest labor force. There will thus be a tendency for labor to migrate to the region that happens to have the largest labor force, and thereby the greatest variety of products, at the outset. This mechanism gives rise to a cumulative process, resulting in concentration and urbanization. The model thus contains an element of (potentially dramatic) demographic change. Such a change could, however, be mitigated by a number of features that were not present in the 1979 model.



## International Trade

Building on Krugman's analysis, a vast literature has developed exploring the implications of returns to scale and monopolistic competition for trade patterns in richer model settings. We now consider some of these further research developments, including the policy and empirical implications of the new theory.

### *Transport costs and trade: the home-market effect*

An important force behind the growth of trade has been the decrease in transport costs. Yet such costs were conspicuously absent in most trade models for a long time. In a second seminal contribution, Krugman (1980) extended his 1979 model by introducing transportation costs. For analytical convenience, these costs were assumed to be proportional to the quantity of goods shipped to another nation (sometimes referred to as "iceberg costs" in the sense that a fraction of the goods melts away before they reach their destination). This allowed him to accord analytical precision to the *home-market effect* – earlier discussed by Corden (1970) – according to which firms tend to concentrate, i.e., locate more than in proportion to market size, in large markets. This explains why preferences matter for trade patterns, a point raised earlier by Linder (1961). Countries tend to export the goods for which they have a large domestic market.

The intuition for the home-market effect is simple. With both increasing returns and transport costs, there is an incentive to concentrate production of a good close to its largest market. By concentrating production in one place, scale economies can be realized, while by locating near the largest market, transport costs are minimized. The home-market effect provides a demand explanation as to why a country can have an advantage in the production of a specific good. Workers will be better off in the larger economy because of a lower price level, as a smaller fraction of total consumption is burdened by transport costs.

### *Further developments of the new trade theory*

Krugman's 1979 and 1980 papers demonstrated that models based on the assumptions of increasing returns to scale and monopolistic competition can explain important patterns of trade observed in the data. The way in which these new elements interacted with the traditional factor-proportions mechanism remained to be analyzed. Integrated models of inter-

industry trade (based on technology gaps and Heckscher-Ohlin differences in factor proportions) and intra-industry trade in differentiated goods (based on increasing returns to scale and monopolistic competition) were provided by Lancaster (1980), Dixit and Norman (1980), Krugman (1981) and, with greater generality, by Helpman (1981) and Helpman and Krugman (1985). Integration of the new and old trade theory was particularly important as it led to testable predictions about cross-country differences in trade patterns. It has formed the basis for extensive empirical research on bilateral trade flows, thereby allowing researchers to evaluate the relevance of the new theory.

Another fact about trade patterns is that the bulk of intra-industry trade occurs in knowledge-intensive products between highly developed countries, often in industries dominated by multinational corporations. Such a pattern was accounted for by Helpman (1984) and Markusen (1984), who developed theories of international trade in knowledge-intensive production sectors dominated by multinational firms, where such firms appear as the market response to fixed R&D costs.<sup>6</sup> These and other developments were incorporated and extended in Helpman and Krugman's 1985 monograph. This book provides a comprehensive exposition of the new trade theory and develops its implications for a variety of issues. It remains the standard reference in the field.

The new trade theory has profoundly affected the analysis of trade policy. The theory yields predictions about the impact of trade liberalization on trade patterns, the location of output and factor remunerations. It can also be used for welfare analysis. Realistic models easily become too complex to be handled analytically and estimated by econometric methods. Instead, a vast literature of calibrated numerical models has emerged. Early examples include Harris (1984), Dixit (1988) and Baldwin and Krugman (1988). Such models are now routinely used, e.g. by the World Bank in assessing the effects of the WTO rounds of trade liberalization.

