

THE NEW ECONOMIC GEOGRAPHY

Armin Schmutzler

Socioeconomic Center, Zürich

Abstract. Recently, the ‘new economic geography’ literature has developed as a theory of the emergence of large agglomerations which relies on increasing returns to scale and transportation costs. This literature builds on diverse intellectual traditions. It combines the insights of traditional regional science with those of modern trade theory and thus attempts to provide an integrative approach to interregional and international trade. The paper surveys this literature and discusses its relation to earlier approaches to similar topics.

Keywords. Agglomerations; Economic geography; Positive externalities

1. Introduction

Perhaps surprisingly, for a long time international economists more or less ignored such concepts as distance, space and transportation costs. Neither Heckscher-Ohlin theory nor new trade theory (e.g. Helpman and Krugman, 1985) relies on the inclusion of these factors. Nevertheless, there were some early attempts to investigate the role of distance between and within nations for the analysis of international trade. For instance, in his ‘Theory of Interregional and International Trade’, Bertil Ohlin himself analyzed how transportation costs affect patterns of trade and specialization. For him, ‘international trade theory cannot be understood except in relation to and as part of the general location theory, to which the lack of mobility of goods and factors has equal relevance’ (Ohlin, 1933, p. 141–142). Nevertheless, for decades to come, the role of distance and space was almost exclusively the remit of regional economists. Recently, this has been changing. Since the publication of Paul Krugman’s ‘Geography and Trade’ in 1991, a burgeoning literature has developed under the heading ‘new economic geography’. In the following, I shall survey the new economic geography literature and evaluate its contribution relative to earlier work on similar topics.

Like earlier works in regional science, the new economic geography deals with variants of one basic question, namely, which factors have influenced and continue to influence the geographical distribution of economic activity? For instance, why did European manufacturing concentrate in such regions as the Midlands, Northern France, the Ruhr Valley and Northern Italy in the early stages of the industrial revolution? Why has this spatial distribution been so

remarkably robust for a long time? What have been the causes of recent changes in the pattern of manufacturing activity?¹ Corresponding questions can be asked for the United States, where, for a long time, manufacturing activity was concentrated in a comparatively small part of the country, in the manufacturing belt between the Great Lakes and New England, before it gradually shifted to the South and the West (see Hoover 1948, ch. 9, Fuchs 1962, Krugman 1991c).

An evident argument in the explanation of geographical concentration is that certain regions enjoy 'first-nature' advantages over others, such as superior endowment with natural resources or transportation facilities such as rivers or harbour.² Often, such factors reasonably explain why some regions face a particular concentration of economic activity. However, there are many cases where regions without obvious natural advantages develop into economic centers. In such cases, additional arguments need to be invoked to understand concentration.

Reduced to its essence, the new economic geography is a theory of the emergence of large agglomerations which relies on increasing returns to scale and transportation costs,³ and emphasizes linkages between firms and suppliers as well as between firms and consumer.⁴ The basic story underlying this type of analysis can be sketched as follows. Increasing returns to scale tend to foster geographical concentration of production of each good. When transportation costs play a role, attractive locations for production are those which are close to markets and suppliers, other things being equal. Finally, concentration of production in some location tends to attract the mobile factors of production. Workers have better job and consumption opportunities where production is concentrated. The resulting concentration of the labour force leads to more demand for consumption goods in that location, which makes this region more attractive for producers. Once a region has a high share of production, this pattern is likely to reinforce itself: a so-called second-nature advantage for the dominant region develops, that is, the region becomes attractive for firms because so many other firms already produce there (rather than because of superior resource endowment). In other words, success breeds success. Working against these centripetal forces which strengthen agglomerations are centrifugal forces. For instance, concentration of productive activities in one region may drive land rents and housing prices up, and may lead to environmental problems. Also, if immobile factors of productions remain in peripheral areas, firms from the center may want to move there to serve these areas. The population and production patterns result from a balancing of these centrifugal forces and the centripetal forces.

I shall proceed as follows. In Section 2, I shall trace the intellectual roots of the new economic geography literature. The remaining sections analyze the new economic geography in more detail. Section 3 summarizes Krugman's model (1991b, 1991c, 1992) which focuses on the linkages between producers and workers/consumers. I also highlight some of the missing elements in this model, which have been the starting point for further work. Section 4 presents some of these additional contributions. Section 5 discusses those modifications of the basic Krugman model that include backward and forward linkages between firms

and their suppliers, in more detail, and explores the applications of the modified framework to international trade theory. Section 6 concludes.

2. The new economic geography — background

To sum up the discussion of the last section, new economic geography models of economic agglomeration have the following elements. First, they emphasize advantages of concentration which are unrelated to natural endowments. Hence, arguments of circular causation play a role, that is, dominance of regions is regarded as a self-reinforcing process that can be sparked off by a small event. Second, the whole approach has a distinct general equilibrium flavor. The interactions between different markets, between firms and their suppliers and customers, and the dual role of workers as production factors and consumers are emphasized. Third, the centripetal forces favoring agglomerations are weakened by counterveiling centrifugal forces. Fourth, microfoundations are important. In particular, positive externalities are not assumed, they are derived from the interplay of transportation costs, increasing returns to scale and factor mobility.

No single one of these aspects is new to spatial economics. This is particularly true for the point that there are potential advantages from geographical concentration of economic activity. For firms within industries, this point has been made by Marshall (1920, ch. 10) who distinguishes between advantages from having a larger local labour pool, from employing common non-traded inputs and from knowledge spillovers.⁵ The Marshallian arguments do not rely on general equilibrium interactions; in fact, they are particularly suitable for explaining small-scale concentration of firms within specific industries. They may explain why cities or small areas without specific natural advantages in the production of certain goods become highly specialized in these goods: the more or less arbitrary decision of a small number of firms to locate in one region may induce others to follow. However, they can probably not explain the existence of vast agglomerations with firms from different industries.

Thus, apart from emphasizing (Marshallian) 'external economies of concentration of a particular industry', scholars such as Ohlin (1933, p. 203) also pointed to the 'economies of concentration of industries in general'. In a similar vein, Hoover (1948, p. 3/4) maintained that '...economic interrelations between different industries and firms play an important part in shaping the pattern of location as a whole' and consequently '...even in the absence of any initial differentiation, ..., - patterns of specialization and concentration of activities would inevitably appear', the reason being not only advantages from concentrating certain kinds of business in relatively few locations', but also from 'proximity of related processes' and from the closeness of consumers and producers.

Second-nature advantages of concentration resulting from interactions between different sectors are also familiar from the development literature of the 1950s and 1960s which aimed at explaining why some regions attract more production and population than others. Stories of circular causation were as popular as they are in the new economic geography literature, and the mechanisms described were similar

to those sketched above. For instance, Hirschman (1963, p. 100) emphasized that there are '...backward linkage effects, i.e., every non-primary activity will induce attempts to supply through domestic production the inputs needed in that activity' and 'forward linkage effects, i.e., every activity that does not by its nature cater exclusively to final demands, will induce attempts to utilize its outputs as inputs in some new activity'. Hirschman's focus is on how best to exploit these linkages for development strategies. Myrdal (1957, ch. 3) used similar kinds of arguments to explain persistent dominance of regions in Third World countries. 'Backwash effects' reinforce the dominance of successful cities, for instance, by inducing selective migration of younger and possibly relatively well qualified workers. In short, the older (mostly qualitative) literature was well aware of the nature of centripetal force.⁶ This literature usually regarded the existence of these forces as problematic, both for the center and the periphery. For instance, Friedmann and Alonso (1963, p. 3) hold that '...centers not only grow so rapidly as to create problems of an entirely new order, but they also act as suction pumps pulling into the more dynamic elements from the more static regions'.

