New Economic Geography: what about the N?*

Gianmarco I.P. Ottaviano and Jacques-François Thisse

10 September 2004

Abstract

Since its very appearance, probably due to its provocative name, New Economic Geography has stirred a debate on whether it is economic geography proper or rather geographical economics. In both cases, its real novelty has been questioned. We focus on this last issue. In particular, we argue that many of the NEG ideas have been around for a long time in the works of economic geographers and location theorists. However, NEG has the fundamental merit of having framed those ideas within a general equilibrium model encompassing most of these ideas. This has drawn economic geography and location theory from the periphery to the center of mainstream economic theory. More importantly, it has made already existing ideas more amenable to empirical scrutiny and policy analysis.

*We thank three referees, K. Behrens, G. Duranton, and T. Tabuchi for helpful comments. This research was supported by the Ministère de l'éducation, de la recherche et de la formation (Communauté française de Belgique), Convention 00/05-262. Both authors are also grateful to the RTN Program of the European Commission for financial support.

Gianmarco I.P. Ottaviano, Università di Bologna, Dipartimento di Scienze Economiche, Strada Maggiore 45, I-40125 Bologna, Italy. Tel. +39-051-2092600. Fax. +39-051-209266. Email: ottavian@economia.unibo.it. FEEM and CEPR.

Jacques-François Thisse, CORE-Université catholique de Louvain, 34 Voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium. Tel. 32-10-474311. Fax. +32-10-474301. Email: thisse@core.ucl.ac.be. CERAS and CEPR.

1 Introduction

That economic activities are unevenly distributed in space is self-evident. To a large extent it is hardly surprising as different areas of the planet exhibit different climates, different degrees of accessibility, and different endowments of productive resources. All these features, which are bundled under the common label of *first nature*, play an important role in explaining economic history and especially the early stages of economic development until the Industrial Revolution included. However, as pointed out by outstanding scholars such as Fernand Braudel and Carlo Cipolla, a driving force of human history has always been the search for safe and cheap ways to reach and ship around raw materials. In other words, since the very start, human kind has tried to soften the constraint of first nature, thus generating a spatial distribution of economic activities that, in many cases, is largely independent from natural advantage. Hence the concept of *second nature*, which emerges as the outcome of human beings' actions to improve upon the first one.

Second nature is the research focus of geographical economics. At a high level of abstraction, geographical economics uses first nature as a 'control variable', that is, it isolates second nature as the component of the spatial variation of economic activities that cannot be explained in terms of first nature. In other words, it is not the overall variation that geographical economists want to understand but rather the 'conditional variation' given identical first nature. The question, then, becomes whether such a residual is large enough to motivate an intellectual effort.

As casual observation would have it (think of the many clusters of activities that are little dependent on natural advantage such as the metropolitan area of Tokyo or the Silicon Valley), the residual turns out to be very large. Indeed, so large that even trade in goods and factors, which as stated above has been a driving force of economic history, can be hardly understood without a theory of the location of economic activities. This is what Ohlin (1933) first argued, thus setting the stage for the so-called 'new economic geography' (henceforth, NEG). NEG has been pioneered by three authors, namely Fujita (1988), Krugman (1991a) and Venables (1996). They all use general equilibrium models with monopolistic competition à la Dixit and Stiglitz (1977) to study the effect of different degrees of transportability of goods and factors on industry location depending on the extent of returns to scale and product differentiation. Their and other NEG scholars' results have been recently synthesized in Fujita et al (1999), Baldwin et al (2003) and Ottaviano and Thisse (2004).

NEG has not passed unobserved. The very name 'new economic geography' seems chosen ad hoc to stir a debate (see, e.g. Isserman, 1996 and Martin, 1999): is NEG 'economic geography' proper or rather 'geographical economics'? In any case, is there anything really 'new' in it? Given what we have written above, it should be clear by now what we think about the first question. *Methodologically NEG belongs to economics.* Nonetheless, in terms of its subject matter, it cannot be considered alien to economic geography proper. Having said that, in the present paper we focus on the second question. In particular, we argue that many of the NEG ideas have been around for a long time in the works of economic geographers and location theorists. However, as already noticed by Brakman et al (2001, p 323), NEG has the fundamental merit of having framed most of those ideas within a general equilibrium model. This has drawn economic geography and location theory from the periphery to the center of mainstream economic theory. More importantly, it has made already existing ideas more amenable to empirical scrutiny and policy analysis. That has been achieved by connecting trade and location theories, a research target put forward several years ago by Ohlin et al (1977). Such an integration has been made possible because, as observed by Krugman (1993, p 111) himself: "the analytical underpinnings of the 'new trade theory' are much more friendly to location theory than those of traditional trade models". One of our objectives is to show why this is so. In addition, through the adoption of monopolistic competition, NEG has overcome the impossibility for the competitive paradigm of standard neoclassical economics to explain cities, local specialization, and trade as second nature. The reason being the presence of space-specific nonconvexities that prevent the existence of a nontrivial competitive equilibrium.

As we see it, the key question of location theory is:[1] why do some particular production activities (such as plants, offices, public facilities, etc.) choose to establish themselves in some particular places of a given space? Such a question can be divided in two sub-questions concerning either a firm or an industry respectively: (i) why does an isolated firm face a location problem and how does it solve it? and (ii) how are firms' locational decisions intertwined? In what follows, we will show that earlier contributions in location theory have developed many of the ideas used in NEG to answer the above questions. To this end, we will borrow at will from Baldwin et al (2003), Fujita and Thisse (2002), Krugman (1995), as well as Ottaviano and Thisse (2004). Let us also make it clear that we do not

[1]Traditionally, the location of consumption activities is studied in urban economics. It is worth noting that NEG integrates both types of activities.

intend to provide an exhaustive survey of what had been accomplished by geographers, economists and regional scientists before NEG. More modestly, we want to show that most of the main ingredients were already there. This will allow us to assess why there is an 'N' in NEG.[2]

The paper is organized as follows. Section 2 surveys the theoretical explanations of the location of a firm, highlighting the seminal contributions of Weber and Lösch. Section 3 deals with industry location. It presents Starrett's impossibility theorem and the alternative strategies to overcome its constraints put forth by Marshall, Thünen, Ricardo, Ohlin, Hotelling, and Myrdal. Section 4 concludes by discussing why NEG models represent a useful addition to already existing theories.

2 The location of a firm

In this section, we start with a discussion of the location choice faced by a firm in isolation from any competitor both in its product and factor markets. In the next section we will see how such choice is affected by competition.