### *Empirical relevance*

As we have seen the development of the new trade theory was largely motivated by the inability of existing models to account for observed empirical patterns, such as the dominance

---

<sup>6</sup> These models are precursors to later studies of endogenous growth by Grossman and Helpman (1989, 1991), where innovations are endogenously produced by profit-maximizing firms, and innovations – typically originating in the most highly developed economies – are spreading to the rest of the world through trade.

of intra-industry trade in the trade between developed countries, and the increase in intra-industry trade resulting from trade liberalization (e.g. the EEC agreement from 1959). Krugman's 1979 (a) simple one-factor model was able to explain these observations. By integrating the new model and the traditional factor-proportions theory, it was possible to formulate more specific hypotheses. The integrated model states that (i) the *volume* of trade between any two countries should increase with the difference in relative factor endowments and decrease with the difference in country size; and (ii) the *share* of intra-industry trade in the total trade between two countries should depend negatively on the difference in capital-labor ratios and positively on size dispersion. Taken literally, the integrated model based on Dixit-Stiglitz preferences (2) and a specified production function even gives predictions about the exact functional forms of the relations between bilateral trade flows, factor endowments and country sizes.

The richness of model predictions inspired a lively empirical literature. Early empirical tests by Helpman (1987) were supportive of the new trade theory. Some later studies, however, showed mixed results; see Leamer and Levinsohn (1995) for a review of the empirical literature as of the early 1990s. For instance, Hummels and Levinsohn (1993, 1995) found that the data on trade flows fit the theory almost too well, when they analyzed both trade between developed countries (for which consumer preferences for product variety and trade in differentiated products were reasonable descriptions of reality) and trade between less developed countries (where trade in monopolistically produced, differentiated goods can hardly be of much importance). Their conclusion was that something else, besides the factor endowments emphasized by the "old" theory, and the increasing returns and differentiated products emphasized by the "new" theory, must lie behind a large part of real-world trade flows. More recent studies tend to be in line with the theoretical predictions. Antweiler and Trefler (2002) note that the model expounded in Helpman and Krugman (1985) gives predictions regarding the factor content of exports and imports, depending on the degree of scale economies. They found that allowing for scale economies improved the fit of the model, and that around one third of all industries could be characterized by increasing returns to scale. Similarly, Evenett and Keller (2002) conclude that trade patterns are best explained by a combination of increasing returns to scale and factor proportions.

## Economic Geography

It has long been recognized that factor mobility and trade may act as substitutes for one another. Impediments to trade could lead to factor-price differences that would induce migration of labor and capital. This had already been analyzed in a Heckscher-Ohlin world by Mundell (1957). As mentioned above, the final section of Krugman (1979a) considers factor mobility in a world of differentiated products and monopolistic competition. In the absence of trade, the larger region would offer its inhabitants higher welfare due to a greater variety of products, thereby providing incentives for migration. The incentives for migration would be stronger, the greater the number of people that moved to the larger region and, absent any impediments to migration, the whole population would in equilibrium end up in the region that happened to be largest at the outset. Note that such a process could be triggered solely by the initial size difference, if there were no inherent differences between regions. If the regions differed, e.g. in labor productivity, it would still be possible to end up in an inefficient equilibrium under certain initial conditions. If the region with the lowest exogenous productivity was given a head start by a larger initial market size, then migration might lead to an equilibrium with the entire population concentrated in the region with low productivity.

It would take another twelve years until these ideas were developed into the so-called *core-periphery* model of Krugman (1991a), the starting point of the new economic geography. To assess the far-reaching nature of this model, some background on spatial economics is called for.