The idea that tendencies towards concentration are reduced by various centrifugal forces is not exactly new either. For instance, Ohlin (1933) stated that deglomeration forces such as transportation costs, land rents and high labour prices limit the extent of agglomeration.⁷ Similarly, in his central place theory which attempts to explain regularities in the distribution of urban centers in Southern Germany, Christaller (1933, p. 28) highlighted the tradeoff between scale economies as sources of agglomeration and high rents and wages as countervailing forces, and asks which geographical patterns are likely to emerge from this interplay. Myrdal's backwash effects were partly offset by 'spread effects': growth in centers may induce growth in some peripheral regions if these regions are needed to supply the centers, for instance because they have good resource bases. More recently, formal models in urban economics have addressed the interaction of centripetal and centrifugal forces. For instance, Henderson (1974) assumes localization economies in the spirit of Marshall as centripetal forces, and he considers urban land rents as centrifugal force.⁸

Like other recent contributions to Urban Economics, Henderson's paper has solid microfoundations. However, locational externalities are merely assumed. While this simplification is helpful to investigate the consequences of externalities, it is inadequate for a study of their causes.

So what, if anything, is new about the new economic geography? Indeed, some critics maintain that the contribution is very limited. Some regional economists, in particular, argue that Krugman (1991c) is old wine in new bottles, and that he restates what has been familiar in the field for decades, but somehow manages to make more people listen than others before him. Indeed, good marketing is certainly an important part of the story. One achievement of the new economic geography literature is that, thanks to these marketing efforts, it has reminded mainstream economists that some (old) ideas from regional economics are interesting. However, there are some reasons to believe that the contribution of the new literature goes beyond this.

First, in a limited sense, the new economic geography adds something to regional economics itself. Even though neither centripetal and centrifugal forces nor general equilibrium considerations and microfoundations are unknown to regional economists, until recently there was no single approach that emphasizes all of these points in a coherent framework. This is mainly because increasing returns to scale are crucial to the explanation of agglomeration pattern.⁹ Therefore, traditional Arrow-Debreu type general equilibrium approaches are unsuitable for issues of economic geography, because they rely on convex technology sets. The new economic geography uses a more recent modeling framework, introduced by Dixit and Stiglitz (1977), which is known for its applications to trade theory (see Krugman 1979). This model sacrifices generality for tractability. It uses a convenient assumption on market structure (monopolistic competition) to avoid the problems associated with price-taking behavior when there are increasing returns to scale, without getting into the intricacies of strategic interaction. It employs very specific functional forms for consumer preferences, as will be described in the next section. Thanks to this ingenious model, the new economic geography offers a convenient framework for the rigorous investigation of macro level spatial economics, which has been missing so far.

Second, and more importantly, with its emphasis on the fact that ‘... countries both occupy and exist in space’ (Krugman 1991c, p. 2), the new economic geography has certainly contributed to a better understanding of international trade theory. In this respect, it goes beyond the most notable intellectual ancestor, Ohlin (1933). While Ohlin made many important points concerning the relationships between transportation costs and trade,¹⁰ the new methods appear to lead to new insights about some particularly relevant economic questions of our time. The most conspicuous recent tendencies in the world economy have been increases in regional integration, fundamental changes in the distribution of manufacturing activity, and a rapid increase in foreign direct investment, and there is little doubt that these processes will continue for some time. As will be sketched in Section 5, the new literature has something to say about the causes and consequences of these developments.

To sum up, many of the ideas from new economic geography are familiar from regional economics and from informal approaches to trade and development theory. The specific modelling approach, however, has some new elements which are worth a more careful discussion.

3. Second nature advantages of agglomeration — the basic model

I shall now present a model by Krugman (1991b, 1991c) that is designed to show how large-scale agglomerations can emerge from the interaction of increasing returns and transportation costs. On a general level, the model allows to investigate what effects different factors have on the robustness of agglomerations. Most of the contributions we shall consider in the following are closely related to this model. I shall discuss it in some detail.

Assumptions

Consider an economy with two sectors, manufacturing and agriculture. The agricultural sector produces a single homogeneous (numéraire) good under perfect competition; the total quantity is C_A . The manufacturing sector is characterized by monopolistic competition. A large number of potential firms can each produce differentiated products $i=1, \dots, I$. These products are symmetric in the sense that consumers do not prefer one product to another one. However, consumers have preference for variety; in particular, starting from any consumption vector a unit of a product that is not yet consumed is always preferred to an additional unit of a product that is already consumed.

A functional form that captures symmetry and preference for variety is given as follows. Let c_i be the quantity consumed of product i . Then the utility of the representative consumer derived from (c_1, \dots, c_I) is given in the CES-form

$$C_M = \left[\sum_{i=1}^I c_i^{\sigma-1/\sigma} \right]^{\sigma/\sigma-1}, \quad (1)$$

where $\sigma > 1$ (see for instance Krugman 1991b). Total utility from consuming a vector (c_1, \dots, c_I) of manufacturing goods and C_A units of the agricultural good is given as

$$U(C_A, C_M) = C_M^\mu C_A^{1-\mu}, \quad (2)$$

where $\mu \in (0, 1)$. Checking through the first-order conditions of the household's maximization problem shows that with this specification of utility, the share of consumer expenditures devoted to manufacturing goods in household equilibrium is μ and the elasticity of substitution between different manufacturing goods is constant (σ).

To simplify further, assume that the only production factor in the economy is labour. However, there are two types of labour, *workers* who produce the manufactured goods and *farmers* who produce the agricultural good. The supply of agricultural (manufacturing) labour is given exogenously as L_A (L_M); farmers never become workers or vice versa. To simplify assume that the share of manufacturing workers in the population equals μ , the share of manufacturing in consumer expenditure. The agricultural sector works with constant returns to scale. Each variety of the manufacturing good is produced with increasing returns to scale, more precisely, with positive fixed costs and constant marginal costs. These costs are identical across different manufacturing products.

Geography enters the model in the simplest conceivable fashion. The economy consists of two distinct regions, each of which is treated as a single point. Suppose for the moment that both types of labour are totally immobile, so that the distribution of workers and farmers across regions is fixed. The transportation of manufactured goods between regions is costly. For analytical convenience, Krugman uses exogenous *iceberg* transportation costs, that is, he assumes that a certain fraction of the goods does not reach its destination.¹¹ More precisely,

suppose τ is a positive constant between 0 and 1. Then, of any unit of the good transported from region 1 to region 2 or vice versa, only $1 - \tau$ units arrive at the destination. Therefore, τ parametrizes transportation costs: the greater the losses from transportation, i.e., the greater the fraction of the 'iceberg' that 'melts' in the process of transportation. Transportation costs for agricultural goods are assumed to be non-existent (see Krugman 1991b, fn. 2 for a discussion of this point).

The behavioral assumptions are as follows. Consumers maximize their utility functions $U(C_A, C_M)$ given their budget constraints. There is free entry for firms; and firms maximize profits. This completes the description of the assumptions.

When do agglomerations arise?

Given this set-up, Krugman shows how manufacturing is distributed across regions. In particular, he investigates under which circumstances agglomerations arise, that is, under which circumstances the entire manufacturing population will concentrate in one region.

Several intermediate results are helpful here. First, for a large number of manufacturing products, the demand elasticity is approximately constant and the same as the elasticity of substitution. As a result, profit maximizing firms set a constant mark-up over marginal costs. Second, because of increasing returns to scale, each firm produces only one product. Third, with free entry profits are zero. Fourth, because of the symmetry of the problem each firm produces at the same output level in equilibrium. The equilibrium output of each firm is a positive function of the fixed costs and the elasticity of substitution, and a negative function of marginal costs. The number of firms in a region is a positive function of its manufacturing labour supply, and a negative function of both fixed and marginal costs. These results are intuitive: with a high elasticity of substitution, consumers do not value variety very much, so there will be a small number of large firms in equilibrium, and this effect will be stronger when fixed costs are high. High marginal costs obviously reduce the output society can produce with a given labour supply, which implies that a small number of firms will be producing low outputs.