2.1 Weber and the minisum location problem

The oldest formal analysis of the location of a firm is the minisum location problem whose mathematical formulation can be traced back to Fermat early in the 17th century. The objective is to locate a plant in the plane by minimizing the weighted sum of Euclidean distances from that plant to a finite number of sites corresponding to the markets where the plant purchases its inputs and sells its outputs (the minisum problem). In the sum, the weights represent the quantities of inputs and outputs bought and sold by the plant, multiplied by the appropriate freight rates (Weber, 1909). This problem has been given an interpretation in terms of forces that should be appealing to the tenet of NEG. Each point is the origin of a force which, because of the advantage of proximity, attracts the facility toward this point with an intensity proportional to its weight. A system of conflicting forces, therefore, emerges and the optimal location is the equilibrium point of this

[2]As observed by a referee, international trade theory is another forerunner of NEG. This should not come as a surprise as we have just seen that long ago Ohlin called for an integration of location and trade theories. Yet, it is fair to say that trade economists have not shown much willingness to acknowledge the connection and continuity between their work and location theory (see Krugman, 1993, 1995, for two noticeable exceptions). This is why we have chosen to stress the connections of NEG with location theory, while mentioning its connections with trade theory whenever relevant to our purpose.

system. By using negative weights, Tellier and Polansky (1989) have added repulsive forces to the minisum problem, the solution of which is now the outcome of the interplay between attractive (agglomeration) and repulsive (dispersion) forces, very much as in NEG.

Even though a great deal of work devoted to firm location theory has concentrated on continuous spaces, an alternative and more relevant line of research focuses on the explicit modeling of the transportation network. More precisely, a *network* is a connected set of the plane defined by the union of a finite number of arcs of a well-defined length such that any two arcs intersect at most at their extremities. The extremities of these arcs are called *vertices* of the network. Without loss of generality, we may assume that any market is a vertex of the network, and that any vertex which is not a market is a node, i.e. a point which is the extremity of at least three arcs of the network. A path in a network is a sequence of arcs that can be traversed continuously without retracing any arc. The distance between any two points of the network is then given by the length of the shortest path linking these two points. Using this framework, Hakimi (1964) has shown that the set of vertices of the network always contains an optimal solution to the minisum problem. This result has several interesting implications. First, it shows that the location of a facility on a network is of a discontinuous nature because the search for an optimal location may always be restricted to a finite set of points. This concurs with Isard (1956, 251) for whom:

> "[s]ubstitution among transport inputs is not in the small but rather in the large, entailing geographic shifts over substantial distance from one focal point to another."

and suggests some form of inertia in firms' locational behavior. Second, the locations to be considered are very distinctive in the sense that they are markets, or nodes, or both. What kind of point will ultimately be selected hinges on the shape of the network, the spatial distribution of markets and the vector of weights. Nevertheless, one can derive special meaningful results. For example, Witzgall (1964) proves that, if there exists a site (called a 'dominant market') such that its weight is greater than or equal to the sum of the weights of all the other sites, then such site is the solution to the minisum problem. This result may explain the locational decision made by seemingly different firms to set up in a large metropolitan area.

Since the 1960s, the more standard microeconomic approach of profit maximization has taken the place of the least transport cost approach. This has allowed to understand the influence of additional economic and geographic factors on firm location. For instance, Sakashita (1967) considers the case of freight rates that are constant or taper off with distance, a condition that characterizes all modern transport technologies. He demonstrates that, in the case of a segment connecting the market and the input source, the firm always chooses to set up at one of the two endpoints and never at an intermediate point. This result has been extended to the case of a network and of several input and output markets (Louveaux et al, 1982).

The fact that the location of a firm follows a discontinuous logic is the counterpart of the inertia observed in locational decisions. Thus, models of firms' location suggest a fundamental feature of the space-economy:

Claim 1 - Bang-bang' behavior - *Firms' locational behavior is either sluggish or catastrophic.*

2.2 Increasing returns vs transport costs

The minisum problem rationalizes the location choice of plants assuming away the crucial decision firms make in the real world on how many plants to run. In other words, in supplying (or sourcing from) a new market a firm can choose between shipments from (or to) some pre-existent distant plant or sales from (or to) a new local plant. The choice is not trivial provided transportation is costly and there are increasing returns to scale at the plant level. Without transport costs, space is immaterial. Increasing returns lead the firm to concentrate its production in a few plants, whereas transport costs raise the issue of where to locate those plants. In Koopmans's (1957, 154) words:

> "without recognizing indivisibilities - in human person, in residences, plants, equipment, and in transportation - urban location problems, down to those of the smallest village, cannot be understood."

That increasing returns to scale generate such indivisibilities has been recognized by various scholars interested in the formation of human settlements. Indeed, the trade-off between increasing returns and transport costs may take quite different forms, thus making the argument applicable to a wide range of scenarios. For instance, in his *Social Life in Ancient Egypt* published in 1923, W. M. Flinders Petrie states (the quote is borrowed from Ullman, 1941):

> "It has been noticed before how remarkably similar the distances are between the early nome capitals of the Delta (twenty-one miles on an average) and the early cities of Mesopotamia (averaging twenty miles

> apart). Some physical cause seems to limit the primitive rule in this way. Is it not the limit of central storage of grain, which is the essential form of early capital? Supplies could be centralised up to ten miles away; beyond that the cost of transport made it better worthwhile to have a nearer centre."

This is reminiscent of what Lösch (1940) writes about twenty years later:

> "We shall consider market areas that are not the result of any kind of natural or political inequalities but arise through the interplay of purely economic forces, some working toward concentration, and other toward dispersion. In the first group are the advantages of specialization and of large-scale production; in the second, those of shipping costs and of diversified production." (p.105 of the English translation)

It is worth noting that the very same trade-off has also been modeled independently by Kuehn and Hambuger (1963) and Stollsteimer (1963) in a planning context. Those authors consider a situation in which the spatial distribution of requirements for a particular commodity is given, the costs incurred for locating the production facilities are fixed, and shipments from the facilities to the consumers are costly. Within this framework, they determine the number and locations of facilities that minimize the sum of production and transport costs (see, e.g. Hansen et al, 1987, for a survey of this literature).

We thus have:

Claim 2 - Proximity vs scale - *Firm location weighs transport cost savings from proximity to customers and suppliers against production cost savings from large scale plants.*

3 The location of an industry

The insights presented so far consider the location decision made by a firm isolated from any competition with other firms. Few industries, however, feature such local monopolists. Therefore, to better describe reality, we need to add more firms to the picture. Unfortunately, such passage from firm to industry location is not smooth because, in an economy sterilized from first nature, the locational decisions of several firms can hardly be studied within a perfectly competitive framework.

3.1 Starrett and the breakdown of the competitive price mechanism

That perfect competition is essentially incompatible with spatial interdependence in the absence of first nature differences across sites has been very elegantly proven by Starrett (1978). To illustrate his result, we use a simple example with two locations forming a homogeneous space (see Fujita and Thisse, 2002, Ch. 2 for a more detailed discussion). By a homogeneous space, we mean that technological alternatives and preferences are the same in both locations. There are two agents, one firm and one consumer who trade with each other and consume land. The consumer provides an amount x of labor to the firm; the firm uses that labor and one unit of land to produce a quantity y of a good for the consumer who also consumes one unit of land (land rents go to the consumer). Land is available in two distinct and separate locations, A and B, such that resources are required to ship labor and goods from one location to the other (transport costs). All markets are perfectly competitive.