### *Spatial economics – some background*

Combining space and competitive equilibrium is a major challenge.<sup>7</sup> Trade theory has traditionally taken the heterogeneity of space as exogenously given and analyzed the trade patterns resulting from differences in factor proportions and technology. Much of regional analysis, starting with Marshall, has recognized that agglomeration of economic activity is driven by economies of scale, while assuming that the scale economies are *external* to the individual firm but internal to the industry or the city, and hence consistent with perfect

---

<sup>7</sup> In fact, Starrett (1978) formulated a spatial impossibility theorem, which states that there cannot exist a spatial equilibrium involving transportation in an Arrow-Debreu economy with homogeneous space and costly transportation. If all activities were perfectly divisible there would exist an equilibrium in which each region operated as an autarchy. In practice, however, this is an uninteresting case, since it implies no trade and no division of labor.

competition. An example would be pure “technological spillovers” between firms in a well-defined area due to transfers of knowledge useful in production. A shortcoming of this approach is that the nature of these external scale economies are typically left rather vague and it is, therefore, hard to measure the externalities in order to test and use the theory empirically. The new economic geography initiated by Krugman broke with this tradition by assuming *internal* economies of scale and imperfect competition. Agglomeration is then driven by *pecuniary externalities* mediated through market prices as a large market allows greater product variety and lower costs. When a household or a firm transacts in a market, there is generally an effect on other agents through impact of the transaction on the price. In those cases where there is some form of market imperfection at the outset, this effect can be viewed as an externality. In Krugman’s work, the presence of scale economies and of monopolistic competition implies a market imperfection, so that the externality can be traced to these fundamental features of the economy.

The forerunning contribution built on external economies of scale is Henderson (1974). Here an industry-specific externality in production causes the marginal cost of a firm to be declining in the level of industry output within an entire city. Equilibrium city size then depends on the trade-off between this externality, which determines the agglomeration gains, and the costs of spatial concentration of activity (such as commuting costs). Cities tend to specialize by industry and those industries in which external scale economies are more substantial tend to be concentrated in larger cities.

A more elaborate model based on Marshallian externalities was later provided by Fujita and Ogawa (1982). Whereas Henderson assumed cities to be monocentric around a central business district (CBD) – as in the classical models of von Thünen (1826) and Alonso (1964) – Fujita and Ogawa solve for land prices, wages and the equilibrium allocation of land to production and housing, while allowing for the possibility of secondary business districts in coexistence with the CBD. A key finding is that cities may undergo drastic structural changes when transport costs and other key parameters change. The impact of the literature which relies on the external-economies-of-scale assumption may have been limited by the lack of explicit micro foundations. Are the externalities due to knowledge spillovers, thin markets for specialized inputs, backward and forward linkages in the production chain, local public goods or to other factors? This is an area of active research; see Duranton and Puga (2004) for an

account of different theoretical approaches and Rosenthal and Strange (2004) for an assessment of the empirical literature.

*The new economic geography – the core-periphery model*

The more recent development of economic geography builds primarily on models of internal, as opposed to external, economies of scale and monopolistic competition. The seminal contribution that launched the new economic geography is the *core-periphery model* developed in Krugman (1991a). Important precursors to Krugman's analysis were published by Abdel-Rahman (1988) and Fujita (1988), who developed models of location within an agglomeration based on Dixit-Stiglitz monopolistic competition and derived equilibrium patterns of location. In these models, however, there is no agricultural sector and no migration across regions.<sup>8</sup>

More than half of the world's population lives in cities. In most countries, a majority of the population lives in a few highly developed urban regions (the core), whereas a minority of the population remains in a mainly agricultural hinterland (the periphery). The model of Krugman (1991a) is aimed at explaining the prevalence of this pattern. It does so by introducing mobility of workers (= consumers) in a model similar to the trade model with transport costs of Krugman (1980). There are two types of products: food, which is assumed homogeneous and produced under constant returns to scale, and manufactured goods, which exist in a large number of varieties, each produced under increasing returns to scale and sold in monopolistically competitive markets. Consumer preferences are expressed by a Cobb-Douglas function in food and the subutility of manufactured goods, the latter given by a Dixit-Stiglitz utility function with a constant elasticity of substitution; cf. eq. (2).