Using these intermediate results, Krugman goes on to analyze the centripetal and centrifugal forces in this model. To this end, he introduces labour mobility. For simplicity, he assumes that, while agricultural labour continues to be immobile, manufacturing labour always moves towards the region that offers the higher present real wage; more precisely, the population of manufacturing workers moves to the high wage location at a speed that is proportional to the present wage differential.¹² In equilibrium, both regions must either offer the same manufacturing wage or the manufacturing population must be concentrated in a region offering the higher real wage. Under which conditions can a concentration of the entire manufacturing activity in one location (the center) arise in equilibrium?

For such a constellation to be robust, no firm must have an incentive to build a plant in the periphery, that is, the location without manufacturing. There are two reasons why such a deviation might be unprofitable, both related to transporta-

tion costs. First, the firm must induce manufacturing workers from the center to work in the periphery. As these workers will have to import most consumption goods from the center, the costs of living are higher in the periphery. As a result, the firm must pay higher wages, which drives up the output price. Second, the majority of the firm's customers live in the center. Serving them from the periphery involves transportation costs, another reason to stay in the center. On the other hand, the agricultural population in the periphery can be served cheaper if the firm produces in the periphery. Basically, an agglomeration equilibrium results when the last, centrifugal, effect is small relative to the first two, centripetal, effects. It remains to be shown what determines whether centripetal or centrifugal forces dominate. Specifically, which effects do transportation costs, the size of the manufacturing sector and consumer preferences for variety have on the robustness of agglomerations?

Without transportation costs, location does not matter in this set-up. There are neither centripetal nor centrifugal forces, so that the locational pattern is indeterminate. If transportation costs are very high, an agglomeration becomes unlikely: it is prohibitively costly to serve the periphery from the center, and deviating from an agglomeration may be a profitable strategy. An agglomeration can only arise if transportation costs are positive, but so small that serving the periphery from the center is a feasible alternative to local production.

The size of the manufacturing sector, as measured through the share of manufacturers in consumer expenditures, or equivalently, through the share of manufacturing workers in the population, has two effects which work in the same direction. For a high share of manufacturing in consumer expenditure, the extra wage necessary to compensate workers for living in the periphery is high, since a large quantity of manufacturing goods have to be imported. In addition, in this scenario, the agricultural population and hence the size of the market in the periphery will be small, which will further weaken the centrifugal forces.

Finally, the elasticity of substitution matters. First note that there is a close relationship between the elasticity of substitution and economies of scale. In equilibrium, $\sigma/(\sigma - 1)$ equals the ratio of average cost to marginal cost, a common measure of economies of scale.¹³ Hence, a low elasticity of substitution tends to go along with high economies of scale, which makes it less attractive to serve the smaller market locally.

To sum up: with higher transportation costs, an increasingly important manufacturing sector, and more significant economies of scale, agglomerations become more robust.

The determinants of the manufacturing pattern

Agglomeration is not the only possible equilibrium in this set-up. Krugman (1992) uses numerical simulations to show which equilibrium constellations can arise as a function of various exogenous factors. For instance, he investigates how transportation costs affect the equilibrium distribution of manufacturing over regions. Suppose region 1 has a slightly larger share of agricultural population,

and the other parameter values are suitably chosen. Then, for relatively high transportation costs, there is an equilibrium such that both regions have some manufacturing, but region 1 has a higher manufacturing share than region 2. This reflects the fact that the larger market is more attractive for manufacturing firms. As transportation costs fall, it becomes more attractive to serve region 2 from the larger market. The share of region 1 grows. For lower values, a new equilibrium emerges where all manufacturing is concentrated in region 1. Finally, for very low values of transportation costs, this equilibrium is unique: the advantages of concentrating production dominate over the advantage of being close to the peripheral market; because of the slight asymmetry in agricultural population, region 1 is the better location for production. Krugman argues that this is roughly consistent with the empirical observation that the development of railroads to Southern Italy in the nineteenth century which exposed the local industry to competition from the North eventually led to its collapse.¹⁴

Manufacturing patterns also depend on the distribution of the farming population. If region 1's share of the agricultural population is sufficiently high, and the other parameters have suitable values,¹⁵ all manufacturing will be concentrated there in the unique equilibrium, because centrifugal forces coming from the desire to serve location 2 are small. As the share decreases, the system goes through a sequence of bifurcations, that is, discontinuous changes in the equilibrium structure. First an additional equilibrium emerges with some manufacturing in region 2. As region 1's share of the agricultural population decreases, this is the only equilibrium. Next, a new equilibrium emerges without any production in region 1. Finally, for a sufficiently low agriculture share, this becomes the only equilibrium.

Krugman argues that the move of the manufacturing industry to California around 1900 might be explained by an increase in the state's agricultural base, where 'agriculture' is taken somewhat loosely to apply to the oil industry, which is definitely not footloose and hence similar to the agricultural sector of his model. Given the initial situation where (almost) all manufacturing was concentrated in the east (region 1), the oil industry had to reach a critical mass for a new equilibrium with some footloose production to emerge in California (region 2).

Both stories are nice illustrations of the principles, not more. Krugman makes no serious attempts to compare them with other possible explanations. Therefore, not everyone agrees about their relevance. In direct response to Krugman, Rauch (1993b) for instance has argued that cheap labour and the invention of air-conditioning played a more important role in the move of American manufacturing to the South and West.¹⁶

The basic model — insights and limitations

Summing up, the main insights of the model are as follows. First, suppose two regions start out (almost) identically by nature, in the sense that no region has a superior resource base or technology or a larger consumer market. Then an agglomeration of manufacturing can develop endogenously in one of the regions. In

this development, history matters. Due to minor historical events, small initial differences may lead to a core-periphery structure. Second, the structure of the equilibrium set depends on transportation costs, agricultural shares, etc. in a fairly complex way: for some parameter values, manufacturing will always be evenly distributed in equilibrium; for others, the equilibrium set contains only the two full agglomeration equilibria; for yet others, all three equilibria exist and are locally stable. Continuous parameter changes may lead to discontinuous change. In particular, integration favors agglomeration. Third, whether agglomeration develops, depends on the interplay of different forces. Positive demand externalities lead to centripetal forces favoring agglomeration. The main centrifugal force identified here is the desire to keep the transportation costs of serving the periphery low.

4. Modifications of the basic model

While the basic Krugman model captures important aspects of the evolution of spatial patterns, it relies on a number of assumptions, which, if relaxed, lead to additional insights. The following seven assumptions are worth exploring in more detail.

- (1) There are no direct negative externalities between firms (e.g. due to pollution or congestion).
- (2) There is no market for housing and land.
- (3) Households are indifferent between regions offering identical wages; in particular, they do not take into account such issues as pollution or the value of amenities (landscape, climate, etc).
- (4) There are only two regions.
- (5) The dynamics of locational choice are not derived explicitly from forward-looking optimizing behavior.
- (6) No region has a superior resource base or technology.
- (7) There are no intermediate goods.

As will argued in more detail below, relaxing assumptions (1)–(3) adds new centrifugal forces. Dropping (4), that is, adding more regions, leads to more complex agglomeration patterns, and it also makes the framework applicable to new questions. Relaxing (5) is important because it is not obvious how strongly the result depends on the ad-hoc adjustment process postulated in the basic model. Without assumption (6) some interesting efficiency questions arise, which are absent in the basic model. Abandoning (7), i.e., introducing intermediate goods leads to additional backward and forward linkages which I shall deal with in section 5. In the remainder of this section, I shall confine myself to modification (1)–(6).

Additional centripetal forces

The centrifugal forces in Krugman's original model reflect the desire of firms to serve the periphery. In this section, we briefly indicate how other authors

have introduced alternative centrifugal forces, such as negative technological externalities, urban land rents, and preferences for quality of life.