If the amount of land available at A, say, is larger than or equal to 2 units, then both the firm and the consumer are able to locate together at location A. Hence, the land rent in A is nonnegative ($R_A \geq 0$) while the rent at the vacant location B is zero ($R_B = 0$). When the land rent R_A is not too high relative to transport costs, this no-transportation configuration is a perfectly competitive equilibrium. On the other hand, if the amount of land in A is less than 2 units, either the firm or the consumer must locate in B so that the land rent in B becomes also nonnegative ($R_B \geq 0$). However, this configuration, which entails some positive transportation cost between locations A and B, is not a perfectly competitive equilibrium. To prove it, we must show that, when the firm and the consumer reside in different locations, at least one of them has an incentive to move, i.e., by changing location either the firm can increase its profits or the consumer can decrease her net expenditures.

The proof is by contradiction. For concreteness, take a situation where the firm is in A while the consumer is in B. The mirror situation can be analyzed in a symmetric way. Suppose that the configuration with the firm in A and the consumer in B is indeed an equilibrium and let p_A, w_A, R_A (p_B, w_B, R_B) be the price of the good, the wage and the land rent in A (B) that correspond to such an equilibrium. Price and wage differences between the two locations measure the wedge between what the buyer pays and what the seller receives net of transport costs. Specifically, $p_B - p_A$ is equal to the transport costs of one unit of the good from A to B, whereas $w_A - w_B$

stands for the extra pay the consumer must be given in order to work in A instead of B.

The firm's profit, which equals the value of output minus the wage bill and the land rent, is $\Pi_A = p_A y - w_A x - R_A$. Consider now what profit the firm would make if it were to produce in B. In so doing, keep in mind that the essence of the theory of competitive markets is that agents make their production or consumption decisions on the basis of prices alone and that these prices are unaffected by their decisions (price-taking behavior). For example, adding considerations such as the firm in A recognizes that if it locates in B will change prices and land rent there, takes us away from that theory. Since space is homogenous, the firm may keep the same input-output structure and, therefore, would gain $\Pi_B = p_B y - w_B x - R_B$. As a result, its incentive to move (which may be positive, negative, or zero) is

$$I_{firm} = \Pi_B - \Pi_A = (p_B - p_A)y - (w_B - w_A)x - (R_B - R_A) \tag{1}$$

As to the consumer, her net expenditure is given by the value of consumption minus wage income plus the net land rent (total land rent $R_A + R_B$ minus the rent of the land she consumes). Therefore, in B net expenditures are $X_B = p_B y - w_B x - (R_A + R_B - R_B) = p_B y - w_B x - R_A$ while in A she would spend $X_A = p_A y - w_A x - R_B$ while keeping the same consumption-labor structure due to space homogeneity. Thus, her incentive to move is

$$I_{consumer} = X_B - X_A = (p_B - p_A)y - (w_B - w_A)x + (R_B - R_A) \tag{2}$$

If the configuration with the firm in A and the consumer in B were indeed an equilibrium, neither agent would be better off by changing location, that is, both I_{firm} and $I_{consumer}$ would be nonnegative. This implies that, if either I_{firm} or $I_{consumer}$ or both are positive i.e. $I_{firm} + I_{consumer} > 0$, then that configuration is not an equilibrium. By adding up (1) and (2), we obtain

$$I = I_{firm} + I_{consumer} = 2[(p_B - p_A)y + (w_A - w_B)x] > 0 \tag{3}$$

which is just equal to twice the total transport costs borne by the firm and the consumer since, as already argued, at the competitive equilibrium the price difference $p_B - p_A$ (the wage difference $w_A - w_B$) must be equal to the transport cost of shipping the good (labor) between the two locations. In other words, when the two agents are separate, the incentive to change location is of the same order of magnitude as transport costs. Therefore, in this example, a competitive equilibrium, when it exists, must be such that both agents co-locate.

Using a more general framework in which the transport sector is explicitly accounted for and allowing for variable land consumption by all agents, Starrett (1978) shows the following result without making any convexity assumption regarding preferences and technologies.

Claim 3 - The Spatial Impossibility Theorem - *Assume an economy with a finite number of locations and a finite number of consumers and firms. If space is homogeneous, transport is costly and preferences are locally nonsatiated, then there is no competitive equilibrium involving transportation.*

What does it mean? If economic activities are perfectly divisible, each activity can operate at an arbitrarily small level without any loss of productive efficiency. Thus, a competitive equilibrium exists and is such that each location operates as an autarchy. However, once economic activities are *not* perfectly divisible, they have an address and the transport of some goods between addresses becomes unavoidable. In Starrett's (1978, 27) words:

> "as long as there are some indivisibilities in the system (so that individual operations must take up space) then a sufficiently complicated set of interrelated activities will generate transport costs"

that is, the fact that shipping goods is costly and that agents have an address generates nonconvexities in production or consumption sets that prevent the competitive setting from handling such a system of exchanges.

As simple as it may look, the logical implications of the Spatial Impossibility Theorem are nonetheless far reaching. In particular, it implies that understanding the spatial distribution of economic activities cannot escape some deviation from the set of assumptions that stand behind the theorem. This allows us to present the various theories of industry location as alternative settings. In particular, in the wake of Marshall, modern urban economics assumes that local externalities exist and are many. Other theories assume instead that space is heterogeneous. This is the case of land use models à la Thünen and neoclassical models of international trade. Finally, spatial competition theory and economic geography rest on imperfectly competitive markets.

3.2 Marshall and local externalities

According to Marshall, second nature may be explained by mutually reinforcing external effects. In his own words (Marshall, 1890, 1920, 225):

> "When an industry has thus chosen a location for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighborhood to one another....A localized industry gains a great advantage from the fact that it offers a constant market for skill....Employers are apt to resort to any place where they are likely to find a good choice of workers with the special skill which they require; while men seeking employment naturally go to places where there are many employers who need such skills as theirs and where therefore it is likely to find a good market."

More generally, Marshallian externalities arise because of (i) mass-production (or, equivalently, increasing returns at the firm level as discussed below), (ii) the formation of a highly specialized labor force and the production of new ideas, both based on the accumulation of human capital and face-to-face communications, (iii) the availability of specialized input services, and (iv) the existence of modern infrastructures.

Building on Weber (1909, ch. 5), Hoover (1936, ch. 6) has proposed what became the now standard classification of agglomeration economies: (a) localization economies, which are external to firms but internal to an industry; and (b) urbanization economies, which are external to industries and depend on the overall scale as well as scope of the economic activity in one location. Localization economies refer to Marshallian externalities of type (ii) and (iii), whereas urbanization economies cover the Marshallian externalities of types (ii), (iii) and (iv) since they typically depend on the presence of public infrastructures and on the size of the agglomeration. This classification has been used extensively in empirical studies as surveyed in Rosenthal and Strange (2004).

While such a classification mainly stresses the supply side of the economic system, agglomeration economies also operate on the demand side. For instance, Haig (1926) argues that people value the large assortment of consumption goods and services offered by cities (*varietas delectat*) so much that migration from a small town to a large city may be explained purely by a consumption motive. In Stahl's (1983, 318) words:

> "in a large city an individual may derive a higher utility from spending a given amount of income than in a small town ...even if the prices for commodities obtainable at both locations are higher in the former than in the latter."

Along the same lines, Lampard (1955) concurs with Haig in that "each city serves a variety of social purposes and meets an array of human needs", and does not just provide goods (an idea modeled by Beckmann in 1976). He also made it clear that a variety-like argument similarly applies to intermediate goods used by firms, while "the principal function of the city today in terms of employment it creates is the provision of services rather than manufactures."