Consider an example of two regions with identical fundamental conditions (preferences and production functions). Goods can be exported from one region to the other. Agricultural goods are traded costlessly (this assumption is relaxed in later works), whereas manufactured goods are subject to (iceberg-type) transport costs. There are two types of individuals: manufacturing workers, who are free to migrate to the region that offers the highest utility

---

<sup>8</sup> The relation between the Fujita and the Krugman models is discussed in Fujita and Krugman (2004). Krugman's analysis also bears some resemblance to Murphy, Shleifer and Vishny (1989a,b), who analyze enlargement of a market through simultaneous expansion of many sectors or the expansion of a leading sector whose income is distributed widely enough. This, in turn, can create a "push" towards industrialization in a developing country by making it profitable to adopt increasing-returns technologies.

level, and peasants, who do not migrate. The problem analyzed is how population and economic activity will be allocated between the two regions. Will there be a concentration of manufacturing into one region? Will the population be split between an industrialized “core” and an agricultural “periphery”?

The model is driven by the location choices of firms and individuals. Firms have an incentive to locate in the larger market to exploit economies of scale in production and to save on transport costs (the home-market effect identified in Krugman, 1980). Individuals have an incentive to move to the larger region, since it offers higher real wages and a larger variety of goods. This tends to increase the difference in size between the markets and strengthen the incentive to migrate both for firms and individuals. Hence, there is an element of circular causality.

To sketch the working of the model and the determination of equilibrium, consider an initial situation with half of the population in two regions. If the regions are identical, this is obviously an equilibrium. But now assume that, by chance, this equilibrium is perturbed by migration, thereby making one region slightly larger than the other. As a result of the initial migration away from equal-sized regions, would there now be incentives for further population divergence? If the home-market effect together with the real-wage effect is strong enough, the initial population perturbation will stimulate further migration to the larger region. This would set in motion a cumulative process, where migration increases the population and the size of the market in the larger region even more, thereby raising the real wage further and thus leading to even more migration etc. Hence, the new equilibrium may be quite different from the original one. In this way, Krugman was able to build a strict model of the process of circular causation discussed much earlier by Myrdal (1957), Hirschman (1958), and others.

However, there are also counteracting forces. If all firms were located in one large region, a single firm that moved to the periphery would become almost like a monopolist in the market for manufacturing goods there. It would sell its product to the local farmers and to its own employees, being disciplined in the goods market by the competition of imports which, in turn, depends on the transport costs. Similarly, it would be disciplined in the labor market by the fact that workers, in order to accept living in the smaller region, must have sufficiently high real wages. Welfare for these manufacturing workers depends on the wage paid by the

firm, the price of the consumption good that this firm produces (which is relatively low, since it is not burdened by transport) and the price they have to pay for imported goods that are burdened by transport costs. All this combined may offer an incentive for not only one, but several firms to locate to the smaller region. There may thus be a move towards a decentralized equilibrium, where industrial production takes place in both regions. Whether this happens or not depends on a complex interplay between transport costs, economies of scale, and preferences.

Other factors, such as congestion costs or rising land prices in the larger region, may also lead to dispersion instead of concentration (features that were absent from Krugman's original model). The comparative-statics results in Krugman's analysis allow us to understand why urbanization, and the move towards a core-periphery structure, would tend to result if transport costs fell or technologies with increasing returns became more prevalent. Arguably, such trends were important during the process of industrialization.

Krugman's arguments also explain why locational patterns can change "catastrophically". Assume that the initial equilibrium is symmetric, with half of the population living in one region and half in the other. If trade costs begin to fall, there may be no immediate effects on migration and the location of production. But once costs fall below a threshold value, a cumulative process could be set off. This regional inequality arises endogenously even if all exogenous conditions are equal. Thus, we see how agglomeration tendencies may suddenly become stronger at a certain point of development. Illustrative solutions for the model, showing how some parameter values give rise to concentration of manufacturing activity while others give rise to decentralization, are studied in Fujita, Krugman and Venables (1999, Chapter 5).