Brakman *et al.* (1994) modify the basic Krugman model by introducing negative technological externalities. To this end, they take a multi-region version of the basic Krugman model described in Section 3, and add the following technological assumption. The fixed and the marginal costs associated with the production of any variety of the good depend positively on the number of firms in the location. This captures the idea of congestion effects. Not surprisingly, the introduction of congestion affects the nature of the equilibrium set. The negative congestion externalities make production in the dominant manufacturing region excessively costly, so that agglomeration is usually not complete and firms will therefore find it profitable to move to the less congested periphery. The findings of the authors are one way to explain the observation that a complete concentration of manufacturing in one region rarely occurs; and that even peripheral regions usually contain some manufacturing.¹⁷

Urban land rents and commuting costs are the centrifugal force in a paper by Livas Elizondo and Krugman (1996). Unlike the models introduced before, this model operates without the assumption that a region is a dimensionless point. Instead, farming and housing are space-consuming activities. With this modification, agglomerations have to offer higher wages to compensate for commuting costs and land rents. To capture this effect, regions are modeled as long, narrow cities, i.e., as intervals of the real line. Production in each region takes place in the center of the interval. Within each city, the population is distributed evenly. There are commuting costs that are increasing in the distance from the center. Land rents offset these commuting costs: agents who live close to the center have to pay higher land rents, so that in equilibrium the sum of land rents and commuting costs is constant within each region. The authors assume that a concentration of manufacturing workers in a region results in a greater length of the city, which drives up commuting costs and rents and therefore reduces the likelihood of agglomeration.

Preferences for quality of life are present in a paper by Asilis and Rivera-Batiz (1994). In their model, consumers have preferences over consumption goods as in the basic model. In addition, however, they also have preferences as to where they want to live relative to the center of the region (which is modeled as an interval); in particular the authors allow for the possibility that consumers have preferences for living far away from the center to escape from pollution. However, these preferences have to be weighed against higher costs of living resulting from the costs of transporting manufacturing goods. Nevertheless, the result of the trade-off may reduce the strong forces for agglomeration present in models without preferences for location.¹⁸

Multi-region models

One of the distinctive features of the basic Krugman model is that there are only two locations. The main insights generated in this framework about the role of

history and the relation between transportation costs, scale economies and agglomeration patterns are robust to the relaxation of this assumption. Yet, new questions can be addressed in multi-region models.

First, with many potential locations, agglomeration and decentralization are not the only possible equilibria. In particular, multiple agglomerations in different regions are conceivable as equilibria. This raises a series of questions: Can anything general be said about the number of agglomerations in equilibrium? Are there any typical patterns in the locations of these equilibria? Second, with more than two regions it makes sense to introduce differences in transportation costs between different regions. For instance, in a three-region model two regions can be regarded as different regions within the boundaries of one state, while the third region may be a different country. The assumption of relatively low transportation costs between the first two regions can then be interpreted as the absence of trade barriers (see e.g. Livas-Elizondo and Krugman 1996). In such a framework, one can investigate how trade barriers affect manufacturing patterns *within* countries. There are two distinct types of multi-region models, namely models with finitely many locations and models with a continuum of locations.

Krugman (1993b, 1994) addresses the case of a finitely number of locations. He generalizes the basic model in a straightforward fashion to an arbitrary finite number of regions (see also Krugman 1992). To solve the resulting equilibrium equations, numerical simulations are necessary. Krugman illustrates this for the case of twelve equidistant regions. Given various initial distributions of manufacturing in these regions and an appropriate generalization of the process of adjustment to regional wage differentials,¹⁹ he checks for the resulting final distributions. It turns out that a great number of equilibrium constellations exist, which share common features. The typical equilibrium involves agglomeration, but usually in more than one region. In most cases, two agglomerations in almost diametrically opposed locations emerge. On rare occasions, there are three agglomerations. The three insights from the two-region model stated at the end of Section 3 are essentially robust to the addition of further regions: asymmetry between regions can emerge endogenously so that history matters; for different parameters different types of equilibria emerge; and the interplay of centrifugal and centripetal forces determines whether or not an agglomeration occurs.

The most interesting idea added by this extension relates to the question of where agglomerations might occur. Simulations suggest there is a 'principle of maximum differentiation': when both transportation costs and increasing returns are strong enough for multiple agglomerations to result in equilibrium, these agglomerations tend to keep away from each other.²⁰ Unfortunately, these models lack transparency — at least in the published versions of the above papers, it is hard to determine what exactly drives the results.

An alternative is the case where the space of possible locations is a continuum, which Krugman (1993a) considers.²¹ He modifies the basic set-up by assuming that manufacturing firms can choose their location in the entire unit interval.²² He assumes that the agricultural population is distributed evenly across space and

that transportation costs are an increasing function of the distance between the firm and the customers.

With this modification, central locations have an initial advantage: the closer an interval is to the center, the lower are the transportation costs of serving an evenly distributed population. Nevertheless, the center is usually not the only possible equilibrium. After noting that for sufficiently high transportation costs, equilibria with only one agglomeration do not always exist, Krugman goes on to show which locations are suitable for such monocentric equilibria in cases where they do exist.²³ The typical set of equilibria which can be calculated in this fashion consists of the center of the unit interval and locations to both side of the center which are sufficiently close to the center.

The intuition is as follows. If only the costs of serving the agricultural population mattered, only the mid-point of the interval could be an equilibrium: in any other situation firms would have an incentive to move to the mid-point to economize on transportation costs. However, if an agglomeration has formed slightly to the left or right of the center, second-nature advantages as described in Section 3 may render such an agglomeration stable. With the entire manufacturing population concentrated in the agglomeration, it is cheaper to serve it from this point than from the middle of the interval. Not surprisingly, the size of the equilibrium set is increasing in the share of manufacturing goods in spending: the more important manufacturing and hence the positive externalities, the less attractive it is to move away from any existing agglomeration. Similarly, a lower elasticity of substitution (and hence a higher degree of scale economies in equilibrium), increases the equilibrium set. In other words, second nature is more likely to dominate first nature if the share of footloose production is high and economies of scale are high.

The introduction of asymmetry between regions therefore helps to obtain a partial answer to the question of *where* agglomerations form: while externalities produce a certain ambiguity in the exact location, relatively central locations are more likely to emerge as industrial agglomerations because of their transportation-cost advantages.

The effects of trade liberalization on the internal geography of a nation

Another example of a framework with more than two regions is the above-mentioned paper by Livas-Elizondo and Krugman (1996). The author's goal is to explain why many of the world's largest cities have developed in Third World countries, sometimes in spite of government effort to encourage decentralization. More precisely, Livas-Elizondo and Krugman give a possible explanation why there is a negative relation between geographical concentration within a country and the degree of trade liberalization of this country. This empirical relation is apparent from an investigation of 85 countries by Ades and Glaeser (1995). It has also been pointed out by Hanson (1994) who attributes the dominance of Mexico City in national manufacturing to the country's import substitution policy, and the recent gains of other regions to the liberalization policy in the eighties.

To explain such phenomena, Livas-Elizondo and Krugman use a modified version of the basic model. Their model works with the assumptions on goods, technology, market structure and preferences that are familiar from the Krugman model. However, there are three regions (long narrow cities), two of which are domestic (Mexico City and the rest of the country), while the third stands for the rest of the world. Labour is perfectly mobile between domestic regions, but not between domestic regions and the rest of the world. The transportation of goods within the country and imports is subject to 'iceberg' costs as defined in Section 3. However, there are two different cost parameters, one for transports within the country and one for imports. The second parameter includes trade barriers as well as ordinary transportation costs.²⁴ The fraction of manufacturing goods provided by the rest of the world is exogenous. As described above, regions are modeled to capture the centrifugal forces due to land rents and commuting costs in agglomerations.