To sum up:

Claim 4 - Localized externalities - *Under perfect competition, industry location can be explained in terms of localized production or consumption externalities.*

3.3 von Thünen and the role of marketplace

The Spatial Impossibility Theorem tells us that the price mechanism does not work in a homogeneous spatial economy. Therefore, such result may not hold in the presence of spatial heterogeneity. One example is the model designed by von Thünen (1826) to explain the pattern of agricultural activities around pre-industrial cities. von Thünen considers a space represented by a plain on which land is homogeneous in all respects except for a marketplace (e.g. a town) where all economic transactions take place. In this framework the allocation of land to the different economic activities (e.g. crops) is shown to emerge as the equilibrium outcome of a perfectly competitive land market. In other words, the existence of a center is sufficient for a competitive land market to structure the use of space by different activities. Consider, for instance, the basic set-up in Fujita and Thisse (2002, ch. 2) in which land is the only factor of production, whereas transport costs are linear in distance. Then, if activities use similar amounts of land per unit of output, the goods facing higher transport costs are produced closer to the marketplace. Conversely, if goods face similar transport costs, activities that are relatively land intensive are located closer to the marketplace.

Even though von Thünen also assumes both land and activities to be perfectly divisible, the key assumption that allows for an analysis within the competitive paradigm is that transactions among firms and households have to occur at a given marketplace. Such an exogenously located marketplace breaks the homogeneity of space, thus removing the locational indeterminacy highlighted by Starrett. The reason is that, as discussed in the foregoing example, the source of indeterminacy is that in a homogeneous space all that matters for the location of an activity is the location of other activities. Under this respect, Thünen's marketplace serves as an 'anchor' to otherwise undetermined locational choices.

We thus have:

Claim 5 - Marketplaces - *Under perfect competition, industry location can be explained in terms of geographical accessibility to given markets for goods and factors.*

3.4 Ricardo and comparative advantage

Spatial homogeneity is also rules out in trade models based on the concept of comparative costs or comparative advantage: countries specialize in the production of the goods in which their opportunity cost, in terms of the implicit sacrifice of other foregone goods, is lowest.[3]

Even though it is mainly known as Ricardian theory, the credit for the classical model of comparative advantage should be equally shared by Mill (1821), Ricardo (1817), and Torrens (1815). Its most clear formulation can be found in two sentences by Mill, as quoted by Irwin (1996, ch. 6):

> "When a country can either import a commodity or produce it at home, it compares the cost of producing at home with the costs of procuring from abroad; if the latter cost is less than the first, it imports. The cost at which a country can import from abroad depends, not upon the cost at which the foreign country produces the commodity, but upon what the commodity costs which it sends in exchange, compared with the cost which it must be at to produce the commodity in question, if it did not import it".

Why then do comparative costs differ across countries? The answer is because space is heterogeneous in terms of technological alternatives. In Torrens' words (the quote is borrowed from Irwin, 1996, ch. 6):

> "If England should have acquired such a degree of skill in manufactures, that, with any given portion of her capital, she could prepare a quantity of cloth, for which the Polish cultivator would give a greater quantity

[3]Our taxonomy of traditional trade models follows Helpman and Krugman (1985) as well as Irwin (1996). According to the latter, "The classical theory developed by Ricardo and others developed around the real costs of production in terms of one factor of production (labor). Neoclassical theory (associated with the Swedish economists Eli Heckscher and Bertil Ohlin and the American Paul Samuelson) focused more on different factor endowments and opportunity costs between countries as driving international trade" (footnote 5, p 177). Specifically, "the theory of comparative costs, or comparative advantage" is "the most important analytical contribution of classical economics relating to the free trade doctrine" (p. 90-91). On the other hand, "the core of modern analysis of trade is the factor proportion theory - the Heckscher-Ohlin model and its extensions" (Helpman and Krugman, 1985; p. 11).

> of corn, than she could, with the same portion of capital, raise from her own soil, then, tracts of her territory, though they should be equal, nay, even though they should be superior, to the lands of Poland, will be neglected; and a part of her supply of corn will be imported from that country".

For classical trade economists it is thus localized knowledge, embodied in the skills of the local population, that breaks spatial homogeneity. Therefore, we have:

Claim 6 - Comparative advantage - *Under perfect competition, industry location can be explained in terms of localized technological knowledge.*

3.5 Ohlin and factor proportions

The neoclassical theory of international trade, heralded by Heckscher (1918) and Ohlin (1933), departs from its classical antecedent by ruling out technological differences among countries. Spatial heterogeneity stems instead from the assumptions that natural resources, and more generally productive factors, are unevenly distributed across countries. Their international immobility generates different relative factor prices under autarchy and this determines the pattern of specialization once trade is liberalized: a country specializes in the production and exportation of the goods that are relatively intensive in its relatively abundant resources.

To summarize:

Claim 7 - Factor proportions - *Under perfect competition, industry location can be explained by the uneven distribution of primary production factors.*

When compared to location theory, one of the most distinctive features of both classical and neoclassical trade theories is that in the latter countries are generally considered as being dimensionless. Such a simplifying assumption has long been considered as a handicap as recognized by Ohlin himself (1968, 97, emphasis in the original):

> "*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*".

The reason being that changes in the transportability of commodities as well as in the mobility of factors *between and within* countries affect the location of economic activities, the distribution of labor and capital, the geography of demand and, therefore, the pattern of trade (Ohlin, 1967, Ch. XI-XII). Indeed, even though Ohlin's contribution is essentially identified with his spaceless factor proportions model (Ohlin, 1967, Part Two, 49-93), he explicitly discusses how the explicit consideration of space affects (and, sometimes, even reverses) some of the main conclusions of the neoclassical paradigm of trade theory (Ohlin, 1967, Part Three, 97-165). Unfortunately, the robustness of Ohlin's spatial insights is often undermined by the lack of a model dealing with both increasing returns and imperfect competition (Ohlin, 1967, Ch. III). This introduces the next topic.

3.6 Hotelling and spatial competition

So far we have seen how the deadlock stressed by the Spatial Impossibility Theorem has been overcome by introducing either nonmarket interactions among agents or spatial inhomogeneities. Both strategies allow one to perform meaningful analyses without departing from perfect competition. The last research strategy we discuss takes an alternative route by assuming instead that firms are not price-takers in their output markets.

In the wake of Hotelling (1929), it is now generally accepted that competition for market areas is a force that would lead sellers to congregate (the so-called principle of minimum differentiation). The two ice-cream men problem provides a neat illustration. Two merchants selling the same ice-cream at the same fixed price compete in location for consumers who are uniformly distributed along a beach; each consumer purchases one cone of ice-cream from the nearer firm. As shown by Lerner and Singer (1937), in equilibrium the two ice-cream sellers choose to locate back to back half way along the beach. This is due to the 'market area effect': each seller's profit is an increasing function of the fraction of the beach to which she has privileged access.