#### *Further developments*

The initial analysis in Krugman (1991a) reached a number of noteworthy and striking results from a quite simple model, based on a number of specific assumptions. Much of the subsequent literature has been devoted to investigating the importance of these assumptions. As an example, Ottaviano, Tabuchi, and Thisse (2002) have developed an agglomeration model with linear rather than Dixit-Stiglitz preferences. They find that the basic insights from the original model remain unaffected. A central assumption of Krugman's original model is that agricultural products may be traded freely at no cost. This assumption is not consistent



with the data; real-world transport costs appear to be at least as high for agricultural goods as for manufactured products. This would neutralize the home-market effect (Davis, 1998). But Fujita, Krugman, and Venables (1999, Chapter 7) have shown that there are similar mechanisms in a world of transport costs in both sectors and differentiated agricultural products. In such cases, a reduction in agricultural transport costs may trigger agglomeration.

The basic core-periphery model has stimulated the economic analysis of spatial issues, thereby integrating economic geography with mainstream economics. A large literature with an emphasis on economies of scale has developed over the last fifteen years, with Krugman himself as a major contributor. One line of development has stressed the importance of input-output linkages among firms as an alternative explanation for agglomeration tendencies. Such linkages may be crucial for understanding the location of economic activities in situations of low labor mobility as is the case, for example, among many European countries. Input-output linkages were first analyzed by Krugman and Venables (1995) and Venables (1996). The main idea here is that the entry of new firms in a region increases the market for upstream suppliers (backward linkages as opposed to the forward linkages that stimulate the migration of workers). When upstream suppliers can produce at lower costs, the costs of downstream producers fall as well, due to increasing returns to scale. A cumulative agglomeration process could then be driven by the interaction between upstream and downstream firms. If labor mobility is low and the supply of labor is inelastic, then concentration of production must lead to rising wages. This has two opposite effects: higher household income leads to an increase in demand, whereas higher wages decrease firm profits and make relocation to the periphery more attractive. As a result, the set of possible equilibria is quite rich.

In other work, Krugman and several co-authors have bridged the gap between the new economic geography literature and the more traditional research in urban and regional economics (Fujita and Krugman, 1995, Fujita, Krugman and Venables, 1999, and Fujita, Krugman and Mori, 1999). These contributions seek, among other things, to answer the fundamental question of where and when new cities emerge. They emphasize how the land requirements of the agricultural sector interact with scale economies of the industrial sector. Provided that population size is not too large, the equilibrium outcome will be a monocentric economy, consistent with the classical work of von Thünen.

### *The empirical evidence*

Recent years have seen the development of a literature that investigates the empirical validity of the new economic geography; see Head and Mayer (2004) for a survey. There have been findings in favour of some propositions of the theory, such as the positive relationship between market size and wages, the relationship between market size and migration (although the effects seem likely to remain local and thus unable to generate a core-periphery pattern in, for example, Europe as whole), and the importance of backward linkages. There is also evidence of the productivity benefits derived from location in densely populated areas, although identifying the mechanism that drives this pattern remains a research challenge. Evidence on the home-market effect is mixed, as is the evidence on the relationship between agglomeration and the importance of both increasing returns to scale and trade costs.

### **Other contributions**

Apart from his work discussed so far, Paul Krugman has made important contributions in other areas. In trade theory, he has also analyzed so-called strategic trade policy, i.e., the incentives for one country to affect its terms of trade by introducing some barriers to trade (for instance, tariffs). This was already a well-established field within the traditional theory with perfect competition among firms. When trade is brought about by specialization due to economies of scale as described in the new trade literature, strategic trade policy is still relevant, and the policy becomes closely connected to issues that involve the regulation of industry. This insight was exploited in theoretical work initiated by Brander (1981) and Spencer and Brander (1983). Krugman also contributed to this literature in Brander and Krugman (1983). The policy conclusions of this approach are explored in the Helpman and Krugman (1989) monograph, *Trade Policy and Market Structure*. This work synthesizes the burgeoning new literature and analyzes a variety of policy issues that arise in models based on the new trade theory. It is shown that these models lead to qualitatively new and interesting effects of standard trade policies: protection could reduce domestic output, import subsidies could improve the terms of trade, and tariffs could reduce domestic prices.