The authors then investigate how the distribution of manufacturing *within* the country depends on international transportation costs. If these costs (trade barriers) are very high, there will be no international trade. In this case, if an industrial agglomeration exists, it will be the only supplier of manufacturing goods for consumption in the country. This will lead to the usual advantages due to positive externalities: manufacturers can supply local workers with cheaper goods than in the periphery, and the local demand is higher. A significant reduction of trade barriers reduces the importance of these centripetal forces: as the economy becomes more dependent on international markets, local demand is less important. In the region with more manufacturing, land rents and commuting costs are high. Low land rents and commuting costs attract firms to the other region because they can pay lower wages. Numerical simulations show that for very low international transportation costs, only the decentralized equilibrium will be sustainable, while for intermediate ranges, there are multiple stable equilibria: agglomerations in both regions and the decentralized equilibrium, where manufacturing is spread evenly across regions, are possible. In this sense, trade liberalization tends to break up geographical concentration within an economy, as the empirical evidence suggest.²⁵

The argument sketched in this section have met with some skepticism (Henderson 1996, Isserman 1996). The critics claim that the result may depend crucially on some of the simplifying assumptions. For instance, real-world centers are usually not only manufacturing centers, but also government centers, financial centers, etc. Taking this into account, centers might benefit from trade liberalization. Also, they argue that the non-tradeability of agricultural goods matters. If this assumption were relaxed, peripheral areas might suffer from imports of agricultural goods, and trade liberalization may weaken these areas. Finally, treating the distance between both regions within the country with the rest of the world as equal may be misleading: for instance, much of the development of Northern Mexico after trade has been liberalized is of course related to the fact that the area is closer to the United States than the center of Mexico.

Despite these qualifications, the paper by Livas-Elizondo and Krugman shows that the relations between trade liberalization and the internal geography of a country are potentially interesting.

Expectations, history and equilibrium selection

The dynamics of agglomeration in the basic model are driven by rather mechanistic assumptions. The share of any region in the manufacturing population increases *gradually* at a rate that is proportional to the *present* real wage differential. A gradual increase only makes sense if there are some costs of adjustment, otherwise workers should relocate whenever there is a non-zero wage differential. If such adjustment costs exist, however, rational workers should only move if they expect the differential to persist. Hence, expectations about future wage differentials are potentially an important factor in the migration decisions. It is not obvious what the dynamics look like in a world where agents choose their locations based on expectations about future wage differentials. One possibility is the emergence of self-fulfilling prophecies. Future wage differentials depend on the migration decisions of other people and hence on their expectations about other agents. If everybody believes everybody else is moving, and positive externalities are important, these beliefs may be self fulfilling.

However, it is not clear under which conditions self-fulfilling prophecies might dominate over history. Under which circumstances is it possible that, even though a region has a relatively low share of manufacturing workers and a correspondingly low current wage, sufficiently many people believe that its share of manufacturing (and wage rates) will increase in the future, leading to a migration that confirms these beliefs?

Krugman (1991c, appendix B) takes a first stab at these issues.²⁶ He considers a model which is closely related to the basic two-region model, though it is not strictly speaking consistent with it. He assumes there are two regions with given total labour force. The wage differential between the two regions is proportional to the difference in labour force sizes, the idea being that the region with the greater labour force benefits more strongly from agglomeration externalities.²⁷ At every moment of time, workers can decide whether or not to move locations. Workers face a moving cost function, which is convex in the rate of people that are moving at any moment in time. Rather than being concerned with present wage differentials, workers consider the discounted present values of future wage differentials, q . Incurring the costs of moving to another region can then be thought of as investing into an asset, the value of which is the present value of the expected wage differential. The assumptions give rise to a two-dimensional dynamic system in q and the share L_1 of region 1 in the labour force. The two equations governing the system are intuitive. First, at any time, the cost of migration of a marginal worker should be equal to the expected gain from changing location. Second, the rate of gain on the 'asset' of being in one rather than in another should be equal to the market interest rate, which is among the exogenous parameters of the model.

The dynamics of this model help to understand the factors that determine the relative importance of expectations and history. Suppose region 1 has a higher share of the manufacturing population than region 2, so that the present wage is higher in region 1. History will be said to dominate if this implies that the share of region 1 is increasing: in that case, the present wages determine expectations about future wages and hence migration decisions.

The qualitative behavior of the system depends on exogenous parameters in the following way. First, self-fulfilling expectations can play a role only if the interest rate is sufficiently low. Otherwise, the future does not influence locational decisions significantly, so there is little scope for expectations to matter. Second, external economies have to be sufficiently strong to generate a high wage differential. With weak linkages between people's decisions, there is not much room for self-fulfilling prophecies. Third, moving costs have to be sufficiently low. Otherwise, the economy will change slowly, so the present state of the world reveals what to expect for a long time. Therefore, it will take a long time for the sign of the wage differential to change, and, due to discounting, the present sign is most relevant for location decisions.

Note that self-fulfilling expectations can matter only for initial population shares that are sufficiently close to the point of equal distribution. If the initial advantage of one region is sufficiently strong, and hence the present wage differential is high, this will determine the dynamics: everybody will eventually move to the dominant region.²⁸ On the basis of his model, Krugman argues that self-fulfilling prophecies are unlikely to generate large-scale core-periphery shifts between regions, because the costs of moving factors rapidly are too high. Expectations are likely to be more important for the explanation of small-scale events, such as changes in the relative importance of cities.

Gali (1995) also considers a world where expectations matter, but he does not deal with the relative significance of history and expectations. He works with an overlapping generations model in an infinite horizon framework. For any given moment of time and given incomes of consumers, the problem is very similar to Krugman's basic model without labour mobility. The first difference is that manufacturing is the only kind of economic activity; there is no agricultural sector. The second difference is that there is no trade between regions. Whichever region has the greater share of population will produce a greater number of goods, thus yielding greater utility to the local population. Every consumer has to make two intertemporal choices at the beginning of his lifetime (two periods). First, whatever his labour endowment in each period, he has to choose the first period savings level. Second, he chooses the region where he lives. This choice is irreversible; there are infinite moving costs. When making this choice, the agent has to consider the present and expected future distribution of the labour force. This is the only difference between the two regions. In terms of utility from consumption, the agent prefers to live in the more populated regions. However, this effect is partly offset by a congestion effect. The author assumes that the utility function has an additional term that stands for the disutility from congestion (traffic congestion, housing rents,

crime, etc.).²⁹ Firms behave in a myopic fashion, maximizing profits in every period.

For this model, Galí shows how the set of possible steady state equilibria depends on the relative importance of congestion effects and demand externalities. What is more important, he shows that stationary sunspot equilibria exist if expectations about the future are sufficiently important in the consumers' location decisions. Broadly speaking, this says that stochastic shifts in the expectations about the future distribution of preferences about locations can be self-fulfilling. This causes random fluctuations in the populations of the regions and hence in their relative well-being (compare Woodford 1986).

Summing up, self-fulfilling expectations can play a role in the development of agglomeration patterns, but this appears to be relevant mainly for the explanation of small-scale events.

Heterogeneous locations

In the model described in Section 3, history matters in the sense that small events can determine in which of two essentially identical regions an industry locates. However, given the initial identity of the regions, this is not important from an efficiency viewpoint. The economy may get locked into one of two equilibria; as neither site has a significant natural advantage, however, this is not necessarily inefficient.³⁰ However, suppose that an agglomeration has developed. Suppose then that some kind of exogenous shock (technological or political) makes the new region more attractive, so that it would be more efficient for some or all of the production to move to the other region. Intuitively, it need not be the case that the economy moves to the new location: if the backward and forward linkages are strong enough, and firms and workers cannot coordinate a move to the new location, the economy might be locked in to the old location. While this issue has not been addressed in the monopolistic competition general equilibrium framework and will therefore not be dealt with in any more detail, it should be noted that there exist reduced-form models of such lock-in phenomena by Arthur (1990) and Rauch (1993a).