However, the principle of minimum differentiation no longer holds once sellers are allowed to compete in prices. The reason is that the desire to relax price competition pushes the sellers to separate along the beach and this always dominates, at least locally, the market area effect (d'Aspremont et al, 1983). For example, Hotelling considers a situation where sellers first simultaneously choose their locations and afterwards their prices. As established by d'Aspremont et al (1979), if the transport costs are quadratic, a unique price equilibrium exists for any location pair and the two firms wish to set up at the endpoints of the beach. More generally, whether sellers cluster half way along the beach or separate depends on the degree of product differentiation as well as the level of transport costs: the less similar the products and the lower the transport costs, the more likely the clustering (de Palma et al, 1985). The reason is weaker price competition and a stronger market area effect.

Finally, once the mobility of customers along the beach is allowed for, sufficiently differentiated products and/or sufficiently low transport costs lead to a spatial equilibrium in which both sellers and customers distribute themselves according to two bell-shaped curves that sustain each other (Papageorgiou and Thisse, 1985).

The foregoing insights can be summarized as follows:

Claim 8 - Spatial competition and market area - *Under imperfect competition, industry location can be explained in terms of the search for privileged access to customers and the desire to relax competitive pressures by other firms.*

Another result, which is reminiscent of gravity theory, is that the number of consumers patronizing a cluster grows with the number of firms in it. For example, when consumers know which varieties are available in the market, but are not aware of which variety is offered where and at which price, they must engage in a costly search process. When firms are spatially dispersed, search is particularly expensive for consumers as they have to travel to many different sites. Travel costs are much lower when firms are clustered. Indeed, after incurring the travel cost to the cluster, consumers can visit any local firm at zero additional cost. Thus, the very same firm is more likely to be visited, and so faces higher expected demand, when clustered than otherwise (see Stahl, 1987, for more details). This generates an agglomeration force that is very much in the spirit of the market size effect featured by the core-periphery model developed by Krugman.

In his review of Chamberlin's (1933) book, Kaldor (1935) claims that a firm affects directly the sales of its neighboring competitors, but not distant ones. The impact of its price reduction is, therefore, not symmetric across its competitors. In other words, there are good reasons to believe that competition across locations is inherently oligopolistic (Eaton and Lipsey, 1977; Gabszewicz and Thisse, 1986). Unfortunately, models of spatial competition are plagued by the frequent nonexistence of an equilibrium in pure strategies (Gabszewicz and Thisse, 1992). Thus, research has faced a modeling trade-off: to appeal to mixed strategies, or to use monopolistic competition in which interactions between firms are weak. For the sake of simplicity, NEG has retained the second option, which is not unreasonable once we address spatial issues at a macro-level. In addition, models of monopolistic competition have shown a rare ability to deal with a large variety of issues related to economic geography, which are otherwise unsatisfactorily treated by the competitive paradigm (Matsuyama, 1995). However, it should be kept in mind that spatial competition should not be missed at the micro-level, if not at the macro-level (see, e.g. Pinkse et al, 2002).

3.7 Myrdal and circular causation

The result according to which, under imperfect competition, the spatial distributions of customers and firms sustain each other highlights the possibility of the economic landscape being molded by feedback mechanisms. Papageorgiou and Thisse (1985, 20) describe such a possibility by these words:

> "Households are attracted by places where the density of firms is high because opportunities there are more numerous ... Firms are attracted to places where the density of consumers is high because there the expected volume of business is large."

In the same spirit, research conducted in migration theory raises the following question: do people follow jobs or do jobs follow people? Since the seminal work of Muth (1971), it has been argued convincingly that this is an egg-and-chicken problem.

The role of circular causation among the location decisions of different agents has been mainly stressed in terms of regional divergence. In particular, it has been argued for long that growth is localized, the reason being that technological and social innovations tend to be spatially clustered while their diffusion across places would be slow. For example, Hirschman (1958, 183) claimed that:

> "we may take it for granted that economic progress does not appear everywhere at the same time and that once it has appeared powerful forces make for a spatial concentration of economic growth around the initial starting points."

This echoed by Myrdal (1957, 26):

> "The main idea I want to convey is that the play of the forces in the market normally tends to increase, rather than to decrease, the inequalities between regions"

as well as by Kaldor (1970, 340) who stresses the role of transport costs in sustaining circular causation in a way that strongly reminds the results of spatial competition models:

"As communication between different regions becomes more intensified (with improvements in transport and marketing organization), the region that is initially more developed industrially may gain from the progressive opening of trade at the expense of the less developed region whose development will be inhibited by it."

Circular causation has an important place in the work of geographers too. For example, Harris (1954, 315) writes:

> "The interrelationship between this [manufacturing belt] and other manufacturing areas and location of markets has been reciprocal; manufacturing has developed partly in areas or regions of largest markets and in turn the size of these markets has been augmented and other favourable conditions have been developed by the very growth of this industry."

The reasons why larger markets attract firms are described by Harris (1954, 326) along Marshallian lines:

> "Areas of high market potential furnish especially suitable conditions for the development of manufacturing. Industries in which economies of scale are important find in the immense nearby markets a particularly favorable environment. The existence of a large and diversified labor force, the presence of many specialized services, the ease of obtaining components or sub-assemblies nearby, the presence of large industrial markets for new parts and gadgets, the ability to deliver quickly to the market, and a host of other factors reinforce their transport advantages.".

whereas the reason why more developed areas attract people is neatly discussed by Haig (1926, 185-186), according to whom:

> "The great bulk of population...must work and must consume most of what they earn where they earn it. With them consumption and production is practically a simultaneous process and must be carried on for the most part in the same place. To them location is of interest both in its effects upon production and in its effects upon consumption."

That is, more people implies more income and, therefore, a larger local market.

A natural implication of circular causation is that even minor events can generate large permanent regional imbalances. This neat idea is put forward by Myrdal (1957, 26-27):

> "Within broad limits, the power of attraction of a centre has its origin mainly in the historical accident that something was once started there, and not in a number of other places where it could equally well have been started, and that start met success."

Thus, we have:

Claim 9 - Circular causation - *Firms and households locational choices sustain each other. As a result, the location of industry can sometimes be explained as the result of historical accidents.*

Note that circular causation may be driven by coordinated expectations about future changes in industry location, which are then self-fulfilling (Rosenstein-Rodan, 1943; Krugman, 1991b).

4 Conclusion

In the previous sections, we have provided a bird's eye view of the legacy of location theory, which can be summarized in five points:

(I) the economic space is the outcome of a trade-off between various forms of increasing returns and different types of mobility costs;

(II) price competition, high transport costs and land use foster the dispersion of production and consumption;

(III) firms are likely to cluster within large metropolitan areas when they sell differentiated products and transport costs are low;

(IV) cities provide a wide array of final goods and specialized labor markets that make them attractive to consumers/workers;

(V) agglomerations are the outcome of cumulative processes involving both the supply and demand sides.

Consequently, the space-economy has to be understood as the outcome of the interplay between agglomeration and dispersion forces - an idea put forward by geographers and regional scientists long ago - within a general equilibrium framework accounting explicitly for market failures - a need neglected by most economic geographers.