Krugman has also made important contributions to the analysis of international monetary economics. A framework of analysis that set a new standard in the study of currency crises was proposed in Krugman (1979b). Here, he assumed that a government is trying to maintain

a fixed exchange rate despite some fundamental imbalance (for instance, the country has a higher long-run inflation rate as compared to the rest of the world) that makes such a peg impossible to maintain in the long run. By buying and selling currency in large amounts, the government can maintain the fixed exchange rate in the short run. Krugman analyzed how the expected future depletion of the government's currency reserve would be taken into account by rational investors, so as to ignite a speculative, early attack on the country's currency. Krugman's simple model captured the essential mechanism of currency crises in a way that has inspired considerable later research.

In a related strand of research, Krugman analyzed the movement of exchange rates within *target zones*. Such zones are relevant when central banks pursue exchange-rate policies which dictate that a currency's value be allowed to float within a specified band, such as those stipulated by the European Exchange Rate Mechanism (ERM) that preceded the adoption of the Euro as a common currency by many EU countries. In particular, Krugman (1991b) formulated the "canonical" model for the analysis of how exchange rates will behave within such a zone. The basic idea is that in the middle of the band, the exchange rate is equally likely to move upwards and downwards; the expected change is thus zero. Closer to, say, the upper end of the band, the exchange rate is more likely to move downwards than upwards, and the expected change is thus negative. This will be taken into account of by rational investors, and option-pricing models can be used to analyze the movements of the exchange rate within the band. In fact, a credible band tends to stabilize exchange rate movements. A large literature has emerged from the basic analysis in Krugman (1991b).

## **Conclusion**

By having integrated economies of scale into explicit general equilibrium models, Paul Krugman has deepened our understanding of the determinants of trade and the location of economic activity. His seminal papers published in 1979 (a) and 1980 were instrumental to the development of the new trade theory, and his 1991 (a) paper inspired the new approach to economic geography. His monographs, co-authored with Helpman and with Fujita and Venables, demonstrate the richness of the new theories.

## References

- Abdel-Rahman, H.M. (1988), "Product Differentiation, Monopolistic Competition, and City Size", *Regional Science and Urban Economics* 18, 69-86.
- Alonso, W. (1964), *Location and Land Use*, Cambridge, MA: Harvard University Press.
- Antweiler, W. and D. Trefler (2002), "Increasing Returns and All That: A View from Trade", *American Economic Review* 92, 93-119.
- Baldwin, R.E. and P. Krugman (1988), "Market Access and International Competition: A Simulation Study of 16K Random Access Memories", in R.C. Feenstra, ed., *Empirical Models for International Trade*, Cambridge, MA: MIT Press.
- Brander, J. (1981), "Intra-Industry Trade in Identical Commodities", *Journal of International Economics* 11, 1-14.
- Brander, J. and P. Krugman (1983), "A Reciprocal Dumping Model of International Trade", *Journal of International Economics* 15, 313-321.
- Chipman, J. (1970), "External Economies of Scale and Competitive Equilibrium", *Quarterly Journal of Economics* 85, 347-385.
- Corden, W.M. (1970), "A Note on Economies of Scale, the Size of the Domestic Market and the Pattern of Trade", in I.A. McDougall and R.H. Snape, eds., *Studies in International Economics*, Amsterdam, North-Holland.
- Davis, D. (1998), "The Home Market, Trade and Industrial Structure", *American Economic Review* 88, 1264-1276.
- Dixit, A. (1988), "Optimal Trade and Industrial Policy for the U.S. Automobile Industry", in R.C. Feenstra, ed., *Empirical Models for International Trade*, Cambridge, MA: MIT Press.
- Dixit, A. and V. Norman (1980), *Theory of International Trade: A Dual General Equilibrium Approach*, Cambridge, Cambridge University Press.
- Dixit, A. and J. Stiglitz (1977), "Monopolistic Competition and Optimum Product Diversity", *American Economic Review* 67, 297-308.
- Duranton, G. and D. Puga (2004), "Micro-Foundations of Urban Agglomeration Economics", in J.V. Henderson and J.-F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier North-Holland.
- Ethier, W.J. (1979), "Internationally Decreasing Costs and World Trade", *Journal of International Economics* 9, 1-25.
- Ethier, W.J. (1982), "National and International Returns to Scale in the Modern Theory of International Trade", *American Economic Review* 72, 225-238.