5. International specialization, backward and forward linkages

In the models treated so far, agglomerations arise because firms benefit from being close to workers, and vice versa, with labor mobility reinforcing initial advantages of a location. Much of the traditional regional economics literature, however, emphasizes vertical linkages between different firms: upstream firms benefit from being close to their customers, and downstream firms benefit from being close to their suppliers. More recently, a second strand within the new economic geography literature has developed that focuses on such vertical linkages. While it shares with the original Krugman approach the emphasis on transportation costs, it does not rely on labor mobility. As a result, it becomes relevant to issues where the relevant regions are parts of different countries. I shall very briefly sketch this literature.

The seminal paper was Venables (1996). Like Krugman, Venables dealt with the effects of decreasing trade costs on economic geography. He modified Krugman's model in several ways. Most importantly, he considered two monopolistically competitive industries, which are in an upstream-downstream relationship. Also, there is no labor mobility. Nevertheless, concentration of the manufacturing industry sales in one of the two (ex-ante) identical regions may result: upstream firms benefit from being in locations with many downstream firms, because thus they can serve customers more cheaply. Conversely, downstream firms benefit from being in a location with many upstream firms, because this decreases input costs.

In this setting, the effects of integration on the likelihood of agglomeration turn out to be non-monotone. As transportation costs decrease from high to medium, agglomeration becomes more likely by familiar arguments. As transportation costs decrease further, the lower wage in the periphery increases its share of total sales of world trade, thus reducing the tendency for concentration.

Krugman and Venables (1995) use a similar model to address the following puzzle. In the 1960's and 1970's, globalization was regarded as deepening the difference between North and South, while more recently, the opposite point of view, that globalization hurts rich countries, has become more common. A variant of the story just given shows that both parties may be right. In earlier periods, when trade costs fell below a critical value, the world economy developed a core-periphery structure. Further falls of trade costs, as experienced more recently, mean that proximity is becoming less important and production costs matter more, which benefits the periphery and reduces the difference between core and periphery.

Krugman and Venables (1996) use a variant of the Venables model to analyze whether increasing integration will make countries more or less similar in their industry structure. The two monopolistically competitive industries are no longer in a clear upstream-downstream relationship. Instead, each industry produces an output that can be used for consumption and as an input. Each industry relies to some extent on the other industry as supplier, but it also uses intermediate goods produced in the sector itself as an input. Crucially, such intra-industry linkages are assumed to be more important than inter-industry linkages: the cost share of inputs from the same industry is higher than the share from the other industry. Finally, while labour is internationally immobile; it can move between different sectors, and it moves towards the sector offering the higher present real wage.

In this setting, with high transportation costs, the industry structure is the same in both countries. For lower transportation costs, the stronger intra-industry linkages lead to specialization: firms of each industry concentrate in one of the two locations.

The arguments are appealing, but it is not obvious whether they are borne out in reality. Krugman and Venables offer very limited evidence. They argue that the United States are obviously more integrated than Europe and have a more specialized industrial structure. On the other hand, e.g. Molle (1997) shows that for certain industry classifications and definitions of specialization, interregional

specialization within Europe has been decreasing over the last few decades. Assuming that Europe has become more integrated in this period, this would seem to be hard to reconcile with the results of Krugman and Venables. At best therefore, the superficial empirical evidence in this case appears to be mixed.

Puga and Venables (1996) have used arguments based on backward and forward linkages to explain the spread of industries across different Asian countries. Markusen and Venables (1996) have integrated geography models with earlier models of multinational activity. This is very natural because most reasonable discussions of multinationals rely crucially on the existence of trade costs. They use modifications of the above approach to analyze the causes and consequences of multinational activity.

In brief, variants of the new economic geography models that rely on vertical linkages between firms rather than on labor mobility offer some promise for a better understanding of some important issues in international economics.

6. Concluding remarks

In this section, I have surveyed recent literature which explains how positive economies of agglomeration can emerge, and what their consequences are. The following main conclusions emerge.

First, history matters in the development of agglomerations. Cumulative processes generated by positive externalities can lead to the development of core-periphery structures even when no region has natural advantages. Second, transportation costs, the strength of scale economies and the importance of footloose industries are important factors determining whether such industrial concentration is likely to develop. Third, continuous changes in such parameters can lead to a discontinuous change in the equilibrium structure. Fourth, there are possible implications for trade: if positive externalities play a role, increasing economic integration affects both the distribution of manufacturing and the geographical distribution within the manufacturing sector. Fifth, there are interactions between the trade policy and the regional structure of an economy: increasing international integration may lead to decreasing concentration within the economy. Sixth, models with transportation costs are helpful to understand the causes and consequences of multinationals.

The results were derived under rather specific assumptions on geography, market structure, etc. While this is enough to demonstrate the theoretical possibility of certain phenomena, it is nevertheless desirable to understand how robust the general insights are.

Of course, some assumptions are crucial to the story: the existence of scale economies, of imperfect competition and transportation costs. Fortunately, these are among the more appealing assumptions of the models surveyed here. A more difficult issue is the exact nature of market structure. It would be reassuring to know that the results do not depend on the specific model of monopolistic competition in an essential way. Without tractable alternative general equilibrium models of imperfect competition, one has to resort to at least some degree of faith

in this matter. Modifications of other assumptions have been shown in Section 4 to leave some of the qualitative results unaffected, but they may influence the size of effects. With significant congestion effects, for instance, the endogenous formation of agglomerations is still possible, but there is usually a non-negligible share of manufacturing in the periphery. Some other modifications, however, are likely to lead to major changes. As argued in Section 4, this includes the simplistic modelling of centers as pure manufacturing centers.

The reception of the new literature by regional scientists has been mixed: while some complain about '... the failure of some new entrants to this debate to appreciate fully the previous literature', others '... welcome the fact that others are interested in traditional regional topics...' (Yezer 1995, p. 154). Whether this renewed interest in economic geography is just a research fad or a development with a long-lasting impact, will depend mainly on the insights the field will give to other subdisciplines, such as trade theory or growth theory. So far, the most important contribution may not be a particular result, but rather the emphasis on a specific approach to economic problems. The mechanisms yielding the characteristic patterns discussed earlier are not unique to economic geography. Positive externalities and the resulting phenomena such as the existence of multiple equilibria and the importance of history have received some attention in other areas of economics.³¹ In some sense, therefore, the main achievement of the economic geography literature may not be a single original contribution, but rather the popularization of ideas that have deserved more attention than they got in the past.

For the concern with regional issues to have a long-lasting impact, it will also be important that economists approach these issues in other terms than the very special modelling framework discussed here. There are some encouraging signs that this is actually happening. For instance, numerous empirical papers covering regional topics have recently been published in top economics journal.³² In a different sense, there is less reason for optimism. The theoretical papers with regional issues that are published in general journals almost exclusively rely on variants of the Dixit-Stiglitz model. In the short run, this may be a blessing. Having a dominant modelling structure simplifies the comparability of different papers. In the long-run, the benefits of this restriction are debatable. Many important regional issues are hard to integrate in the Dixit-Stiglitz framework. To approach issues such as the relationship between geography and innovation, for instance, partial equilibrium approaches may be more tractable; with only one fashionable modelling approach we might end up not talking about such issues at all.

Finally, of course, the new economic geography literature has one great shortcoming: so far, it has hardly generated any policy recommendation.³³ Maybe it is wise not to jump to such conclusions before the theoretical and empirical foundations have been developed further. In the long run, however, the new economic geography will have to produce more results in this direction, or it will face the same criticism of policy irrelevance that has been raised against regional science in the tradition of Isard (1956) for some time (see Huntoon 1995, Isserman 1995).

Acknowledgements

I am grateful to an anonymous referee and to Malte Faber, Switgard Feuerstein, Hans Gersbach, Peter Meusburger, Thomas Müller, Andreas Polk, Annette Schiller and Thomas Schröder for helpful comments.