Those five points capture also the main ingredients of NEG as classified by Baldwin et al (2003). These include: circular causality, endogenous asymmetry, catastrophic agglomeration, locational hysteresis, self-fulfilling expectations, and home-market magnification. The first is our circular causation. Endogenous asymmetry means that, as highlighted in Claim 9, small historical accidents may generate regional imbalances even in an otherwise homogeneous space. Catastrophic agglomeration means that, as in Claim 1, firm location changes in a discontinuous way. Locational hysteresis implies that, as in Claim 9, spatial development can be locked in by historically inherited patterns. Self-fulfilling expectations are to be contemplated as a caveat to Claim 9. Finally, whereas Claims 2 and 5 stress the attraction of firms to larger markets, home-market magnification gets stronger as transport costs fall. This concurs with the insights of spatial competition models in Claim 8 and of Kaldor's cite on the effects of regional integration.

The key features of NEG were, therefore, already there and there is, hence, little new about them. The crucial aspect, however, is that before NEG those features were not knitted together and were often available only in specialized studies. In addition, they were (at best) developed within partial equilibrium models. Therefore, *what was missing was a general equilibrium framework with imperfect competition connecting these various insights and allowing for a detailed study of their interactions.*

The importance of microeconomic mathematical models can be understood by taking, for instance, the following analysis by von Thünen on whether or not industrial firms are better off by locating in major cities (especially in the capital). von Thünen (1826) starts by describing the main dispersion forces at work (the quotes are borrowed from Fujita and Thisse, 2002, ch. 1):

> " 1. Raw materials are more expensive than in the country towns on account of the higher cost of transport. 2. Manufactured articles incur the cost of haulage to the provincial towns when they are distributed to the rural consumers. 3. All necessities, especially firewood, are much more expensive in the large town. So is rent for flats and houses, for two reasons (1) construction costs are higher because raw materials have to be brought from a distance and are consequently more expensive, and (2) sites that may be bought for a few thalers in a small town are very dear. Since food, as well as fuel and housing, cost so much more in the large town, the wage expressed in money, must be much higher than in the small one. This adds appreciably to production costs." (pp 286-287 of the English translation)

This list is surprisingly comprehensive. In particular, the impact of high land rents and high food prices on monetary wages in large cities is explicitly spelled out. von Thünen then turns to the main forces that stand behind industrial agglomerations:

> " 1.Only in large-scale industrial plants is it profitable to install labour-saving machinery and equipment, which economise on manual labour and make for cheaper and more efficient production. 2. The scale of an industrial plant depends on the demand for its products. ...4. For all these reasons, large scale plants are viable only in the capital in many branches of industry. But the division of labour (and Adam Smith has shown the immense influence this has on the size of the labour product and on economies of production) is closely connected with the scale of an industrial plant. This explains why, quite regardless of economies of machine-production, the labour product per head is far higher in large than in small factories. ...7. Since it takes machines to produce machines, and these are themselves the product of many different factories and workshops, machinery is produced efficiently only in a place where factories and workshops are close enough together to help each other work in unison, i.e. in large towns. ...Economic theory has failed to adequately appreciate this factor. Yet it is this which explains why factories are generally found communally, why, even when in all other respects conditions appear suitable, those set up by themselves, in isolated places, so often come to grief. Technical innovations are continually increasing the complexity of machinery;and the more complicated the machines, the more the factor of association will enter into operation." (pp 287-290 of the English translation)

Observe that the combination of von Thünen's agglomeration forces 1, 2 and 4 almost coincides with Krugman's (1991a) "basic story" for the emergence of a core-periphery structure. Furthermore, if we combine these forces with the last one (7), which is about inter-industry linkages and technological spillovers, we get another fundamental explanation for the emergence of industrial agglomerations.

Even though von Thünen's work took place at the very beginning of the Industrial Revolution in Germany, it would be hard to imagine a more explicit description of the forces shaping the industrial landscape. Nonetheless, despite its originality, von Thünen's analysis falls short of results regarding the interactions among the various forces at work. In particular, it is not clear which agglomeration or dispersion forces one should expect to dominate depending on the underlying economic conditions. This limits our understanding in, at least, three respects. First, from the theoretical point of view, mathematical models are very useful tools in that they allow one to generate possible explanations of phenomena involving a large number of variables, such as cities or regional agglomerations. To be sure, economic models often imply strong simplifications. However, without making such simplifications, it would be hopeless to uncover the forces that really matter in shaping the space-economy. Furthermore, the interactions between the product and labor markets, which are central to the spatial organization of the economy, must be studied in the context of a general equilibrium model.[4]

At this stage, economic geographers might wonder: *why do geographical economists care so much about general equilibrium?* It is worth addressing this issue in some detail. First, it should be clear even to the reader unfamiliar with economic theory that prices are the main variables of a market economy. As economic decisions by firms and households are guided by price signals, prices determine the allocation of economic resources to different uses and thus overall welfare. Second, it should also be clear even to that reader that what happens in a specific market typically affects other markets. This is a straightforward implication of the fact that firms and households operate in many markets simultaneously. For example, increasing oil prices reduce household real income and, therefore, the quantities they are able to buy on other markets. Hence, it should not come as a surprise that economists have spent so much time understanding how prices may - or may not - clear a system of interdependent markets, which is the purpose of general equilibrium analysis.[5]

We need such a setting because *the spatial pattern of production and consumption* within some well-defined space-economy is, ultimately, what economic geographers (we believe) and geographical economists (for sure) want to explain.[6] Clearly, the decisions of many firms and many households to locate in a particular region both determines and is determined by the local prices of goods and factors. This makes those decisions necessarily interdependent and, therefore, they must be studied within a general

[4] Another illustration of the usefulness of the economic methodology for spatial analysis is provided by Duranton and Puga (2004) and Durlauf (2004), who survey what has recently been accomplished in the two very difficult areas, namely urban agglomeration economies and neighborhood effects. These two papers focus on a much smaller spatial scale in which nonmarket interactions are more important than in NEG.

[5] After almost 50 years, Koopmans (1957) remains a prominent reference regarding the relevance of general equilibrium.

[6] Martin (1999) is right when he claims that geographical economists are not sufficiently careful in defining their spatial unit of reference. The definition of the spatial unit of reference is a domain for the development of fruitful interactions between the two groups of scholars.

equilibrium framework.

A key result for developing such a framework is the Spatial Impossibility Theorem, which argues that a system of spatially separated interdependent markets cannot be cleared by a perfectly competitive price mechanism. Thus, if one wants to explain reality through general equilibrium analysis, there is only one way through: to acknowledge that markets are imperfect. This could be due to production and/or consumption externalities. However, we think that such externalities are likely to be critical in the small but not essential in the large. Our opinion is backed by the many empirical analyses undertaken during the last decade and all suggesting that spillover effects are very localized (Audretsch and Feldman, 2004; Rosenthal and Strange, 2004). Accordingly, we prefer to stress another source of imperfection: the market power of firms facing increasing returns at the plant level.