- Evenett, S.J. and W. Keller (2002), "On Theories Explaining the Success of the Gravity Equation", *Journal of Political Economy* 110, 281-316.
- Fujita, M. (1988), "A Monopolistic Competition Model of Spatial Agglomeration: Differentiated Product Approach", *Regional Science and Urban Economics* 18, 87-124.
- Fujita, M. and H. Ogawa (1982), "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations", *Regional Science and Urban Economics* 12, 161-196.
- Fujita, M. and P. Krugman (1995), "When is the Economy Monocentric? Von Thünen and Chamberlin Unified", *Regional Science and Urban Economics* 25, 505-528.
- Fujita, M. and P. Krugman (2004), "The New Economic Geography: Past, Present and the Future", *Papers in Regional Science* 83, 139-164.
- Fujita, M., P. Krugman and T. Mori (1999), "On the Evolution of Hierarchical Urban Systems", *European Economic Review* 43, 209-251.
- Fujita, M., P. Krugman and T. Venables, (1999), *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA: MIT Press.
- Grossman, G. and E. Helpman (1989), "Product Development and International Trade", *Journal of Political Economy* 97, 1261-1283.
- Grossman, G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.
- Grubel, H. and P. Lloyd (1975), *Intra-industry Trade*, London, McMillan.
- Harris, C. (1954), "The Market as a Factor in the Localization of Production", *Annals of the Association of American Geographers* 44, 315-348.
- Harris, R. (1984), "Applied General Equilibrium Analysis of Small Open Economies with Scale Economies and Imperfect Competition", *American Economic Review* 74, 1016-1032.
- Head, K. and T. Mayer (2004) "The Empirics of Agglomeration and Trade", in J. V. Henderson and J. F. Thisse, eds., *Handbook of Regional and Urban Economic*, Vol. 4, Amsterdam: Elsevier North-Holland.
- Helpman, E. (1981), "International Trade in the Presence of Product Differentiation, Economies of Scale, and Monopolistic Competition", *Journal of International Economics* 11, 305-340.
- Helpman, E. (1984), "A Simple Theory of International Trade with Multinational Corporations", *Journal of Political Economy* 94, 451-471.

- Helpman, E. (1987), "Imperfect Competition and International Trade: Evidence from Fourteen Industrial Countries", *Journal of the Japanese and International Economies* 92, 451-471.
- Helpman, E. and P. Krugman (1985), *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Cambridge, MA: MIT Press.
- Helpman, E. and P. Krugman (1989), *Trade Policy and Market Structure*, Cambridge, MA: MIT Press.
- Henderson, J.V. (1974), "The Sizes and Types of Cities", *American Economic Review* 64, 640-656.
- Hirschman, A.O. (1958), *The Strategy of Development*, New Haven, CN: Yale University Press.
- Hummels, D. and J. Levinsohn (1993), "Product Differentiation as a Source of Comparative Advantage", *American Economic Review Papers and Proceedings* 83, 445-449.
- Hummels, D. and J. Levinsohn (1995), "International Trade and Monopolistic Competition: Reconsidering the Evidence", *Quarterly Journal of Economics* 110, 799-836.
- Kemp, M. (1964), *The Pure Theory of International Trade*, Englewood Cliffs, NJ: Prentice-Hall.
- Krugman, P. (1979a), "Increasing Returns, Monopolistic Competition and International Trade", *Journal of International Economics* 9, 469-479.
- Krugman, P. (1979b), "A Model of Balance-of-Payment Crises", *Journal of Money, Credit and Banking* 11, 311-325.
- Krugman, P. (1980), "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review* 70, 950-959.
- Krugman, P. (1981), "Intra-industry Specialization and the Gains from Trade", *Journal of Political Economy* 89, 959-973.
- Krugman, P. (1991a), "Increasing Returns and Economic Geography", *Journal of Political Economy* 99, 483-499.
- Krugman, P. (1991b), "Target Zones and Exchange Rate Dynamics," *Quarterly Journal of Economics* 106, 669-682.
- Krugman, P. (1999), "Was it All in Ohlin?", unpublished notes.
- Krugman, P. and A. Venables (1995), Globalization and the Inequality of Nations, *Quarterly Journal of Economics*, 110, 857-880.