Notes

1. For instance, why did the Ruhr and Saar regions' share of German manufacturing employment decline in the nineteen eighties, while the corresponding shares of the southern states Baden-Württemberg and Bayern increased?
2. In a broader sense such advantages could also arise from regional differences in governmental policies, e.g., with respect to taxes, subsidies, etc.
3. Here and in the following transportation costs should more precisely be understood as 'transfer costs' or 'trade costs', that is, they should include all the factors that drive a wedge between prices in different locations, for instance higher monitoring and marketing costs, and of course, when boundaries are relevant, import duties and non-tariff barriers to trade (see Baldwin 1994, p.43).
4. This is a fairly narrow, but not unusual definition. There have been many other important recent contributions to economics with a strong regional component, both theoretical and empirical. For reasons of space and coherence, I shall confine myself to new economic geography in the more limited sense.
5. These arguments for localization have received a lot of attention in the empirical literature recently. In particular, the importance of technological spillovers has been investigated by Jaffe (1989), Jaffe *et al.* (1993), Acs *et al.* (1994), Feldman (1994), Audretsch and Feldman (1996) and Audretsch and Stephan (1996). Also, more recently, additional reasons for localization have been analyzed in more detail. For instance, Schulz and Stahl (1996) argue that retailers in differentiated markets may seek proximity to each other even though this increases competition: such proximity reduces search costs for consumers with differentiated tastes, and therefore makes it more likely that they search for a retailer at all. A similar point has been made by Kim (1990): he argues that agglomeration generally reduces search costs on input markets. Finally, Stahl and Varaiya (1978) consider the signalling effects of agglomerations: the fact that some firms do business in an area may be interpreted by others as a signal of a favorable business climate, and it may induce them to follow.
6. As Krugman (1992) points out, similar ideas are captured in the concept of a region's market potential which is well-documented in geography. For instance, Harris (1954) describes this potential, a measure of the region's attractiveness for new manufacturing firms, as a weighted sum of the incomes of all regions, with weights that are decreasing with distance. As income depends positively on the presence of other manufactures, manufacturing attracts manufacturing.
7. Hoover was also well aware of this (e.g. Hoover 1948, ch.7).
8. Hotelling models that are familiar from the Industrial Organisation literature emphasize another centrifugal force: the desire of firms to stay away from others in order to reduce competition (d'Aspremont *et al.* 1979).
9. Without increasing returns, firms could produce the demand of each region locally without sacrificing scale advantages.
10. For instance, he developed a simple transportation cost argument which explains some degree of intra-industry trade.

11. A fully satisfactory treatment would include transportation activities as an input supplied by profit-maximizing agents in the economy, with transportation cost as an endogenous variable; see Ohlin (1933, ch. 8) for some thoughts on this issue.
12. For a discussion of this assumption, see Section 4.
13. Essentially, with a low elasticity of substitution, consumers value variety highly. Hence, firms will respond by producing a great number of varieties, and they will therefore forsake economies of scale. More precisely, free entry implies that price equals average costs. Profit maximization implies that the price is obtained from marginal costs by multiplication with the mark-up factor, which can be shown to be $\sigma/(\sigma - 1)$. Hence, the ratio of average costs to marginal costs equals $\sigma/(\sigma - 1)$ (see Krugman 1991c, p. 104).
14. Myrdal (1957, p. 28) argues similarly, but points out the abolishment of tariffs after Italian unification instead of a reduction of transportation costs in general.
15. For instance, transportation costs should not be too high, so that it is possible to serve both locations from one place.
16. See also Hoover 1948 (ch. 9) and Fuchs 1962 for detailed empirical work on these issues.
17. There are of course competing explanations, the simplest one of which relies on initial asymmetries between regions. For instance, if each region has sufficiently strong natural advantages (resources or closeness to export markets) in some industries, at least these industries will stay in an otherwise underdeveloped region.
18. In principle, preferences for proximity to the center may also exist. For instance, Matsuyama and Takahashi (1993) consider the idea that agglomeration advantages can result from the possibility of consuming non-traded local public goods that makes cities *more* attractive; they call this the 'bright light — big city' effect.
19. Recall that, in the two region model, workers moved towards the region offering the higher present real wage.
20. This resembles familiar results from partial equilibrium location theory (d'Aspremont *et al.* 1979, Tirole 1988, ch.7).
21. Continua are not always the appropriate unit of analysis in the study of locations. To model the locational choices of firms who are facing existing regional distributions of manufacturing, networks consisting of finitely many points and arcs can be more appropriate choice sets (see Thisse 1993).
22. Krugman also shows that the results sketched below generalize to the more realistic case of a two-dimensional object (the unit disk).
23. Fujita (1993) and Fujita and Krugman (1994) deal in more detail with non-monocentric equilibrium configurations.
24. For notational simplicity, transportation costs for exports are set to zero. This does not affect the main results.
25. Note that this is not inconsistent with the Krugman model of Section 3: There, a reduction in *transportation costs between* two regions could lead to an agglomeration in one of these regions. Here we are considering the transportation costs between the two regions and the rest of the world, so it is not surprising that the effect is different.
26. His model is based on Krugman (1991a), which considers the related problem of shifts between sectors within a region, given that there are Marshallian localization externalities in one of the sectors. There is an error in this model, which does, however, not affect the qualitative conclusions presented here (see Fukao and Benabou 1993).
27. This formulation makes the model applicable to situations where agglomeration advantages come from arbitrary sources. In particular, we could interpret the model as describing the development in a single industry, where Marshallian externalities of localization are important.

28. The size of the interval of initial shares that is compatible with this kind of dominance of expectations depends on the three parameters (interest rate, strength of linkages, moving costs) in the same way as above.
29. This differs slightly from the model by Brakman *et al.* (1994). Recall that there congestion resulted in negative externalities between firms.
30. It may of course still be relevant from an equity point of view. In the words of Rauch (1993 a, p. 844), '...certain regions or countries may be condemned to agricultural poverty, not because of lack of human and physical resources or poor governmental management, but simply because of bad luck'.
31. See for instance the development models of Murphy, Shleifer and Vishny (1989), the product standard models of Farrell and Saloner (1985) and the technology adoption models of Arthur (1989). Similar structures are also familiar from the natural sciences, where they come under the heading of 'self-organization' (Prigogine and Stengers, 1984).
32. For instance Glaeser *et al.* (1992), Ades and Glaeser (1995), Henderson *et al.* (1995), Audretsch and Feldman (1996), Ciccone and Hall (1996), Hanson (1996).
33. See, however, Martin and Rogers (1995).

References

- Acs, Z. J., Audretsch, D. B. and Feldman, M. P. (1994) R&D-spillovers and recipient firm size. *Review of economics and statistics*, 76, 336–345.
- Ades, A. F. and Glaeser, E. L. (1995) Trade and circuses: Explaining urban giants. *Quarterly Journal of Economics*, 110, 195–228.
- Arthur, W. B. (1989) Increasing returns, competing technologies and lock-in by historical events. *Economic Journal*, 99, 116–131.
- Arthur, B. (1990) Silicon valley locational clusters: When do increasing returns imply monopoly? *Mathematical Social Sciences*, 19, 235–251.
- Asilis, C. M. and Rivera-Batiz, L. A. (1994) Geography, trade patterns and economic policy. In *The Location of Economic Activity: New Theories and Evidence*, Consorcio da Zona Franca de Vigo and CEPR.
- Audretsch, D. B. and Feldman, M. P. (1996) The geography of innovation and production. *American Economic Review*, 86, 630–640.
- Audretsch, D. B. and Stephan, P. E. (1996) How localized are networks in biotechnology?. *American Economic Review*, 86, 641–652.
- Baldwin, R. E. (1994) *Towards an Integrated Europe*, CEPR, London.
- Brakman, S., Garretsen, H., Gigengack, R., Marrewijk, C. V. and Wagenvoort, R. (1994) *Congestion and Industrial Location*, mimeo, University of Groningen.
- Christaller, W. (1933), *Die Zentralen Orte in Suddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung der Siedlungen mit städtischen Funktionen*, Fischer, Jena.
- Ciccone, A. and Hall, R. E. (1996) Productivity and the density of economic activity. *American Economic Review*, 86, 54–70.
- d'Aspremont, C., J. Gabszewicz and J. F. Thisse (1979) On hotelling's stability in competition. *Econometrica*, 47, 1145–1151.
- Dixit, A. K. and Stiglitz, J. E. (1977) Monopolistic competition and optimum product diversity. *The American Economic Review*, 67, 297–308.
- Farrell, J. and Saloner, G. (1985) Standardization, Compatibility, and Innovation. *American Economic Review*, 76, 940–955.
- Feldman, M. (1994) *The Geography of Innovation*, Kluwer, Boston.