Here is the reason for the success of the Dixit-Stiglitz model in NEG (as well as in other economic fields such as trade and growth where the competitive model had shown its limits). The Dixit-Stiglitz model encompasses imperfect competition and increasing returns and, thus, offers an alternative to the Arrow-Debreu framework that has dominated economics for so long. When we deal with the location of firms and consumers, it is unavoidable to work at a very disaggregate level as locational decisions are agent-specific and based on price and wage differences (ignoring amenities and nontradeable goods available at each location). As seen above, what happens on the product market is crucial to determine what wage firms may pay their workers, whereas individual incomes are fundamental in determining the price level that prevails on the product market. In other words, both markets are interrelated and how they interact in a region also affects the way they operate in another, and vice versa. This is precisely what the spatialized version of the Dixit-Stiglitz model developed by Krugman allows us to accomplish.

More precisely, firms and households being free to choose their location, *the size of local markets is endogenous in NEG*. This, in turn, allows it to explain *why* and *how* agglomerations of different types may emerge (Fujita and Thisse, 2002). To the best of our knowledge, NEG is the first body of theory that provides full-fledged microeconomic models of agglomeration of a system of spatial markets with mobile agents and/or factors. For this reason, NEG may be viewed as a reformulation of the research program launched by Lösch, which economists, geographers, and regional scientists all together did not push sufficiently far before Krugman's contribution. The reason is that they lacked the appropriate analytical tools to address such questions, *especially when dealing with market structure and the formation of prices.* In this respect, it is worth noting that the main difference between, on the one hand, fairly old - but still useful - approaches such as the gravity model or Harris's market potential and NEG on the other hand, lies in the role played by prices and markets: competition is central to NEG whereas prices are just not accounted for in gravity or market potential models (not either in standard location theory à la Weber). However, economic geographers and location theorists cannot ignore prices in their analyses because they are *fundamental* for understanding how a spatial market-economy operates.

Nonetheless, many challenges mark NEG's road ahead. From a theoretical point of view, NEG offers a very poor framework in terms of spatial structure. Its extension to the case of a system of n asymmetrically distributed regions is called for (see Behrens et al, 2004, for a first attempt in this direction). Another major weakness of NEG lies in the fact that it rests on a very specific model, so exploring alternative frameworks is needed to test the robustness of its main conclusions. This is why we have proposed another model that, fortunately, leads to results qualitatively similar to those obtained by Krugman (Ottaviano et al, 2002; Ottaviano and Thisse, 2004). Finally, NEG does not account for the nonmarket interactions that are part of the explanation of the success or failure of regions (Duranton and Puga, 2004; Scott and Storper, 2003). Recall, however, that NEG was not designed for that, its main purpose being to provide an integrated framework explaining the *macrospatial* organization of economic activities.

From an empirical point of view, the test of the existing explanations of industry location is generally crippled by identification problems due to observational equivalence. In such a context, mathematical models should help design simultaneous-equation regressions in which theoretically driven cross-equation constraints can be used to identify alternative theories (see, e.g. Rosenthal and Strange, 2004 as well as Head and Mayer, 2004).

From a policy point of view, mathematical models highlight the direct and indirect effects of public intervention. For example, NEG models have clarified that all sorts of non-regional policies can have 'regional side-effects', that is, a potentially large impact on the location of economic activities and, thus, on the geographical distribution of wealth (Ottaviano, 2003). More evidence is provided in the fourth volume of the *Handbook of Regional and Urban Economics*, where several chapters address such issues. All in all, despite several shortcomings, we find it fair to say that NEG has been pretty successful in updating Lösch's program as well as in integrating location theory and international economics in a way that Ohlin would probably not have disapproved.

To sum up, here lies the main contribution of NEG: "to combine old ingredients through a new recipe". Having said so, we expect the ideas discussed in the foregoing to become 'common knowledge'. If there is no shame in rediscovering science, there is a certain waste of resources that, hopefully, will be avoided in the future. But then, where do we go from here? Well, this is another story. However, we believe that, for some common research agenda to emerge, geographical economists and economic geographers should, at least, agree on the fact that NEG is sufficiently tasty to be part of the menu.

References

[1] Audretsch D, Feldman M , 2004, "Knowledge spillovers and the geography of innovation" in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) pp 2714-2739

[2] Baldwin RE, Forslid R, Martin Ph, Ottaviano GIP, Robert-Nicoud F, 2003 *Economic Geography and Public Policy* (Princeton University Press, Princeton)

[3] Beckmann MJ, 1976, "Spatial equilibrium in the dispersed city" in *Mathematical Land Use Theory* Ed YY Papageorgiou (Lexington Books, Lexington) pp 117-125

[4] Behrens K, Lamorgese A, Ottaviano GIP, Tabuchi T, 2004, "Testing the home market effect in a multi-country world: the theory" CEPR Discussion Paper No. 4468

[5] Brakman S, Garrensein H, Van Marrewijk Ch, 2001 *An Introduction to Geographical Economics* (Cambridge University Press, Cambridge)

[6] Chamberlin EH, 1933 *The Theory of Monopolistic Competition* (Harvard University Press, Cambridge MA)

[7] d'Aspremont C, Gabszewicz JJ, Thisse J-F, 1979, "On Hotelling's "Stability in Competition"" *Econometrica* **47** 1045-1050

[8] d'Aspremont C, Gabszewicz JJ, Thisse J-F, 1983, "Product differences and prices" *Economics Letters* **11** 19-23

[9] de Palma A, Ginsburgh V, Papageorgiou YY, Thisse J-F, 1985, "The principle of minimum differentiation holds under sufficient heterogeneity" *Econometrica* **53** 767-781

[10] Dixit AK, Stiglitz JE, 1977, Monopolistic competition and optimum product diversity *American Economic Review* **67** 297-308

[11] Duranton G, Puga D, 2004, "Micro-foundations of urban agglomeration economies" in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) 2064-2117

[12] Durlauf SN, 2004, "Neighborhood effects" in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) 2173-2242

[13] Eaton BC, Lipsey RG, 1977, "The introduction of space into the neoclassical model of value theory" in *Studies in Modern Economics* Eds M Artis, A Nobay (Basil Blackwell, Oxford) pp 59-96

[14] Fujita M, 1988, "A monopolistic competition model of spatial agglomeration: a differentiated product approach" *Regional Science and Urban Economics* **18** 87-124

[15] Fujita M, Krugman P, Venables AJ, 1999 *The Spatial Economy. Cities, Regions and International Trade* (The MIT Press, Cambridge MA)

[16] Fujita M, Thisse J-F, 2002 *Economics of Agglomeration. Cities, Industrial Location, and Regional Growth* (Cambridge University Press, Cambridge)

[17] Gabszewicz JJ, Thisse J-F, 1986, "Spatial competition and the location of firms" in *Location Theory* Eds JJ Gabszewicz, J-F Thisse, M Fujita, U Schweizer (Harwood Academic Publishers, Chur) pp 1-71

[18] Gabszewicz JJ, Thisse J-F, 1992, "Location" in *Handbook of Game Theory with Economic Applications, Volume 1* Eds RE Aumann, S Hart (North-Holland, Amsterdam) pp 281-304

[19] Haig RM, 1926, "Toward an understanding of the metropolis. I. Some speculations regarding the economic basis of urban concentration" *Quarterly Journal of Economics* **40** 179-208

[20] Hakimi SL, 1964, "Optimum location of switching centers and the absolute centers and medians of a graph" *Operations Research* **13** 462-475

[21] Hansen P, Labbé M, Peeters D, Thisse J-F, 1987, "Facility location analysis" in *Systems of Cities and Facility Location* Eds P Hansen, M Labbé, D Peeters, J-F Thisse, JV Henderson (Harwood Academic Publishers, Chur) pp 1-70

[22] Harris C, 1954, “The market as a factor in the localization of industry in the United States” *Annals of the Association of American Geographers* **64** 315-348

[23] Head K, Mayer T, 2004, “The empirics of agglomeration and trade” in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) 2609-2669

[24] Heckscher E, 1918 *The Continental System: An Economic Interpretation* (English translation: Clarendon Press, Oxford)

[25] Helpman E, Krugman PR, 1985 *Market Structure and Foreign Trade* (The MIT Press, Cambridge, MA).