- Lancaster, K. (1980), "Intra-industry Trade under Perfect Monopolistic Competition", *Journal of International Economics* 10, 151-171.
- Leamer, E. and J. Levinsohn (1995), "International Trade Theory: The Evidence", in G. Grossman and K. Rogoff, eds., *Handbook of International Economics*, Vol. 3, Amsterdam: Elsevier Science, 1339-1344.
- Linder, S.B. (1961), *An Essay on Trade and Transformation*, New York, John Wiley.
- Markusen, J. (1984), "Multinationals, Multi-Plant Economies, and the Gains from Trade", *Journal of International Economics* 16, 205-226.
- Matthews, R.C.O. (1949), "Reciprocal Demand and Increasing Returns", *Review of Economic Studies* 17, 149-158.
- Melvin, J. (1969), "Increasing Returns to Scale as a Determinant of Trade", *Canadian Journal of Economics and Political Science* 2, 389-402.
- Mundell, R. (1957), "International Trade and Factor Mobility", *American Economic Review*, 47, 321-335.
- Murphy, K. M., A. Shleifer, and R. W. Vishny (1989a), "Income Distribution, Market Size, and Industrialization", *Quarterly Journal of Economics*, 104, 537-564.
- Murphy, K. M., A. Shleifer, and R. W. Vishny (1989b) "Industrialization and the Big Push", *Journal of Political Economy* 97, 1003-1026.
- Myrdal, G. (1957), *Economic Theory and Underdeveloped Regions*, London: Duckworth.
- Negishi, T. (1969), "Marshallian External Economies and Gains from Trade between Similar Countries", *Review of Economic Studies* 36, 131-135.
- Ohlin, B. (1933), *Interregional and International Trade*, Cambridge, MA: Harvard University Press.
- Ottaviano, G., T. Tabuchi and J.-F. Thisse (2002), "Agglomeration and Trade Revisited", *International Economic Review* 43, 409-436.
- Pred, A. R. (1966), *The Spatial Dynamics of U.S. Urban-Industrial Growth, 1800-1914: Interpretive and Theoretical Essays*, Cambridge, MA: MIT Press.
- Rosenthal, S. S. and W. C. Strange (2004), "Evidence on the Nature and Sources of Agglomeration Economics", in J. V. Henderson and J.-F. Thisse, eds., *Handbook of Regional and Urban Economics*, Vol. 4, Amsterdam: Elsevier North-Holland.
- Spencer, B. and J. Brander (1983), "International R&D Rivalry and Industrial Strategy", *Review of Economic Studies* 50, 707-722.

- Starrett, D. (1978), "Market Allocations of Location Choice in a Model with Free Mobility", *Journal of Economic Theory* 17, 21-37.
- Venables, A. J. (1996), "Equilibrium Locations of Vertically Linked Industries", *International Economic Review*, 37, 341-359.
- von Thünen, J.H. (1826), *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg: Perthes. (English translation: *The Isolated State*, Oxford: Pergamon, 1966).