- Friedmann, J. and Alonso, W. (1964) *Regional Development and Planning*, MIT Press, Cambridge, MA.
- Fuchs, V. R. (1962) *Changes in the Location of Manufacturing in the United States since 1929*, Yale University Press, New Haven and London.
- Fujita, M. (1993), Monopolistic competition and urban systems. *European Economic Review*, 37, 308–315.
- Fujita, M. and Krugman, P. (1994) Urban systems and regional development. In *The Location of Economic Activity: New Theories and Evidence*, Consorcio da Zona Franca de Vigo and CEPR.
- Fukao, K. and Benabou, R. (1993) History versus expectations: a comment. *Quarterly Journal of Economics*, 58, 535–542.
- Gali, J. (1995) Expectations-driven spatial fluctuations, *Regional Science and Urban Economics*, 25, 1–19.
- Glaeser, E. L., Kauai, H. D., Scheinkman, J. A. and Shleifer, A. (1992) Growth in Cities. *Journal of Political Economy*, 100, 1126–1152.
- Hanson, G. H. (1994) Increasing returns and the regional structure of wages. In *The Location of Economic Activity: New Theories and Evidence*, Consorcio da Zona Franca de Vigo and CEPR
- Hanson, G. H. (1996) Localization economies, vertical organization, and trade. *American Economic Review*, 86, 1266–1278.
- Harris, C. (1954) The market as a factor in the localization of industries. *Annals of the Association of American Geographers*, 64, 315–348.
- Helpman, E. and P. Krugman (1985) *Market structure and foreign trade*, MIT Press, Cambridge, MA.
- Henderson, J. V. (1974) The sizes and types of cities. *American Economic Review*, 64, 640–656.
- Henderson, J. V. (1996) Ways to think about urban concentration: neoclassical urban systems versus the new economic geography. *International Regional Science Review*, 19, 21–36.
- Henderson, J. V., Kuncoro, A. and Turner, M. (1995) Industrial development in cities, *Journal of Political Economy*, 103, 1067–1090.
- Hirschman, A. (1963) *The strategy of economic development*, Yale University Press, New Haven.
- Hoover, E. (1948) *The Location of Economic Activity*, Mc-Graw-Hill, New York.
- Hotelling, H. (1929) Stability in competition. *Economic Journal*, 39, 41–57.
- Hunton, L. (1995) Can regional science find its place in the policy arena? *International Regional Science Review*, 18, 233–236.
- Isard, W. (1956) *Introduction to Regional Science*. Prentice-Hall, Englewood Cliffs, 1956
- Isserman, A. M. (1995) The history, status and future of regional science: an american perspective. *International Regional Science Review*, 17, 249–296.
- Isserman, A. M. (1996) It's obvious, it's wrong, and anyway they said it years ago? Paul Krugman on Large Cities. *International Regional Science Review*, 19, 21–36.
- Jaffe, A. B. (1989) Technological opportunity and spillovers of R&D: evidence from firms. Patents, Profits and Market Value. *American Economic Review*, 79, 957–970.
- Jaffe, A. B., Trajtenberg, M. and Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 63, 577–598.
- Kim, S. (1990), Labour heterogeneity, wage bargaining, and agglomeration economies. *Journal of Urban Economics*, 21, 109–126.
- Krugman, P. (1979) Increasing returns, monopolistic competition, and international trade, *Journal of International Economics*, 9, 469–479.
- Krugman, P. (1991a) History versus expectations. *Quarterly Journal of Economics*, 56, 652–667.

- Krugman, P. (1991b) Increasing returns and economic geography. *Journal of political economy*, 99, 183–199.
- Krugman, P. (1991c) *Geography and Trade*. MIT Press.
- Krugman, P. (1992) *A Dynamic Spatial Model*, NBER Working Paper #4219.
- Krugman, P. (1993a) First nature, second nature, and metropolitan location. *Journal of Regional Economics*, 129–144.
- Krugman, P. (1993b) On the number and location of cities. *European Economic Review*, 37, 293–298.
- Krugman, P. (1994) Complex landscapes in economic geography. *American Economic Review*, 84 (Papers and Proceedings), 412–416.
- Krugman, P. and Venables, T. (1995) Globalization and the inequality of nations. *Quarterly journal of economics*, 110, 857–880.
- Krugman, P. and Venables, T. (1996) Integration, specialization, and adjustment. *European Economic Review*, 40, 959–968.
- Livas-Elizondo, R. and Krugman, P. (1996) Trade policy and the third-world metropolis. *Journal of Development Economics*.
- Markusen, J. R. and Venables, A. J. (1996) The theory of endowment, intra-industry and multi-national trade. mimeo, London School of Economics.
- Marshall, A. (1920) *Principles of Economics*, 8th ed., MacMillan, London.
- Martin, P. and Rogers, C. A. (1995) Industrial location and public infrastructure. *Journal of International Economics*, 39, 335–351.
- Matsuyama, K. and Takahashi, T. (1993) *Self-Defeating Regional Concentration*. mimeo, Northwestern University.
- Molle, W. (1997) The regional economic structure of the european union: an analysis of long-term developments. In K. Peschel (ed.): *Regional Growth and Regional Policy within the Framework of European Integration*, Physica, Heidelberg.
- Murphy, K., Shleifer, A. and Vishny, R. (1989) Industrialization and the big push. *Journal of Political Economy*, 97, 1003–1026.
- Myrdal, G. (1957) *Economic Theory and Underdeveloped Regions*. Duckworth, London.
- Ohlin, B. (1933) *Interregional and International Trade*, Harvard University Press. Cambridge, MA.
- Prigogine, I. and Stengers, I. (1984) *Order out of Chaos*. New York, Bantam.
- Puga, D. and Venables, T. (1996) *The Spread of Industry: Spatial Agglomeration in Economic Development* Discussion Paper No. 279, Centre for Economic Performance, London.
- Rauch, J. E. (1993a) Does history matter only when it matters little? The case of city-industry location. *Quarterly Journal of Economics*, 58, 843–867.
- Rauch, J. E. (1993b) Review of geography and trade by Paul Krugman. *Journal of International Economics*, 34, 195–198.
- Schulz, N. and K. Stahl (1996) Do consumers search for the highest price? Equilibrium and monopolistic optimum in differentiated product markets. *RAND Journal of Economics*, 27, 542–562.
- Stahl, K. and Varaiya, P. P. (1978) Economics of information: examples in location and land use theory. *Regional Science and Urban Economics*, 8, 43–56.
- Thisse, J. F. (1993) Oligopoly and the polarization of space. *European Economic Review*, 37, 299–307.
- Tirole, J. (1988) *The theory of industrial organization*, MIT Press, Cambridge, MA.
- Venables, A. J. (1996) Equilibrium locations of vertically linked industries, *International Economic Review*, forthcoming.
- Woodford, M. (1986) Stationary sunspot equilibria in a financed constrained economy. *Journal of Economic Theory*, 40, 128–137.
- Yezer, A. (1995) Intellectual space for regional science. *International Regional Science Review*, 18, 153–157.