[26] Henderson JV, 1988 *Urban Development. Theory, Fact and Illusion* (Oxford University Press, Oxford)

[27] Hirschman AO, 1958 *The Strategy of Development* (Yale University Press, New Haven, CN)

[28] Hoover EM, 1936 *Location Theory and the Shoe and Leather Industries* (Harvard University Press, Cambridge MA)

[29] Hotelling H, 1929, “Stability in competition” *Economic Journal* **39** 41-57

[30] Irwin DA, 1996 *Against The Tide. An Intellectual History of Free Trade* (Princeton University Press, Princeton)

[31] Isard W, 1956 *Location and Space-economy* (The MIT Press, Cambridge MA)

[32] Isserman AM, 1996, ““It’s Obvious, It’s Wrong, and Anyway They Said It Years Ago”? Paul Krugman and large cities” *International Regional Science Review* **19** 37-48

[33] Kaldor N, 1935, “Market imperfection and excess capacity” *Economica* **2** 35-50

[34] Kaldor N, 1970, “The case for regional policies” *Scottish Journal of Political Economy* **17** 337-348

[35] Koopmans TC, 1957 *Three Essays on the State of Economic Science* (McGraw-Hill, New York)

[36] Krugman PR, 1991a, "Increasing returns and economic geography" *Journal of Political Economy* **99** 483-499

[37] Krugman PR, 1991b, "History versus expectations" *Quarterly Journal of Economics* **106** 651-667

[38] Krugman PR, 1993, "On the relationship between trade theory and location theory" *Review of International Economics* **1** 110-122

[39] Krugman PR, 1995 *Development, Geography, and Economic Theory* (MIT Press, Cambridge MA).

[40] Kuehn AA, Hambuger MJ, 1963, "A heuristic program for locating warehouses" *Management Science* **9** 643-666

[41] Lampard EE, 1955, "The history of cities in the economically advanced areas" *Economic Development and Cultural Change* **3** 321-342

[42] Lerner A, Singer HW, 1937, "Some notes on duopoly and spatial competition" *Journal of Political Economy* **45** 145-186

[43] Lösch A, 1940 *Die Räumliche Ordnung der Wirtschaft* (Gustav Fischer, Jena). English translation *The Economics of Location* (Yale University Press, New Haven CN, 1954)

[44] Louveaux F, Thisse J.-F., Beguin H, 1982, "Location theory and transportation costs" *Regional Science and Urban Economics* **12** 529-545

[45] Marshall A, 1890 *Principles of Economics* (Macmillan, London) 8th edition published in 1920

[46] Martin R, 1999, "The new 'geographical turn' in economics: some critical reflections" *Cambridge Journal of Economics* **23** 65-91

[47] Matsuyama K, 1995, "Complementarities and cumulative process in models of monopolistic competition" *Journal of Economic Literature* **33** 701-729

[48] Mill J, 1821 *Elements of Political Economy* (Baldwin, Cradock and Joy, London).

[49] Muth RF, 1971, "Migration: chicken or egg?" *Southern Economic Journal* **3** 295-306

[50] Myrdal G, 1957 *Economic Theory and Underdeveloped Regions* (Duckworth, London)

[51] Ohlin B, 1933 *Interregional and International Trade* (Harvard University Press, Cambridge MA) Revised version published in 1968

[52] Ohlin B, Hesselborn PO, Wijkman PM, eds., 1977 *The International Allocation of Economic Activity* (Macmillan, London)

[53] Ottaviano GIP, 2003, "Regional policy in the global economy: insights from New Economic Geography" *Regional Studies* **37** 665-674

[54] Ottaviano GIP, Tabuchi T, and Thisse JF, 2002, "Agglomeration and trade revisited" *International Economic Review* **43** 409-436

[55] Ottaviano GIP, Thisse JF, 2004, "Agglomeration and economic geography" in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) 2564-2608

[56] Papageorgiou YY, Thisse J-F, 1985, "Agglomeration as spatial interdependence between firms and households" *Journal of Economic Theory* **37** 19-31

[57] Pinkse J, Slade ME, Brett C, 2002, "Spatial price competition: a semiparametric approach" *Econometrica* **70** 1111-1153

[58] Ricardo D, 1817, *The Principles of Political Economy and Taxation* (John Murray, London)

[59] Rosenstein-Rodan PN, 1943, "Problems of industrialization of Eastern and South-Eastern Europe" *Economic Journal* **53** 202-11

[60] Rosenthal S, Strange W, 2004, "Evidence on the nature and sources of agglomeration economies" in *Handbook of Regional and Urban Economics. Volume 4* Eds J.V. Henderson, JF Thisse (Amsterdam, North-Holland) 2120-2171

[61] Sakashita N, 1967, "Production function, demand function and location theory of the firm" *Papers and Proceedings of the Regional Science Association* **20** 109-129

[62] Scott AJ, Storper M, 2003, "Regions, globalisation, development" *Regional Studies* **37** 579-593.

[63] Starrett D, 1978, "Market allocations of location choice in a model with free mobility" *Journal of Economic Theory* **17** 21-37

[64] Stahl K, 1983, "A note on the microeconomics of migration" *Journal of Urban Economics* **14** 318-326

[65] Stahl K, 1987, "Theories of urban business location" in *Handbook of Regional and Urban Economics, Volume 2* Ed. ES Mills (Amsterdam, North-Holland) pp 759-820

[66] Stollsteimer JF, 1963, "A working model for plant numbers and locations" *Journal of Farm Economics* **45** 631-645

[67] Tellier LN, Polansky B, 1989, "The Weber problem: frequency of different solution types and extensions to repulsive forces and dynamic processes" *Journal of Regional Science* **29** 387-405

[68] von Thünen JH, 1826 *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie* (Perthes, Hamburg) English translation *The Isolated State* (Pergammon Press, Oxford)

[69] Torrens D, 1815 *Essay on the External Corn Trade* (J. Hatchard, London)

[70] Ullman E, 1941, "A theory of location for cities" *American Journal of Sociology* **46** 853-864

[71] Venables AJ, 1996, "Equilibrium locations of vertically linked industries", *International Economic Review* **37** 341-359

[72] Weber A, 1909 *Über den Standort der Industrien* (J.C.B. Mohr, Tübingen) English translation: *The Theory of the Location of Industries* (Chicago University Press, Chicago, 1929)

[73] Witzgall C, 1964 *Optimal location of a central facility: mathematical models and concepts* National Bureau of Standard Report 8388, Washington D.C.